

Lecture Notes
Physics



Analytical Mechanics

Lagrange Mechanics for Advanced Students

SANDRA P. KLEVANSKY AND
RICHARD H. LEMMER



HEIDELBERG
UNIVERSITY PUBLISHING

Analytical Mechanics

Lecture Notes **Physik**



Analytical Mechanics

Lagrange Mechanics
for Advanced Students

SANDRA P. KLEVANSKY AND RICHARD H. LEMMER


HEIDELBERG
UNIVERSITY PUBLISHING

About the Authors

Sandra P. Klevansky is a retired Director of the Heidelberg Graduate School for Physics and Department of Physics and Astronomy, and adjunct Professor for Theoretical Physics at Heidelberg University. Her interests have ranged from theoretical nuclear and particle physics to non-Hermitian formulations of quantum field theory.

Richard H. Lemmer (deceased) was Professor of Mathematical Physics at the University of the Witwatersrand in Johannesburg. His international career had many stations, one of the longer being 10 years at the Massachusetts Institute of Technology (MIT), USA, before he returned to his country of birth, South Africa. He has made many significant contributions to nuclear and particle physics.

ORCID®

Sandra P. Klevansky  <https://orcid.org/0000-0003-4208-9472>

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.dnb.de>.



This book is published under the Creative Commons Attribution License CC BY-SA 4.0. The cover is subject to the Creative Commons License CC BY-ND 4.0.

Published at Heidelberg University Publishing (heiUP), 2026

Heidelberg University/Heidelberg University Library
Heidelberg University Publishing (heiUP)
Grabengasse 1, 69117 Heidelberg, Germany
<https://heiup.uni-heidelberg.de>

The electronic open access version of this work is permanently available on Heidelberg University Publishing's website: <https://heiup.uni-heidelberg.de>

urn: urn:nbn:de:bsz:16-heiup-book-1719

doi: <https://doi.org/10.17885/heiup.1719>

Text © 2026, Sandra P. Klevansky and Richard H. Lemmer

ISSN 2566-4816

eISSN 2512-4455

ISBN 978-3-96822-362-9 (Softcover)

ISBN 978-3-96822-361-2 (PDF)

For our families

Contents

<i>List of Figures</i>	9
<i>List of Tables</i>	13
<i>Preface</i>	15
<i>Chapter 1 Principles of Dynamics</i>	17
<i>Chapter 2 Particle Dynamics</i>	59
<i>Chapter 3 Rigid Body Dynamics</i>	103
<i>Chapter 4 Small Oscillations</i>	145
<i>Chapter 5 Fluid Motion</i>	205
<i>Chapter 6 Relativistic Mechanics</i>	223
<i>Chapter 7 Hamiltonian Mechanics</i>	247

List of Figures

- 1.1 The particle at P moves with velocity \mathbf{v} with respect to Σ and velocity \mathbf{v}' with respect to Σ' . 19
- 1.2 The family of curves $\bar{f}(t) = \frac{1}{2}gt_2^2(\frac{t}{t_2})^\beta$ for $\beta = \frac{1}{2}, 1, 2, 3, 4$. 25
- 1.3 The function S (multiplied by the constant $8/mg^2t_2^3$) is shown as a function of the parameter β . 25
- 1.4 A particle of mass m constrained to move in a vertical circular path in the gravitational field of the earth. 34
- 1.5 The particle of mass m constrained to a vertical circular path and acted on by gravity. 39
- 1.6 Bead on a rotating hoop. 44
- 1.7 Illustration of two possible paths of the representative point for a system with three degrees of freedom. 48
- 1.8 Application of Fermat's principle to the motion of a particle. 50

- 2.1 The potential V of a harmonic oscillator. 63
- 2.2 The potential V with a barrier. 63
- 2.3 The contour of integration C . 64
- 2.4 A particle moving in a trajectory measured from the fixed point O . 66
- 2.5 A possible central orbit. 68
- 2.6 The potential U showing positive and negative energies. 70
- 2.7 A potential $U \sim 1/r^2$ that allows for a positive energy only. 71
- 2.8 The particle of mass m moving in a vertical circular path, and acted on by gravity. 71
- 2.9 Attractive and repulsive motion. 74
- 2.10 Bound state in the effective potential. 75
- 2.11 The geometry of a scattering problem. 76
- 2.12 Geometry of a scattering center. 77
- 2.13 Geometry for scattering by a square well. 79
- 2.14 Relative and center-of-mass coordinates for the two-body problem. 80
- 2.15 Vectors in the center-of-mass system. 82
- 2.16 Vectors in the center-of-mass system: B moves to C . 83
- 2.17 Cylindrical coordinate system with external magnetic field \mathbf{B} indicated. 91

- 2.18 The angular momentum vector \mathbf{L} precesses about the direction of the field \mathbf{B} . 92
- 2.19 Coordinate vectors. 93
- 2.20 Coordinate axes x', y', z' for describing motion relative to the earth's surface. 95
- 2.21 Foucault's pendulum. 96
- 2.22 Projected path of the bob of Foucault's pendulum. 97
- 3.1 Coordinates of body-fixed system relative to the space-fixed system. 104
- 3.2 The origins of the body-fixed and space-fixed systems coincide, $O = O'$. 109
- 3.3 The body-fixed and space-fixed origins coincide ($O = O'$). In addition, the z axes are common $z = z'$. The x and y axes are rotated by ϕ to x' and y' . 110
- 3.4 Upper panel: Active point of view. $A(\phi)$ rotates \mathbf{Q} closer to the x axis, Lower panel: Passive point of view. $A(\phi)$ rotates the x axis closer to the vector \mathbf{Q} . 114
- 3.5 Euler angles. 115
- 3.6 Rotation of the top with space-fixed axes. 122
- 3.7 Cuspidal motion of top. 126
- 3.8 Directional change of angular momentum. 129
- 3.9 Rotating top ((a) and (b)) and rotating egg ((c) and (d)). 132
- 3.10 The tippe top. 133
- 3.11 Sketch of two cylinders touching at a tangent point. 142
- 4.1 Coupled pendulums. 145
- 4.2 A linear XY_2 molecule. 157
- 4.3 Normal modes of a linear XY_2 molecule. 160
- 4.4 The molecule X_3 . 162
- 4.5 Normal modes of an X_3 molecule. 164
- 4.6 A rotation about \mathbf{e}_3 through an angle α . 171
- 4.7 Reflections. P is taken into P' by σ_h , into P'' by σ_v . 172
- 4.8 Symmetry operations for the X_3 molecule. 173
- 4.9 A rotation of the point P through an angle α in the $x - y$ plane. 175
- 4.10 Reflection of the point P through a plane passing through \mathbf{e}_3 . 176
- 4.11 Action of C_3 on X_3 in a displaced configuration. 177
- 4.12 Structure of the \mathbf{W} matrix for NH_3 . 186
- 4.13 Linear chain of vibrating particles. 190
- 4.14 Geometry connecting two neighbouring particles. 191
- 4.15 Transverse normal modes of three particles in a linear chain. 192
- 5.1 Coordinate system for discussing water waves on an open sea. 212
- 5.2 Deviations from the surface are designated by η . 213
- 5.3 Phase velocity - wavelength relations for water waves. 215
- 5.4 Initial disturbance of pond. 216

- 5.5 Contour C for the integral. 219
- 6.1 The inertial frames Σ and Σ' . 224
- 6.2 The energy-momentum relation for a relativistic free particle. 238
- 6.3 Experimental confirmation of the relativistic relation between the speed and the kinetic energy of relativistic electrons. The open circles denote the experimental points. 239
- 7.1 Area-preserving ellipses in the phase space of a free particle, shown at three equal intervals of time. System points inside the circular boundary at $t = 0$ remain inside the corresponding elliptic boundary as time increases. 278

List of Tables

- 4.1 A group multiplication table for example (ii) of the text. 165
- 4.2 A character table. 170
- 4.3 Character table for C_{3v} . 174
- 4.4 Character table for C_{3v} and characters for the NH_3 molecule. 186

Preface

Probably as many reasons have been given for writing books on classical mechanics as there are books on the subject. So perhaps we will be forgiven for not adding to the list of reasons as we add to the already excellent and extensive list of books devoted to this subject.

The present text is based in a large part on lectures given on the subject of classical mechanics to senior students. It represents a physicist's view of mechanics, presented to an audience of potential physicists (not mathematicians) who would one day earn their keep in physics. With these initial conditions it became clear that emphasizing the application of the principles of mechanics to problems usually encountered by the physicists is to be preferred over an exposition that stresses the more mathematical aspects of mechanics in their own right. So, some readers may well judge a given mathematical proof, let us say in the calculus of variations, for example, as lacking rigor. That is fine we say. Let that person dig deeper into books on that particular subject. But it is well to remember that there are various levels of rigor in mathematical demonstrations (the last one being *rigor mortis* as was once pointed out to us!) within a subject whose foundations rest in the final analysis on experience and experiment. These remarks must in no way be construed as condoning a careless attitude in *using* mathematics in the process of reasoning in mechanics. We simply stress that mechanics is not mathematics and one should not confuse the two.

It is also well to remark on the choice of material and the emphasis within that choice, that is to be found here. Since this book is aimed at senior students who have already had some substantial experience in the methodology of mechanics, up to and including Lagrangian mechanics, our attitude is one of leading the already converted into the more elegant formulations of the subject, based on variational principles of various kinds. This approach leads naturally to the introduction of the Lagrange function and Lagrange's equations again. Once the Lagrange function, its properties, and the Lagrange equations have been established, we take pains to tackle every new problem from this vantage point. Thus, the build-up from point mechanics to rigid bodies to small vibrations to fluids all happens from a unified point of view.

Special Relativity is a traditional and essential part of mechanics today. However, it is also a somewhat less familiar element, and consequently the subject of a more elementary discussion. The same comment is true for the discussion of fluid motion. We end the book with Hamiltonian mechanics, including a discussion of perturbations and Hamilton-Jacobi theory which is usual, as well as a demonstration of the introduction of *uncertainty* in mechanics, due to Max Born, which is unusual.

Apart from Chapter 1, which is essential for understanding the rest of the book, each succeeding chapter is more or less independent of its predecessor.

A comment on problems in mechanics is also relevant. The text contains a fair number of worked examples that usually serve as illustrations of the principle on which the problem is based. A list of problems will also be found at the end of each chapter. The purpose of most of these problems is again to illustrate some aspect not fully discussed in the text itself. It has not been our purpose to compile an exhaustive list of problems, although we urge the reader to find and do as many as they can lay their hands on. A very useful source of physical problems in classical mechanics is G.L. Kotkin and V.G. Serbo, *Exploring Classical Mechanics*, (Oxford University Press, Oxford, 2020).

In writing a book of this nature, one is inevitably influenced by teachers of, and books on the same subject. As with most textbooks, this one also grew out of coursework. All students in a course play a larger role than they often think in influencing the purpose (and horizons) of that course, and one of us (SPK) was in fact one of the subjects on which this course was practised. This book has taken an extremely long time in its production, due to our other obligations, with the unfortunate consequence that RHL has in the meantime sadly passed away. Consequently, I am hoping that, looking down from his place in physics heaven, he approves of all the changes that have been made.

Finally a special word of thanks is due to all those student assistants who have painstakingly helped put this manuscript together.

S.P. Klevansky

Heidelberg and Johannesburg

2025

Chapter 1 Principles of Dynamics

1-1 Preamble

Classical mechanics is a truly vast subject. That being so, the material that appears in any text of this size and purpose has of necessity been through a severe selection process. The choice of topics to be presented here has been strongly influenced by a perceived relevance to modern physics. This might sound unfair to our subject: The builders of classical mechanics certainly never did intend their work to be seen only as stepping stones on the way to quantum mechanics and field theory. Yet much of it is.

We have also attempted to keep the discussion of mechanics as uniform as possible by using a Lagrangian formulation throughout. In many instances this restricts one to considering only conservative systems, or at least systems for which a Lagrange function can be constructed. This is hardly a disadvantage. Most of the systems of interest in physics are of this type; furthermore the generalizations to cover particular non-conservative systems are usually straightforward.

1-2 The Ingredients of Mechanics

We start by considering the motion of a single particle under the action of given forces, thereby introducing the two main ingredients of mechanics, the concepts of *particle* and *force*. A *qualitative* idea of a force is quite natural and instinctive, based on everyday experience. Gravitational forces and their effects are certainly familiar to anyone who has ever tried to defy them. But other forces also exist in nature that are not so easily experienced; electromagnetic forces that are responsible for the structure of matter in bulk, and forces which govern the dynamics inside the atomic nucleus and its components itself. What we will have to say in the rest of this book is generally applicable for any force system. In practice, of course, the natural forces that occur in classical mechanics are basically of two types: gravitational and electrical (even the action of a spring, or the friction on a rough surface can in the last analysis be traced back to the properties of the material involved, i.e. to the electrical

forces between atoms). Examples of man-made forces in engineering applications like steam pressure, hydraulic systems, etc. abound of course.

Experience tells us that forces have both magnitude and direction: they are to be represented as *vector quantities*. But how does one measure a force, or compare the action of different forces? This is the *quantitative* aspect of the problem. The answer is also obvious: we must do an experiment. For this purpose we introduce the other ingredient mentioned at the beginning of this section: the idea of a *particle*, or *mass point*¹, whose internal structure is characterized by a single parameter m called its *mass*. The extent to which actual material bodies in the laboratory approximate particles is determined by to what extent properties other than their mass play a role in determining their motion.

¹ Sometimes also described as a *material point*.

To investigate the motion of a particle, we must measure its velocity \mathbf{v} and changes in its velocity as a function of time, $\dot{\mathbf{v}}$. In order to perform such measurements the particle must be located in some *frame of reference* by a position vector \mathbf{r} . As the particle moves, \mathbf{r} changes. The rates of change of $\mathbf{r}(t)$ and $\mathbf{v}(t)$, given by

$$\mathbf{v} = \frac{d}{dt}\mathbf{r}(t), \quad \dot{\mathbf{v}} = \frac{d^2}{dt^2}\mathbf{r}(t),$$

represent the *velocity* and *acceleration* of the particle at time t . However, it is not the velocity \mathbf{v} , but rather the product $\mathbf{p} = m\mathbf{v}$, called the *linear momentum*, that plays a central role in determining the motion of the particle. For experience shows that \mathbf{p} stays constant if no forces act on the particle, i.e. the motion is governed by the condition

$$\mathbf{p} = \text{constant}. \quad (1.1)$$

This is *Newton's first law of motion* or law of inertia. The effect of an applied force \mathbf{F} acting on the particle is to change the momentum \mathbf{p} . The rate of change, $\dot{\mathbf{p}} = d\mathbf{p}/dt$, is dictated by *Newton's second law of motion*,

$$\dot{\mathbf{p}} = \mathbf{F} \quad (1.2)$$

This equation is the basic law governing the motion or *dynamics* of material systems. It is often called the *principle of linear momentum*.

We saw that it was necessary to specify a frame of reference in which to measure \mathbf{p} and hence $\dot{\mathbf{p}}$. Are the laws of motion (1.1) and (1.2) valid irrespective of what reference frame is used? Experience shows that this is not the case. Newton's first two laws of motion only assume the simple forms given in special frames of reference, called *inertial frames*. In fact, (1.1) can be turned around to provide a definition of an inertial frame: a free particle moves with constant momentum in an inertial frame. In particular, a free particle initially at rest remains at rest in such a frame².

² A better definition of an inertial frame of reference can be given in General Relativity: The gravitational field vanishes in the immediate neighborhood of an inertial frame, whether or not this field is caused by gravitating masses or the non-uniform motion of the frame of reference.

The next question is very important. How many inertial frames are there? Suppose the point O is the origin of a set of rectangular axes O_{xyz} that form an inertial frame Σ . Now introduce a second frame made up of rectangular axes $O'_{x'y'z'}$ moving in a straight line with constant velocity \mathbf{V} with respect to Σ . A particle at P which has position coordinate $\mathbf{r} = OP$ in Σ has position coordinate $\mathbf{r}' = O'P$ in Σ' , see Fig. 1.1, where

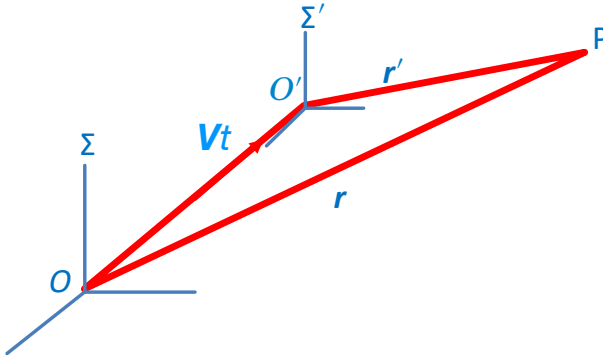


Figure 1.1: The particle at P moves with velocity \mathbf{v} with respect to Σ and velocity \mathbf{v}' with respect to Σ' .

$$\mathbf{r} = \mathbf{r}' + \mathbf{V}t \quad (1.3)$$

if Σ and Σ' coincide at $t = 0$. Hence,

$$\mathbf{v} = \mathbf{v}' + \mathbf{V} \quad (1.4)$$

where \mathbf{v}' is the velocity of the particle as measured in Σ' . Since \mathbf{V} is constant in time, $\dot{\mathbf{v}} = \dot{\mathbf{v}}'$ or $\dot{\mathbf{p}} = \dot{\mathbf{p}}'$. Thus, Newton's laws of motion hold in Σ' if they hold in Σ , if we assume that the force is the same in Σ and Σ' , i.e. $\mathbf{F} = \mathbf{F}'$. All frames Σ' moving with constant velocity in a straight line are inertial frames, i.e. there are infinitely many inertial frames if there is one such frame, and Newton's laws are valid in all these frames. This is *Galileo's principle of relativity*. Stated another way, Galileo's principle asserts that the laws of motion are invariant in form under the transformation of coordinates (1.3). Equation (1.3) is called a *Galilean transformation*. Notice the tacit assumption

$$t = t' \quad (1.5)$$

in (1.4): the passage of time in the two frames of reference is the same. This is the classical point of view of the absolute nature of time in Galilean relativity. It will undergo a profound modification when we discuss the theory of relativity.

In the laboratory, the question of an inertial reference frame is easily settled: for most practical purposes a reference frame attached rigidly to the earth's surface will suffice. The effects coming from the earth's rotation are small and can be corrected for when necessary.

Laws 1 and 2 must be supplemented by two further statements. These are contained in

Law 3: *Action and reaction are equal*, or the forces exerted by two bodies in mutual interaction are equal and opposite, and

Law 4: The superposition principle for forces asserts that *forces add like vectors*. The content of this statement is simply that when several forces act simultaneously on a particle, each force causes changes in motion as if it alone were present; the other forces do not interfere with its action. It is a remarkable fact that these four laws of motion summarize the whole body of experience of experimental mechanics.

1-3 Systems of Particles

So far, we have spoken of a single particle and its motion under the action of an applied force. Now, consider what happens if the four laws of motion are applied to a system of interacting particles. We fix our attention on particle a of this system, moving with momentum \mathbf{p}_a . The forces on a arise from

- (i) the sum $\sum_{b \neq a} \mathbf{F}_{ab}$ of all internal forces acting on a due to its interaction with all the other particles in the system. \mathbf{F}_{ab} is the force exerted by b on a . By law 3, it is equal and opposite to the force exerted by a on b , $\mathbf{F}_{ab} = -\mathbf{F}_{ba}$.
- (ii) the sum of all external forces $\mathbf{F}_a^{(e)}$ that are applied from outside the system. The motion of particle a is thus determined by

$$\dot{\mathbf{p}}_a = \mathbf{F}_a^{(e)} + \sum_b \mathbf{F}_{ab} \quad (1.6)$$

so that

$$\sum_a \dot{\mathbf{p}}_a = \sum_a \mathbf{F}_a^{(e)} + \sum_a \sum_{b \neq a} \mathbf{F}_{ab}$$

summing over all particles. Now, $\sum_a \mathbf{p}_a = \mathbf{P}$ is the total linear momentum of the system, $\sum_a \mathbf{F}_a^{(e)} = \mathbf{F}$ the total *external* force on the system, and $\sum_a \sum_{b \neq a} \mathbf{F}_{ab} = 0$ because the internal forces cancel in pairs. Hence, the principle of linear momentum for a composite system of particles reads

$$\dot{\mathbf{P}} = \mathbf{F}. \quad (1.7)$$

Notice that the internal forces cannot effect changes in the total momentum; they simply drop out. In fact this equation has a simple meaning. If we locate particle a with the position vector $\mathbf{r}_a(t)$ in an inertial frame, the center of mass of the system is located at \mathbf{R} :

$$M\mathbf{R} = \sum_a m_a \mathbf{r}_a(t), \quad M = \sum_a m_a \quad (1.8)$$

so that

$$\sum_a m_a \dot{\mathbf{r}}_a = \sum_a \mathbf{p}_a = M\dot{\mathbf{R}} = \mathbf{P}. \quad (1.9)$$

Thus, \mathbf{P} is the total momentum of the system. If the external forces are removed, the momentum \mathbf{P} remains constant no matter what motions the individual particles perform under their mutual interactions. Notice in passing that the definition of total momentum in (1.9) holds whether or not the particles are interacting.

Next, we derive the *principle of angular momentum*. The *angular momentum* of particle a about the origin O of the coordinate system for \mathbf{r}_a is defined by the vector product³.

$$\mathbf{r}_a \times \mathbf{p}_a = \mathbf{L}_a. \quad (1.10)$$

The time rate of change of \mathbf{L}_a is

$$\dot{\mathbf{L}}_a = (\dot{\mathbf{r}}_a \times \mathbf{p}_a) + (\mathbf{r}_a \times \dot{\mathbf{p}}_a). \quad (1.11)$$

The first vector product vanishes since $\dot{\mathbf{r}}_a$ and \mathbf{p}_a are parallel. For the second, we insert the dynamical information on $\dot{\mathbf{p}}_a$ from (1.6) and get

$$\dot{\mathbf{L}}_a = \mathbf{N}_a + \mathbf{N}'_a \quad (1.12)$$

where

$$\mathbf{N}_a = (\mathbf{r}_a \times \mathbf{F}_a^{(e)}) \quad \text{and} \quad \mathbf{N}'_a = (\mathbf{r}_a \times \sum_{b \neq a} \mathbf{F}_{ab}) \quad (1.13)$$

are the force moments or *torques* about 0 of the external and internal forces acting on the a th particle. As in the case of linear momentum, we can define the total angular momentum, $\sum_a \mathbf{L}_a = \mathbf{L}$, the total external torque $\sum_a \mathbf{N}_a = \mathbf{N}$ and find from (1.12) that

$$\dot{\mathbf{L}} = \mathbf{N} \quad (1.14)$$

since $\sum_a \mathbf{N}'_a = 0$ if \mathbf{F}_{ab} is directed along ab . This equation expresses the principle of angular momentum for the composite system.

The angular momentum and torque vectors \mathbf{L} and \mathbf{N} refer to the same fixed point in an inertial frame of reference Σ , and change in value if this point is changed. However, according to Galileo's principle of relativity, the principle of angular momentum should hold in any inertial frame. We express this analytically by saying that (1.14) is invariant in form under the Galilean transformation. To verify this, note that, by (1.3) and (1.4), the angular momentum $\mathbf{L} = \sum_a \mathbf{r}_a \times \mathbf{p}_a$ in Σ is related to the angular momentum \mathbf{L}' in the moving frame Σ' through

$$\mathbf{L} = \mathbf{L}' + \left(\sum_a m_a \mathbf{r}'_a \right) \times \mathbf{V} + \mathbf{R} \times \left(\sum_a \mathbf{p}'_a \right) + \mathbf{R} \times m_a \mathbf{V}, \quad (1.15)$$

where $\mathbf{L}' = \sum_a \mathbf{r}'_a \times \mathbf{p}'_a$ and $\mathbf{R} = \mathbf{V}t$ is the distance between the origins of Σ and Σ' after a time t . Since \mathbf{R} and \mathbf{V} are parallel, the last term in (1.15) drops away. The time derivative of \mathbf{L}' is simple to calculate:

$$\dot{\mathbf{L}}' = \dot{\mathbf{L}} - \mathbf{R} \times \sum_a \dot{\mathbf{p}}'_a = \sum_a (\mathbf{r}_a - \mathbf{R}) \times \mathbf{F}_a = \mathbf{N}', \quad (1.16)$$

³ The reader unfamiliar with standard vector operations may consult D.E. Rutherford, *Vector Methods*, Eighth Edition, Oliver and Boyd, Ltd., Edinburgh and London, 1954.

where $\mathbf{N}' = \sum_a (\mathbf{r}_a - \mathbf{R}) \times \mathbf{F}_a = \sum_a \mathbf{r}'_a \times \mathbf{F}_a$ is the torque measured about the origin in Σ' . Thus, the *form* $\dot{\mathbf{L}} = \mathbf{N}$ is valid in any inertial frame, even though the *values* of \mathbf{L} and \mathbf{N} depend on the frame.

The principle of angular momentum is also valid relative to the *center of mass* of the moving system (this is *not* necessarily an inertial frame: the center of mass can accelerate). To prove this statement, simply reinterpret the coordinate \mathbf{R} in (1.15) as the center of mass coordinate in Σ , thereby attaching the frame Σ' to the center of mass of the system. The primed quantities are then measured relative to the center of mass which moves with velocity \mathbf{V} . Now both $\sum_a m_a \mathbf{r}'_a$ and $\sum_a \mathbf{p}_a$ are zero. However, the last term in (1.15) survives since \mathbf{R} and \mathbf{V} are not necessarily parallel anymore. The time derivative of \mathbf{L}' in the center of mass system becomes

$$\begin{aligned} \dot{\mathbf{L}}' &= \dot{\mathbf{L}} - \mathbf{R} \times \sum_a m_a \dot{\mathbf{V}} = \sum_a \mathbf{r} \times \mathbf{F}_a^{(e)} - \mathbf{R} \times \sum_a \dot{\mathbf{p}}_a \\ &= \sum_a \mathbf{r}'_a \times \mathbf{F}_a^{(e)} = \mathbf{N}'_a, \end{aligned} \quad (1.17)$$

using (1.7) in reverse for $\dot{\mathbf{P}} = \sum_a \dot{\mathbf{p}}_a$. The torque, or moment, \mathbf{N}' of the external forces now refers to the center of mass. In applying the principle of angular momentum, we may take moments about a fixed point in space (in an inertial frame of course), or about the center of mass. The principle does not hold in a frame of reference performing an arbitrary motion in space.

If the external torque is removed, the total angular momentum \mathbf{L} stays constant. While superficially similar to the statement of constant linear momentum \mathbf{P} if $\mathbf{F} = 0$, the conservation of angular momentum has quite different consequences: If a system initially has zero total (linear) momentum, it stays zero throughout the subsequent motion. Internal forces cannot move the center of mass. If the angular momentum is initially zero, this is maintained throughout the subsequent motion. However, this does not say that the orientation of the system in space is maintained; the internal components can vary their relative orientation at will with the help of mutual interactions. A cat released from an inverted position makes use of internal muscular couplings between its posterior and anterior regions to drop feet first onto the floor! (See Prob. 1-2).

The principles of linear and angular momentum contain all the necessary information to study the motion of particle systems under given forces. We would only have to apply these principles to an endless variety of special situations in order to build up a complete scene of mechanics. To be quite specific, let us look at the motion of a single particle under a prescribed force \mathbf{F} . The equation of motion is given by (1.2). This single vector equation is short-hand for the three component equations

$$\dot{p}_x = F_x, \quad \dot{p}_y = F_y, \quad \dot{p}_z = F_z, \quad (1.18)$$

if we choose a rectangular reference frame O_{xyz} and identify the components of \mathbf{p} and \mathbf{F} along the axes by subscripts. For m a constant, $\dot{p}_x = m\dot{x}$, etc., where (x, y, z) are the coordinates of the particle at any time, and

$$m\ddot{x} = F_x, \quad m\ddot{y} = F_y, \quad m\ddot{z} = F_z. \quad (1.19)$$

These three second order differential equations in time may be integrated (numerically if need be) subject to known initial conditions of position and velocity to find the position $x(t), y(t), z(t)$ of the particle at all times. This process is called *solving for the motion*. Such *brute force* methods have their uses of course. However, we will show only passing interest in them since it turns out that often a great deal can be learned about the properties of a mechanical system *without* integrating the equations of motion directly. The latter point of view concerns itself in part with various invariance properties under coordinate transformations the system may possess and how these properties affect the dynamics. A formulation of mechanics that makes such invariance properties particularly transparent and moreover allows one to exploit them to advantage is to be found in the *principle of least action*⁴ or *Hamilton's principle*.

We base the developments in these lectures on such a point of view. One of the many advantages is a formulation of dynamics that discusses with equal ease, and in a uniform fashion, the dynamics of such widely distinct systems as the motion of a planet, the oscillations of electrons in a solid, or the pattern of ripples around an insect swimming on the surface of a pond. Furthermore, the tendency in modern physics has been to assign the action principle a fundamental role in the dynamics of systems that have no classical counterpart. Consequently, the methods that have evolved by treating classical systems from this point of view are useful in many unrelated fields of research in theoretical physics.

1-4 The Action Principle

Newton's laws of motion are a differential formulation of mechanics. They determine the change in momentum produced by forces acting on a system for a time dt . Other formulations of mechanics are available that concern themselves with the motion of a system throughout a finite time interval from t_1 to t_2 . Such formulations are called *integral principles* of mechanics. A particularly important principle of this type is the *principle of least action*, or *Hamilton's principle*. It simply states that a mechanical system moves between t_1 and t_2 in such a way that the integral

$$S = \int_{t_1}^{t_2} L dt \quad (1.20)$$

assumes a minimum⁵ value. The function L in this integral is called the *Lagrangian function* of the system, and depends on the coordinates

⁴ The name, principle of least action, is often restricted to refer to a principle enunciated by Maupertius in 1747, see Sec 1-8. However, we will use the phrase *generically* to refer to any of the several extremum principles that are available in mechanics.

⁵ More precisely, a stationary value S has a minimum value only if t_1 and t_2 are close together (see, for example, E.T. Whittaker, *A Treatise on the Analytical Dynamics of particles and rigid bodies*, p. 250, fourth edition, Cambridge University Press, London, digital printing 1999).

that describe the system as well as their time derivatives and on t . We will see later how L can be found when we derive this principle from Newton's laws.

The way the action principle has been stated above does not make clear how the value of S is to be judged or on what it depends. A simple example illustrates the basic idea. Consider the motion of a particle in one dimension. Its position $x = f(t)$ can be found at any time t by integrating Newton's second law of motion twice with respect to time. The value of the Lagrange function may thus be calculated as a function of t too, giving $L(x, \dot{x}) = L[f(t), \dot{f}(t)]$ from which we may calculate S . Obviously, S depends on the functional relation between x and t , given by $x = f(t)$ in this case. We say S is a *functional* of f and write

$$S = S[f] = \int_{t_1}^{t_2} L[f(t), \dot{f}(t)] dt.$$

Therefore S changes if the functional relation between x and t changes.

If we replace the function $f(t)$ by *another function* $\bar{f}(t)$ that evolves differently with time, and set $x = \bar{f}(t)$, we get a different value for $L = L[\bar{f}(t), \dot{\bar{f}}(t)]$ and consequently another value for $S = S[\bar{f}]$. The principle of least action then asserts that ⁶ $S[f] < S[\bar{f}]$ if $x = f(t)$ describes the actual motion. The relation $x = \bar{f}(t)$, that could hold between the particle position and elapsed time if additional forces were present, but does not under the actual forces present, is termed a *virtual motion*. We may go from the actual to the virtual motion by making a *virtual displacement* in the coordinates. This means adding an *arbitrary function* of time, $\delta x(t)$ to the actual position $x = f(t)$, so that $\bar{f}(t) = f(t) + \delta x(t)$ or

$$\delta x(t) = \bar{f}(t) - f(t). \quad (1.21)$$

The functions $\delta x(t)$ are usefully regarded as *small*, i.e. $\bar{f}(t)$ and $f(t)$ are not to differ much in form ⁷. In comparing the values of S in actual and virtual motions, we must be careful to let the virtual displacements $\delta x(t)$ vanish at both ends of the time interval under consideration,

$$\delta x(t_1) = \delta x(t_2) = 0, \quad (1.22)$$

that is, the actual and virtual motions must coincide at these two times. The necessity of such boundary conditions in time on the virtual displacements will be proven presently.

To get some feel for the action principle let us test it out on the following simple problem, before starting into a general proof. A particle falls from rest under gravity g . Its position is $x = f(t) = \frac{1}{2}gt^2$ after a time t has elapsed. This is the actual motion. We compare it to a virtual motion prescribed by

$$x = \bar{f}(t) = \frac{1}{2}gt_2^2 \left(\frac{t}{t_2}\right)^\beta, \quad \beta > 0,$$

⁶ Recall that strictly speaking, it may be an extremum.

⁷ More precisely, $\delta x(t)$ must be uniformly small, have a uniformly small first derivative, and satisfy (1.22).

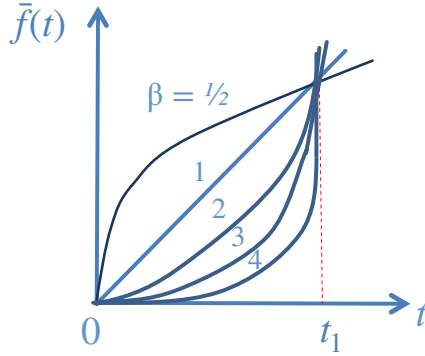


Figure 1.2: The family of curves $\bar{f}(t) = \frac{1}{2}gt_2^2(\frac{t}{t_2})^\beta$ for $\beta = \frac{1}{2}, 1, 2, 3, 4$.

as shown in Fig. 1.2.

The condition $\beta > 0$ ensures that $f(t)$ and $\bar{f}(t)$ agree at $t = t_1 = 0$ and $t = t_2$, so that $\delta x(t)$ vanishes at the endpoints of the motion. Since the Lagrange function for free fall under gravity is [see (1.34) and Prob. 1-3]

$$\begin{aligned} L &= \frac{1}{2}m\dot{x}^2 + mgx \\ &= m\left(\frac{1}{2}gt_2\right)^2\left[\frac{1}{2}\beta^2\left(\frac{t}{t_2}\right)^{2\beta-2} + 2\left(\frac{t}{t_2}\right)^\beta\right], \end{aligned}$$

we have

$$S[\bar{f}] = \int_0^{t_2} L dt = \frac{1}{8}mt_2(gt_2)^2\left[\frac{\beta^2}{2\beta-1} + \frac{4}{\beta+1}\right].$$

For various values of β , the functions $\bar{f}(t)$ form a family of curves that intersect the path of the actual motion at $t = t_1 = 0$ and $t = t_2$. We try various forms of \bar{f} by varying β . The resulting values of S are shown in Fig. 1.3. The minimum value of S at $\beta = 2$ confirms that $\bar{f} = \frac{1}{2}gt^2$ is indeed the actual motion. The other stationary point below $\beta = -1$ is excluded by the condition $\beta > 0$, i.e. by the boundary conditions on $\delta x(t)$.

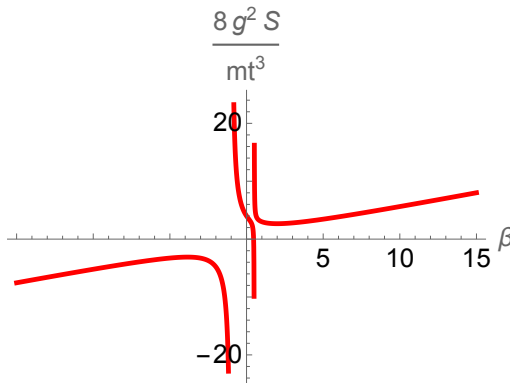


Figure 1.3: The function S (multiplied by the constant $8/mg^2t_2^3$) is shown as a function of the parameter β .

Now consider an arbitrary, interacting system of n particles, possibly also subject to external forces. We wish to derive the principle of least action for this system. To that end, introduce cartesian coordinates (x_1, x_2, x_3) for particle 1, (x_4, x_5, x_6) for particle 2, etc. and the force components (F_1, F_2, F_3) , acting on particle 1, (F_4, F_5, F_6) acting on particle 2, etc. This notation is not as clumsy as it looks, since we may write Newton's law of motion as

$$F_k - m_k \ddot{x}_k = 0 \quad (1.23)$$

for each particle in the system by simply letting $k = 1, 2, \dots, 3n$. The masses m_k are equal in triplets, $m_1 = m_2 = m_3, m_4 = m_5 = m_6$, etc., giving the masses of particles $1, 2, \dots, 3n$ and so forth.

Now subject each coordinate $x_k(t)$ to a virtual displacement, $x_k(t) \rightarrow x_k(t) + \delta x_k(t)$. By doing so, we force the system to perform a virtual motion described by the functions

$$x_k(t) + \delta x_k(t), \quad k = 1, 2, \dots, 3n. \quad (1.24)$$

Notice that the virtual displacements $\delta x_k(t)$ are *instantaneous*. They have nothing to do with the displacement $dx_k = \dot{x}_k dt$ occurring in time dt during the actual motion. Corresponding positions during the actual and virtual motions are tagged by the same value of t . However, the virtual displacements are certainly *functions* of time. Therefore, not only positions but also velocities change in a virtual displacement. The change in velocity $\delta \dot{x}$ is found from the difference

$$\delta \dot{x}_k = \frac{d}{dt}(x_k(t) + \delta x_k(t)) - \frac{d}{dt}x_k(t) = \frac{d}{dt}(\delta x_k). \quad (1.25)$$

Since $\delta \dot{x}_k = \delta(dx_k/dt)$, we see that the operations δ and d/dt actually *commute* for the virtual displacements we are considering. This fact will be important in what follows.

We now form the expression

$$\sum_k (F_k - m_k \ddot{x}_k) \delta x_k, \quad (1.26)$$

where the sum over k runs over all coordinates. This sum is actually *zero* by Newton's second law. Before using this fact let us perform a mathematical transformation on the acceleration terms by writing down the identity

$$\sum m_k \ddot{x}_k \delta x_k = \frac{d}{dt} \left(\sum_k m_k \dot{x}_k \delta x_k \right) - \sum_k m_k \dot{x}_k \frac{d}{dt} (\delta x_k).$$

We interchange d/dt and δ in the last term in accordance with our findings in (1.25) and get

$$\sum m_k \ddot{x}_k \delta x_k = \frac{d}{dt} \left(\sum_k m_k \dot{x}_k \delta x_k \right) - \delta \left(\sum_k \frac{1}{2} m_k \dot{x}_k^2 \right).$$

The sum

$$T = \sum_k \frac{1}{2} m_k \dot{x}_k^2 \quad (1.27)$$

is the total *kinetic energy* of the system and $\delta(\sum_k \frac{1}{2} m_k \dot{x}_k^2) = \sum_k m_k \dot{x}_k \delta \dot{x}_k$ its change, or *variation* in a virtual displacement. Our expression (1.26) thus reads

$$\sum_k F_k \delta x_k + \delta T - \frac{d}{dt} (\sum_k m_k \dot{x}_k \delta x_k) = 0 \quad (1.28)$$

if we use Newton's laws. We may also identify the first sum in this equation. The expression

$$\sum_k F_k \delta x_k = \delta W \quad (1.29)$$

is just the work done by the forces F_k in the virtual displacement. It is called the *virtual work*. Notice that any forces which do no work drop out at this stage. Such workless forces are present when a system is made to move in a definite way by *constraints* on its motion. We postpone discussion of such forces of constraint, except to note that they do not contribute to the virtual work δW . Forces which do contribute to δW are termed *applied forces*.

Returning to (1.28) we now integrate over time from t_1 to t_2 . Then

$$\int_{t_1}^{t_2} \delta W dt + \int_{t_1}^{t_2} \delta T dt - [\sum_k m_k \dot{x}_k \delta x_k]_{t_1}^{t_2} = 0,$$

and if we only employ virtual displacements that vanish at both ends of the time interval,

$$\delta x_k(t_1) = \delta x_k(t_2) = 0, \quad \text{for all } k, \quad (1.30)$$

then

$$\int_{t_1}^{t_2} \delta W dt + \int_{t_1}^{t_2} \delta T dt = 0 \quad (1.31)$$

is seen to be a consequence of Newton's laws of motion plus the boundary condition on the $\delta x_k(t)$.

We can transform this result somewhat further by observing that, since time is not varied in calculating δT , we can move δ outside the integral sign and write

$$\int_{t_1}^{t_2} \delta T dt = \delta \int_{t_1}^{t_2} T dt.$$

Similarly, we might be tempted to write

$$\int_{t_1}^{t_2} \delta W dt = \delta \int_{t_1}^{t_2} W dt.$$

However, this procedure is *ambiguous* since only δW is well-defined and not the total work W . In general W depends on the path followed by

the system between t_1 and t_2 . Only for systems in which the work is independent of the path can we use this transformation. Such systems are called *conservative*, and possess *potential energy* V equal to the negative of the work done in moving from t_1 to t_2 . They occur sufficiently often in physics to merit rewriting (1.31) specially for them as

$$\delta \int_{t_1}^{t_2} (T - V) dt = 0, \quad (1.32)$$

since

$$\int_{t_1}^{t_2} \delta W dt = - \int_{t_1}^{t_2} \delta V dt = -\delta \int_{t_1}^{t_2} V dt.$$

The structure of (1.32) suggests that the function

$$L = T - V \quad (1.33)$$

plays a fundamental role in dynamics. It does. This is the Lagrange function introduced at the beginning of this section. The condition (1.32)

$$\delta \int_{t_1}^{t_2} L dt = 0, \quad \text{where } L = T - V \quad (1.34)$$

asserts that the integral

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

is stationary for the actual motion of the system, subject to the boundary conditions (1.30). The integral S itself is called the *action*. The conditions (1.34), or $\delta S = 0$ for short, emphasize that not S itself but its *stationary character* is important, i.e. the value of S is unchanged to first order if the coordinates $x_k(t)$ describing the actual motion are replaced by coordinates $x_k(t) + \delta x_k(t)$.

The derivation we have presented gives the impression that the variational principle (1.34) only holds when $L = T - V$. This is not so. Equation (1.34) is true for the motion of any system for which a Lagrange function can be constructed, even if this is not equal to $T - V$. For completeness, we also make the obvious remark that the more general principle (1.31) is available whether a Lagrange function can be constructed or not.

The condition $\delta S = 0$ is called a *variational principle* for determining the functions $x_k(t)$. This principle is a compact formulation of Newtonian dynamics that emphasizes a remarkable fact: a *single* scalar function L completely determines the dynamics of a system. In fact, (1.34) embodies an entirely new point of view of mechanics. A. Sommerfeld summarizes its essence on p. 209 of his beautiful treatise on Mechanics⁸ *find the Lagrange function!* We here re-emphasize Sommerfeld's admonition by pointing out that (1.34) indicates that L is indeed the key that unlocks the door to the dynamics of classical (and presumably also

⁸ A. Sommerfeld, *Lectures on Theoretical Physics, Vol. I, Mechanics*, Academic Press Inc., New York, 1952.

quantum) systems. How unique is L ? The variational principle answers this question immediately. For example, the condition $\delta S = 0$ is unaffected by (i) adding a constant to L , (ii) multiplying L by a constant, (iii) adding the total time derivative of any function $f(x_1, x_2, x_3, \dots, t)$ of the coordinates and time to L . Changes (i) and (ii) merely affect the zero of energy and units used in a particular problem. To appreciate (iii) we write $L' = L + df/dt$ and find

$$\delta \int_{t_1}^{t_2} L' dt = \delta \int_{t_1}^{t_2} L dt + [\delta f]_{t_1}^{t_2},$$

where δf is the variation of f in a virtual displacement. This must vanish at t_1 and t_2 by the boundary conditions on $\delta x_k(t)$, so L' and L satisfy the same variational principle.

1-5 Lagrange's Equations

In the previous section, we showed that the variational principle $\delta S = 0$ is a consequence of Newton's equations of motion. Now we turn the question around and ask what equations of motion are implied by this principle. At first sight, this seems a silly thing to do because we should surely just get back to the equations of motion we started out from! In a sense this is true and in another sense it is not. The point is that, while we derived the condition $\delta S = 0$ from Newton's laws of motion expressed in cartesian coordinates, the result must be true in *any* set of coordinates, since S only depends on the scalar function L and hence is a scalar itself.

The advantage of starting out from the variational principle is then that we may pick any system of coordinates that we fancy. Let $\{q_1, q_2, \dots, q_n\}$ or q_k for short, be such a set of n *generalized* coordinates. We may pass to them via *point transformations* of the form

$$x_k = f_k(q_1, q_2, \dots, q_n, t) \quad (1.35)$$

for each x_k . The time derivatives of the x_k transform like $\dot{x}_k = \dot{f}_k$, or

$$\dot{x}_k = \sum_l \frac{\partial f_k}{\partial q_l} \dot{q}_l + \frac{\partial f_k}{\partial t}. \quad (1.36)$$

The last factor, $\partial f_k / \partial t$, is only present if the transformation is time dependent.

How does the Lagrange function respond to such a transformation? Since L is a scalar, its value is unchanged of course. However, its *functional* dependence on the coordinates and their time derivatives changes. For example, the Lagrange function given in (1.33) reads

$$L = \sum_k \frac{1}{2} m_k \dot{x}_k^2 - V(x_1, x_2, \dots) \quad (1.37)$$

in cartesian coordinates, and changes into

$$L = \sum_k \frac{1}{2} m_k \dot{f}_k^2 - V(f_1, f_2, \dots)$$

under the transformations (1.35) and (1.36). Thus, the potential energy becomes a function of the q_k , the kinetic energy (through \dot{f}_k) a function of the \dot{q}_k and the q_k . If the transformation is time-dependent, both T and V become a function of time also. We write

$$L = L(q_1, q_2, \dots; \dot{q}_1, \dot{q}_2, \dots, t), \quad \text{or} \quad L(q, \dot{q}, t)$$

for short, thereby completing our definition of the variational principle in (1.34): L is to be considered a function of q , \dot{q} and t . The time derivatives \dot{q}_k are called the *generalized velocities*.

The actual motion of the system is given by knowing the coordinates q_k as functions of the time: $q_k = q_k(t)$. It is the functions $q_k(t)$ that we now *seek*. The stationary character of S tells us how to find them. For if we subject the $q_k(t)$ to virtual displacements the value of S should not change to first order in these displacements. Changing $q_k(t)$ to $q_k(t) + \delta q_k(t)$ induces a corresponding change in $\dot{q}_k(t)$ to $\dot{q}_k(t) + \delta \dot{q}_k(t)$, so that L varies by an amount

$$\begin{aligned} \delta L &= L(q + \delta q, \dot{q} + \delta \dot{q}, t) - L(q, \dot{q}, t) \\ &= \sum_k \frac{\partial L}{\partial q_k} \delta q_k + \sum_k \frac{\partial L}{\partial \dot{q}_k} \delta \dot{q}_k. \end{aligned}$$

Consequently S changes by

$$\int_{t_1}^{t_2} \delta L dt = \int_{t_1}^{t_2} \sum_k \left[\frac{\partial L}{\partial q_k} \delta q_k + \frac{\partial L}{\partial \dot{q}_k} \delta \dot{q}_k \right] dt.$$

This the final form for the condition $\delta S = 0$. We can meet this condition *separately* for each q_k , *provided that these coordinates are independent*. For if this is the case, we can choose all virtual displacements to be zero barring the k th one, $\delta q_k(t)$. For it, we choose a function of time that not only vanishes at t_1 and t_2 but everywhere inbetween, except in a small region Δt around some time t . The sum drops away in the above equation, which now specializes to

$$\int_t^{t+\Delta t} \left[\frac{\partial L}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) \right] \delta q_k(t) dt = 0. \quad (1.38)$$

Since $\delta q_k(t)$ can be an arbitrary function of time in t , we must conclude that

$$\left[\frac{\partial L}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) \right] = 0. \quad (1.39)$$

By choosing each δq_k to be non-vanishing in turn, we find this equation holds for each independent coordinate. Thus, the action S is rendered stationary if each $q_k(t)$ satisfies (1.39). These are the celebrated

Lagrange equations of motion (1788) for systems possessing a Lagrange function.

The extension to systems not possessing a Lagrange function is straightforward. We merely adopt the general form of the variational principle given in (1.31). We have already noted that the kinetic energy T becomes a function of q and \dot{q} in generalized coordinates. Therefore, the contribution from δT to its time integral is

$$\int_{t_1}^{t_2} \delta T dt = \int_{t_1}^{t_2} \sum_k \left[\frac{\partial T}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_k} \right) \right] \delta q_k(t) dt, \quad (1.40)$$

using the properties and boundary conditions of the $\delta q_k(t)$ once more. We have no information on δW in terms of the virtual displacements δq_k , so we simply write

$$\delta W = \sum_k Q_k \delta q_k, \quad (1.41)$$

thereby *defining* new force components Q_k . The Q_k are called *generalized forces*. We add the time integral of δW to (1.40) and set

$$\int_{t_1}^{t_2} \sum_k \left[Q_k + \frac{\partial T}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_k} \right) \right] \delta q_k(t) dt = 0 \quad (1.42)$$

to find the actual motion.

As before, we can meet this condition separately for each q_k , provided they are independent, by putting

$$Q_k + \frac{\partial T}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_k} \right) = 0 \quad (1.43)$$

for all k . These are often called *generalized Lagrange equations*.

The forces Q_k introduced above can be determined by inspection. We write δW in its alternative forms

$$\sum_k F_k \delta x_k \quad \text{and} \quad \sum_k Q_k \delta q_k$$

and eliminate the virtual displacements δx_k in favor of the δq_k . The transformation is

$$\delta x_k = \sum_l \frac{\partial f_k}{\partial q_l} \delta q_l$$

from (1.44) (notice there is no term $(\partial f_k / \partial t) \delta t$; the time does *not* vary!). The two expressions for δW are identical if

$$Q_k = \sum_l F_l \frac{\partial f_l}{\partial q_k}. \quad (1.44)$$

Finally, if some of the forces contributing to δW are derivable from a potential V ,

$$\delta W = \sum_k Q_k \delta q_k - \delta V. \quad (1.45)$$

The variational principle (1.42) is modified accordingly to read

$$\int_{t_1}^{t_2} \sum_k \left[Q_k + \frac{\partial L}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) \right] \delta q_k(t) dt = 0,$$

implying that

$$Q_k + \frac{\partial L}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) = 0 \quad (1.46)$$

for all k . Equation (1.46) is a *hybrid* Lagrange equation, where some of the forces have been included through V in a Lagrange function $L = T - V$. On comparing their contribution to δW in (1.42), we have

$$\sum_k Q'_k \delta q_k = -\delta V = -\sum_k \frac{\partial V}{\partial q_k} \delta q_k$$

or that

$$Q'_k = -\frac{\partial V}{\partial q_k}. \quad (1.47)$$

A word on the dimensions of the q_k and the Q_k that have been introduced: The q_k need not have dimensions of length, but may individually assume any convenient dimension. Hence Q_k does not always have the dimensions of force. In fact, its dimension is determined by the q_k it associates with, since each term $Q_k \delta q_k$ in the sum (1.41) must have the dimensions of work, that is, energy.

The generality of the results found in this section cannot be overemphasized. The Lagrange equations we derived are true for *any* set of generalized coordinates. There is one very important proviso: the q_k must all be independent of each other. Otherwise we cannot make the essential step from the variational principle to the Lagrange equations it implies for each coordinate. How then should the q_k be chosen; how many do we need? These questions are closely related to the number of coordinates a system requires and the degrees of freedom that it has. This relationship is discussed in the next section.

1-6 Constraints

A system has as many *degrees of freedom* as there are independent coordinates required to specify its instantaneous state completely. For example, a system of N particles moving arbitrarily in space has $3N$ degrees of freedom, since three cartesian coordinates are required to position each particle in space. However, if a system moves subject to constraints, the number of degrees of freedom is reduced. A *constraint* is simply some additional condition that is imposed on the way a system shall move. A wheel rolling without slipping, a ride on a roller coaster, or simply that N atoms move together as a rigid body, are examples of motions performed under constraints. No slipping means the point on the rim of the wheel in contact with the surface is momentarily at rest;

the roller coaster forces the rider to follow a particular path in space; the relative separation between atoms must stay constant during the motion of a rigid body. Forces are needed to enforce these constraints. Friction causes the wheel not to slip, guide rails (exerting reactions) define the path of the roller coaster, while electrical forces between atoms give the rigid body its rigidity. Such forces are called *forces of constraint*. As the above examples show, we may always replace them by their effects on the motion of a system in the form of constraints. One should realize, however, that considering the constraints instead of the forces that enforce them is an idealization of Nature. No rigid body is perfectly rigid. The electrical forces between atoms can never exactly "freeze" them into permanent positions in a solid. When a solid is subjected to stresses, the relative positions of its constituent atoms *do* change (as when a sound wave passes through it). If such changes are unimportant for our purposes, we idealize the situation by considering the solid to be *rigid*. In this way we circumvent the enormous, but for us irrelevant problem of the internal motion of the constituent atoms making up a rigid body.

Forces of constraint have one very important property in common: they do no work in a virtual displacement that is consistent with the constraints on the system. A general proof of this assertion is difficult⁹, but our three examples of constrained motion illustrate the workless nature of these forces: the wheel does not "scratch" the surface it is rolling on, the guide rail reaction is perpendicular to the direction of motion of the roller coaster, the internal forces do no work when a rigid body is displaced as a whole. In this connection the role of frictional forces should be noted with care. *Static* friction, as for the rolling wheel, is a force of constraint; it makes rolling possible. However, *sliding* or *kinetic* friction comes into play if the wheel slips on the surface. The "rolling" constraint falls away and friction becomes an applied force. It does work during a virtual displacement of the system and thus contributes to δW .

Quite generally, then, the distinguishing feature about forces of constraint is that they have to drop out of the variational principle because they do no work. We do not need to know them to calculate δW . However, their effects must show up somewhere, and they do, in that the number of independent coordinates are reduced through the constraints imposed by these forces. Thus, the $3N$ cartesian coordinates x_k for an N -particle system are not all independent if the system has constraints on its motion: r constraint conditions leave it with only $3N - r$ degrees of freedom and as many independent coordinates. The Lagrange equations of the previous section *will therefore not hold for all the x_k* unless the constraining forces are explicitly included among all the forces. There is a simple way around this difficulty: express the x_k in terms of a smaller number of coordinates that *are* independent by introducing $3N$ transfor-

⁹ Lagrange attempts one in his *Mechanique Analytique, Tome premier, Quatrieme edition, Librairie Scientifique et Technique, A. Blanchard, Paris, 1965.*

mations

$$x_k = f_k(\underbrace{q_1, q_2, \dots}_{3N-r \text{ coordinates}}, t) \tag{1.48}$$

to new coordinates q_k . We might well ask: can these transformations to a smaller number of independent coordinates always be constructed? The answer depends on the type of constraint involved. If the constraints are conditions among the coordinates themselves that do not involve their time derivatives, such transformations are always possible. Constraints of this type are termed *holonomic* (from the Greek word *holos*, or whole) and can be displayed in the general form¹⁰

$$F_\alpha(x_1, x_2, \dots, t) = 0 \tag{1.49}$$

where the label $\alpha = 1, 2, \dots, r$ distinguishes between r different functions F . We are thus provided with r relations between $3N$ coordinates x_k , so only $3N - r$ of them are independent. If we now choose the $3N$ functions $f_k(q_1, q_2, \dots, t)$ in (1.48) such that the r relations (1.49) hold as *identities*, i.e. $F(f_1, f_2, \dots, t) = 0$, the reduction of the number of coordinates in the problem is complete. The Lagrange function can be expressed in terms of the q_k and the \dot{q}_k , and satisfies

$$\frac{\partial L}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) = 0$$

in these coordinates. We have succeeded in eliminating the constraints entirely by means of a point transformation.

¹⁰ In the literature these constraints are further classified as *rheonomous* (fluid) or *scleronomous* (fixed) depending on whether they contain the time or not.

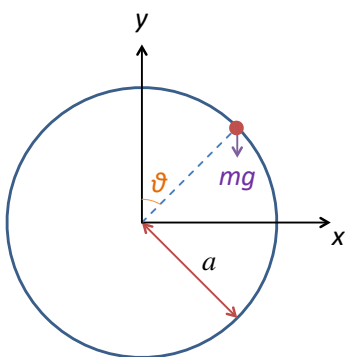


Figure 1.4: A particle of mass m constrained to move in a vertical circular path in the gravitational field of the earth.

A simple example illustrates these ideas, see Fig. 1.4. A particle of mass m moves in a circle in a vertical plane. Find its equations of motion if gravity is the only applied force. The motion has one degree of freedom, since the cartesian coordinates (x, y) of the particle must satisfy

$$F(x, y) = x^2 + y^2 - a^2 = 0$$

choosing the center of the circle of radius a as the origin. Taking the y axis vertical, the potential energy of the particle in any position is gy per unit mass, so the Lagrange function is

$$L = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) - mgy.$$

The Lagrange equations

$$\begin{aligned}\frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) &= -m\ddot{x} = 0; \\ \frac{\partial L}{\partial y} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right) &= -mg - m\ddot{y} = 0\end{aligned}$$

are *incorrect*; x and y are not independent. However, if we first eliminate x and \dot{x} by the relations

$$x = \sqrt{a^2 - y^2}; \quad \dot{x} = -\frac{y\dot{y}}{\sqrt{a^2 - y^2}} \quad (1.50)$$

following from the constraint condition, then

$$L = \frac{1}{2}m\frac{a^2}{a^2 - y^2}\dot{y}^2 - mgy$$

and

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

that is

$$-g\left(1 - \frac{y^2}{a^2}\right) - \frac{y\dot{y}^2}{a^2 - y^2} - \ddot{y} = 0 \quad (1.51)$$

gives the correct equation of motion for y . The elimination of x and \dot{x} amounts to choosing a new variable $q = y$ in (1.48), where

$$x = f_1(y) = \sqrt{a^2 - y^2}; \quad y = f_2(y) = y$$

so that f_1 and f_2 automatically satisfy the constraint condition $F(x, y) = 0$.

We have purposely stuck to the rather clumsy choice of $q = y$ as our independent variable to emphasize that it is not necessary to introduce the "obvious" angular variable θ giving the angular position of the particle from the vertical. But it is much more convenient to do so. The choice $q = \theta$, where

$$x = f_1(\theta) = a \sin \theta; \quad y = f_2(\theta) = a \cos \theta$$

gives

$$L = \frac{1}{2}m(a\dot{\theta})^2 - mga \cos \theta,$$

from which the much simpler equation of motion in θ ,

$$\frac{\partial L}{\partial \theta} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) = ma(g \sin \theta - a\ddot{\theta}) = 0$$

follows.

Constraints that do not fall into the above category are termed *non-holonomic*. Their elimination depends specifically on the form they take. Non-holonomic constraints of special interest in mechanics are of the differential type

$$\sum_k F_{\alpha k} dx_k + G_\alpha dt = 0, \quad \alpha = 1, 2, \dots, r, \quad (1.52)$$

where the $3n \times r$ functions $F_{\alpha k}$ and the r functions G_α can only depend on the x_k and the time. The trouble with such constraints is that we cannot eliminate them from the problem as we could holonomic constraints. Introducing new coordinates q_k by point transformations like (1.35) or (1.48) merely turns (1.52) into similar differential constraints on the q_k :

$$\sum_k F_{\alpha k} dx_k + G_\alpha dt = \sum_k F'_{\alpha k} dx_k + G'_\alpha dt, \quad (1.53)$$

where the $F_{\alpha k}$ and G_α in the second expression are new functions

$$F'_{\alpha k} = \sum_l F_{\alpha l} \frac{\partial f_l}{\partial q_k}; \quad G'_\alpha = G_\alpha + \sum_k F_{\alpha k} \frac{\partial f_k}{\partial t}$$

that depend on the q_k and the time. To see what constraints among the coordinates themselves are implied by such differential relations, we would first have to integrate the latter with respect to time. But because differentials of the coordinates as well as the coordinates themselves appear, we cannot do this until the coordinates are known functions of time, i.e. until the dynamical problem has been solved.

Lagrange introduced a method for dealing with non-holonomic constraints like (1.52) or (1.53). The method depends on the stipulation that in using the variational principle we must only allow virtual displacements that are consistent with the constraints on the system. Suppose that r conditions of the type (1.53) constrain the motion of a system that is described by n generalized coordinates q_k . Since the time is not varied during a virtual displacement, this means that the displacements in the q_k must obey (1.53) in the form (we drop the primes on the $F'_{\alpha k}$ and G'_α from now on)

$$\sum_k F_{\alpha k} \delta q_k = 0; \quad \delta t = 0. \quad (1.54)$$

The δq_k are not independent any more; equations (1.48) provide r subsidiary conditions they must satisfy. This is where the effects of constraints come in. We can no longer regard the δq_k as independent, and the form of Lagrange equations derived in Sec. 1-5 will not follow from the action principle. The method of *Lagrange multipliers* circumvents this difficulty: multiply the r relations (1.54) by arbitrary functions λ_α (one for each relation) and add. The result is still zero:

$$\sum_\alpha \lambda_\alpha F_{\alpha k} \delta q_k = 0. \quad (1.55)$$

We add this sum to the integrand of (1.47), which then reads

$$\int_{t_1}^{t_2} \sum_k \left[\sum_{\alpha} \lambda_{\alpha} F_{\alpha k} + \frac{\partial L}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) \right] \delta q_k dt = 0. \quad (1.56)$$

The r relations in (1.54) leave us only $n - r$ independent virtual displacements to choose at will. Call these δq_k with $k = 1, 2, \dots, n - r$. The remaining virtual displacements δq_k for $k = n - r + 1, n - r + 2, \dots, n$ are dependent on the values of the first $n - r$ displacements. We eliminate the dependent displacements from consideration in (1.56) by forcing their cofactors to be zero, that is, by choosing

$$\sum_{\alpha} \lambda_{\alpha} F_{\alpha k} + \frac{\partial L}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) = 0$$

for $k = n - r + 1, n - r + 2, \dots, n$. These r equations are to be regarded as conditions on the functions λ_{α} . If these conditions are met, the sum in (1.56) will only contain terms involving the independent virtual displacements δq_k for $k \leq n - r$, and may be written as

$$\int_{t_1}^{t_2} \sum_{k=1}^{n-r} \left[\sum_{\alpha} \lambda_{\alpha} F_{\alpha k} + \frac{\partial L}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) \right] \delta q_k(t) dt = 0. \quad (1.57)$$

This condition is satisfied by letting each q_k satisfy

$$\sum_{\alpha} \lambda_{\alpha} F_{\alpha k} + \frac{\partial L}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) = 0 \quad (1.58)$$

for $k = 1, 2, \dots, n - r$. Both sets of equations (1.56) and (1.58) are identical in form but very different in origin. Together they provide n equations for the n unknown coordinates q_k and r unknown functions λ_{α} . Therefore, we need further r equations to make the problem determinate. These are supplied by (1.53). Notice that we must employ the constraint conditions as they unfold with time, since we are interested in the actual motion. The virtual displacements have served their purpose; they are out of the picture now. Equations (1.58) with $k = 1, 2, \dots, n$ are the Lagrange equations in the presence of non-holonomic constraints (1.53). The same procedure adds the sum $\sum_{\alpha} \lambda_{\alpha} F_{\alpha k}$ to the two other versions of the Lagrange equations that we derived in Sec. 1-5.

The functions λ_{α} are called undetermined or Lagrange multipliers. They are introduced as a device to take care of the constraints on the system while the $n - r$ virtual displacements in (1.57) are allowed to vary independently. What is their physical significance? Referring to (1.58), we see that the motion is unchanged if the constraints are removed, but additional forces

$$Q_k'' = \sum_{\alpha} \lambda_{\alpha} F_{\alpha k} \quad (1.59)$$

are introduced. The forces Q_k'' are not included in the Lagrange function; they must therefore represent the forces of constraint that enforce the

constraints imposed by (1.53). If so, the Q_k'' should do no work in a virtual displacement. They don't. Their contribution to the virtual work is

$$\sum_k Q_k'' \delta q_k = \sum_{\alpha k} \lambda_\alpha F_{\alpha k} \delta q_k$$

which is zero by (1.55). However, *time-dependent* constraints do work on the system in time dt during the actual motion. A proof of this statement and its physical origin waits in the next section.

One remark remains to be made: holonomic constraints are a special kind of constraint. For if we calculate the change of the holonomic constraints (1.49) with time, there results

$$dF_\alpha = \sum_k \frac{\partial F_\alpha}{\partial x_k} dx_k + \frac{\partial F_\alpha}{\partial t} dt = 0 \quad (1.60)$$

which is a special case of (1.52) with $F_{\alpha k} = \frac{\partial F_\alpha}{\partial x_k}$ and $G = \frac{\partial F_\alpha}{\partial t}$. Thus, holonomic constraints can also be treated with Lagrange multipliers; we do not *have to* eliminate them. In fact, if the forces of constraint are sought we *must not* eliminate them. Once the functions λ_α are determined, these forces are given by

$$Q_k'' = \sum_\alpha \lambda_\alpha \frac{\partial F_\alpha}{\partial x_k}. \quad (1.61)$$

The example of a particle moving in a circle under gravity that we discussed is a case in point. If we want to calculate the forces of constraint necessary to maintain the circular path we must keep both coordinates x and y in the Lagrange function

$$L = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) - mgy$$

but restrict the virtual displacements in x and y . For circular motion this restriction is

$$\lambda(x\delta x + y\delta y) = 0$$

after multiplying by the single Lagrange multiplier λ required to enforce it. The two modified Lagrange equations

$$\lambda x + \frac{\partial L}{\partial x} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) = \lambda x - m\ddot{x} = 0 \quad (1.62)$$

$$\lambda y + \frac{\partial L}{\partial y} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{y}} \right) = \lambda y - mg - m\ddot{y} = 0 \quad (1.63)$$

plus the constraint condition

$$x\dot{x} + y\dot{y} = 0 \quad (1.64)$$

are three equations that determine x , y and λ according to the general theory of this section. Multiply the Lagrange equations by x and y respectively and add to determine λ . We get

$$\lambda a^2 = mgy + m(x\ddot{x} + y\ddot{y}) = mgy - m(\dot{x}^2 + \dot{y}^2) \quad (1.65)$$

after using the condition $x^2 + y^2 = a^2$ and its second time derivative to obtain the result on the right. Since $v^2 = \dot{x}^2 + \dot{y}^2$ is just the velocity squared of the particle, λ is finally given by

$$\lambda a = mg \frac{y}{a} - \frac{mv^2}{a} = R.$$

Now (1.59) identifies the components of the constraint force along x and y as

$$Q_x'' = \lambda x = \left(mg \frac{y}{a} - \frac{mv^2}{a} \right) \frac{x}{a} = R \frac{x}{a} \quad (1.66)$$

$$Q_y'' = \lambda y = \left(mg \frac{y}{a} - \frac{mv^2}{a} \right) \frac{y}{a} = R \frac{y}{a} \quad (1.67)$$

of magnitude $R = \lambda a$. Since Q_x'' and Q_y'' are in the same ratio as x to y the direction of R is along the radius vector to the particle from the center of the circle. These relations are also shown in Fig. 1.5.

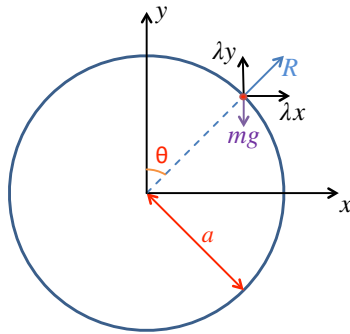


Figure 1.5: The particle of mass m constrained to a vertical circular path and acted on by gravity.

To find the motion of the particle, we either proceed as before and eliminate the holonomic constraint, or we can eliminate λ directly from the equation of motion in x and y : Taking the latter course of action, we multiply (1.62) by x , (1.63) by y , subtract the former equation from the latter and express the derivatives as a total derivative. The result

$$\frac{d}{dt} [m(xy\dot{y} - \dot{x}y)] = -mgx$$

is equivalent to the principle of angular momentum (1.14) applied to the present problem. Now eliminating x and \dot{x} by using the constraint equations (1.50) gets us back to the equation of motion

$$\ddot{y} + \frac{yy\dot{y}^2}{a^2 - y^2} + g\left(1 - \frac{y^2}{a^2}\right) = 0$$

found previously for y and given in (1.51).

But working in x and y coordinates is just a difficult way to solve a simple problem. We can always go to any other coordinates that

are more convenient. So, if the polar coordinates (r, θ) are introduced instead, where

$$x = f_1(r, \theta) = r \sin \theta, \quad y = f_2(r, \theta) = r \cos \theta$$

things become much simpler. The Lagrange function transforms into

$$L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - mgr \cos \theta$$

and the constraint condition into

$$r - a = 0$$

by direct substitution. Again, we must not eliminate the holonomic constraint $r = a$ from L if the force of constraint is desired. Instead, we introduce a multiplier λ' to take care of the constraint on the variation of r ,

$$\lambda' \delta r = 0.$$

Then,

$$\lambda' + \frac{\partial L}{\partial r} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{r}} \right) = \lambda' - mg \cos \theta + mr\dot{\theta}^2 - m\ddot{r} = 0 \quad (1.68)$$

$$\frac{\partial L}{\partial \theta} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}} \right) = mgr \sin \theta - \frac{d}{dt} (mr^2\dot{\theta}) = 0 \quad (1.69)$$

$$\dot{r} = 0 \quad (1.70)$$

are the equations determining r , θ and λ' . But since $\dot{r} = 0$, we get $\ddot{r} = 0$ by differentiation and $r = a$ by integration. The equation in r , (1.68), gives

$$\lambda' = mg \cos \theta - ma\dot{\theta}^2$$

which, together with (1.59), identifies the constraint force for the r th coordinate (i.e. in the radial direction) as

$$Q_r'' = R = mg \cos \theta - ma\dot{\theta}^2. \quad (1.71)$$

Since $a\dot{\theta} = v$ and $y = a \cos \theta$ this is the same result as obtained previously, cf. (1.66) and (1.67). The equation for θ does not depend on R (or λ) and is identical with the equation $ma\ddot{\theta} = mg \sin \theta$ found before.

Notice that the physical origin of R does not have to be specified in the problem we have just discussed, only its effect.

1-7 Properties of L

We return to systems that have a Lagrange function and introduce the quantity

$$p_k = \frac{\partial L}{\partial \dot{q}_k} \quad (1.72)$$

into the Lagrange equations (1.39). These equations then become

$$\dot{p}_k = \frac{\partial L}{\partial q_k}. \quad (1.73)$$

The p_k are called *canonical momenta* and the pair (q_k, p_k) are referred to as canonical variables¹¹.

¹¹ The reason for this terminology will become clear in Chap. 7.

The version (1.73) of the Lagrange equations is reminiscent of Newton's form of the equations of motion, to which it reduces in cartesian coordinates. In that case

$$\frac{\partial L}{\partial x_k} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_k} \right) = -\frac{\partial V}{\partial x_k} - m_k \ddot{x}_k = F_k - m_k \ddot{x}_k = 0,$$

using the Lagrange function given by (1.37) and the fact that $\sum F_k \delta x_k = -\delta V$ is the work done in a virtual displacement.

Equations (1.72) and (1.73) have further important consequences that relate to the question of conservation laws. We explore these next. As a system moves, the coordinates q_k and velocities \dot{q}_k change with time. However, it often happens that one or more functions of these quantities stay constant in time during motion. Such functions are said to be conserved, or to obey conservation laws. It is of utmost importance to discover such conservation laws because they can often provide a complete picture of how a system moves. A powerful and systematic way of doing so involves studying the invariance properties of the Lagrange function under coordinate and time translations. For suppose L does not depend on a particular coordinate q_s . Then, changing q_s infinitesimally to $q_s + \Delta q_s$ cannot induce any change in L , so that

$$L(q_s + \Delta q_s, \dot{q}_s, t) - L(q_s, \dot{q}_s, t) = \frac{\partial L}{\partial q_s} \Delta q_s = 0,$$

indicating that

$$\frac{\partial L}{\partial q_s} = 0. \quad (1.74)$$

Therefore, the Lagrange equation for the coordinate q_s reads

$$\dot{p}_s = 0, \quad p_s = \frac{\partial L}{\partial \dot{q}_s} = \text{constant}. \quad (1.75)$$

The canonical momentum associated with a missing coordinate q_s is conserved. Observe carefully that this statement *is only true if q_s is an independent coordinate*, i.e. if all constraints have been eliminated from L . Otherwise, constraint forces enter (1.75) and spoil the conservation of p_s . The canonical momentum that is conserved may either be a linear or an angular momentum component, or some more complicated entity. The dimension of p_s depends on the dimension of q_s . Most commonly q_s is either a distance or an angle. Then, p_s is either a linear or an angular momentum. Notice in passing that the replacement $q_s \rightarrow q_s + \Delta q_s$ to

probe L for its conserved momenta is *not* a virtual displacement; the velocity \dot{q}_s is not allowed to change during this probe. Coordinates like q_s that do not appear in L are called *cyclic* or *ignorable*.

Another important conservation law is connected with the time-dependence of L . During the motion of a system from time t to $t + dt$ the value of L varies on two accounts: the change wrought by changes in q_k and \dot{q}_k , and the change due to an explicit dependence on time. In symbols,

$$dL = \sum_k \frac{\partial L}{\partial q_k} dq_k + \sum_k \frac{\partial L}{\partial \dot{q}_k} d\dot{q}_k + \frac{\partial L}{\partial t} dt,$$

where the last term contributes only if L depends explicitly on time. This equation is a mathematical statement for dL . We now put in the dynamics by inserting the value of $\partial L/\partial q_k$ from the Lagrange equation (1.73), and find the relation

$$d\left(\sum_k p_k \dot{q}_k\right) = dL - dt \frac{\partial L}{\partial t} \quad (1.76)$$

for the total change dL in time dt . Equivalently,

$$dH = -dt \frac{\partial L}{\partial t}, \quad (1.77)$$

where H is the function

$$H = \sum_k p_k \dot{q}_k - L. \quad (1.78)$$

If $\partial L/\partial t = 0$, i.e. if L does not contain the time explicitly, (1.77) shows that H is conserved. In that case H is called Jacobi's first integral of motion (later on H , whether conserved or not, will have another, more illustrious name).

The two relations, (1.77) and (1.78), have a familiar physical content if: (i) L is given by $T - V$ and does not depend explicitly on time, (ii) the kinetic energy T is a *homogeneous* quadratic function of the \dot{q} , and (iii) the potential V is only a function of the q . Then

$$\sum_k p_k \dot{q}_k = \sum_k \frac{\partial L}{\partial \dot{q}_k} \dot{q}_k = \sum_k \frac{\partial T}{\partial \dot{q}_k} \dot{q}_k = 2T \quad (1.79)$$

by Euler's theorem on homogeneous functions¹², so that (1.77) reduces to

$$dT + dV = 0, \quad \text{or} \quad dT = -dV = dW. \quad (1.80)$$

This is just the equation of differential energy balance in time dt : the change in kinetic energy equals the work done $-dV = dW$ by the applied forces. Equation (1.78) is the time-integrated version of this statement, or the conservation of the total energy E ,

$$H = 2T - (T - V) = T + V = E = \text{constant}. \quad (1.81)$$

¹² See for example R.P. Gillespie, *Partial Differentiation*, Oliver and Boyd, Edinburgh and London, 1951, p. 28.

We can thus establish whether the laws of momentum and energy conservation hold for any given system by simply examining the structure of its Lagrange function! This fundamental feature of L will be used repeatedly in our future discussions.

Two additional points deserve mention. If L is independent of time, V must also be. Otherwise, an additional time derivative shows up on the right hand side of (1.77) and upsets the conservation of H :

$$dH = dt \frac{\partial V}{\partial t}.$$

The other point concerns V being independent of time, but depending on the \dot{q}_k in addition to the q_k . Such a dependence replaces $\sum_k p_k \dot{q}_k$ in (1.79) by

$$\sum_k p_k \dot{q}_k = 2T - \sum_k \frac{\partial V}{\partial \dot{q}_k} \dot{q}_k$$

so that, while

$$H = (T + V) - \sum_k \frac{\partial V}{\partial \dot{q}_k} \dot{q}_k \quad (1.82)$$

is still conserved, it does not equal the total energy $T + V$ anymore (which is *not* conserved!). The conservation of H and conservation of energy are two separate issues. Physically, the important quantity is H , since like L it plays a fundamental role in dynamics, too. We explore this facet of H in Chap. 7. The results we have just found also hold true, with one important qualification, when the system is subject to constraints. The qualification relates to *time-dependent* constraints. The effect of constraints is to inject the additional forces of constraint $Q'_k = \sum_\alpha \lambda_\alpha F_{\alpha k}$ into the relation (1.76) for dL . That relation is changed to read

$$d\left(\sum_k p_k \dot{q}_k\right) - \sum_{\alpha,k} \lambda_\alpha F_{\alpha k} dq_k = dL - dt \frac{\partial L}{\partial t}.$$

The second term on the left would be zero in a virtual displacement. However, dq_k is the actual displacement of q_k in time dt . Therefore, $\sum F_{\alpha k} dq_k = -G_\alpha dt$ from the constraint equations (1.53), and

$$dH + dt \sum_\alpha \lambda_\alpha G_\alpha = -dt \frac{\partial L}{\partial t} \quad (1.83)$$

replaces (1.77). If now $H = T + V$ and L is independent of time, we get

$$dT = dW + dW', \quad dW' = -dt \sum_\alpha \lambda_\alpha G_\alpha \quad (1.84)$$

in place of the usual law $dT = dW$ or differential energy balance. Time-dependent constraints do work on the system! The work dW of applied forces only accounts for part of the change in kinetic energy. The rest is supplied by the forces of constraint working at a rate $dW'/dt = -\sum_\alpha \lambda_\alpha G_\alpha$.

We can visualize the physical origin of this additional energy by means of our particle moving in a circle under gravity again. Imagine the particle is a bead sliding on a hoop and rotate the hoop with a pre-assigned angular velocity ω about the vertical y axis. The motion then takes place in three dimensions and we must supply the azimuthal angle ϕ to locate the bead in spherical polar coordinates (r, θ, ϕ) , see Fig. 1.6.

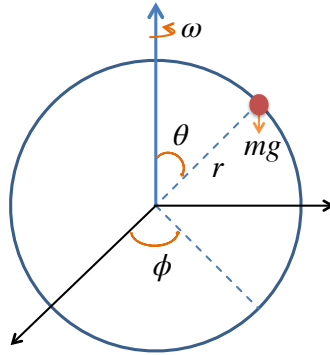


Figure 1.6: Bead on a rotating hoop.

The bead now has an additional velocity component $r \sin \theta \dot{\phi}$ perpendicular to the plane of the hoop, so the Lagrange function reads

$$L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2 \theta \dot{\phi}^2) - mgr \cos \theta$$

subject to the two constraints

$$r - a = 0, \quad \phi - \omega t = 0,$$

ensuring that the bead stays on the hoop, and turns with it. To illustrate the work done by the second constraint that is time-dependent, we must not eliminate it. The first one, $r = a$, may go however, so

$$L = \frac{1}{2}ma^2(\dot{\theta}^2 + \dot{\phi}^2 \sin^2 \theta) - mga \cos \theta \quad (1.85)$$

is the Lagrange function for a system with two coordinates θ and ϕ , subject to the one time-dependent constraint

$$d\phi - \omega dt = 0,$$

in order to identify with (1.53). The constraint

$$\lambda \delta\phi = 0$$

on virtual displacements of ϕ is taken care of by a single multiplier λ in the equation of motion for ϕ , which then determines λ . The work done by the time-dependent constraint is

$$-dt \sum_{\alpha} \lambda_{\alpha} G_{\alpha} = \omega \lambda dt$$

according to (1.84). But we also know the hoop forces the bead to rotate with it by pushing the bead with a reaction $Q''_{\phi} = \lambda$. However, since ϕ is an angle, Q''_{ϕ} will be a *moment* of the reaction R , so

$$Q''_{\phi} = \lambda = Ra \sin \theta$$

and the work done in time dt is

$$dt \omega Ra \sin \theta = R(a \sin \theta d\phi),$$

which is just the work done by R in pushing the bead an angle $d\phi = dt\omega$ in the horizontal plane. The machine (or person!) cranking the hoop around has to supply this energy. Therefore, the energy balance for the motion of the bead is

$$dT = mga \sin \theta d\theta + Ra \sin \theta d\phi,$$

where $dW = mga \sin \theta d\theta$ is the work done by gravity.

But suppose we also eliminated the second constraint $\phi - \omega t$ from L so that it only contains the independent variable θ :

$$L \rightarrow L' = \frac{1}{2}ma^2(\dot{\theta}^2 + \omega^2 \sin^2 \theta) - mga \cos \theta.$$

How does this Lagrange function *know* that it describes a system which does not convert work into kinetic energy according to the equation $dT = dW$? We notice that condition (ii) from (1.79) to hold is violated: the kinetic energy in L' is not a homogeneous function in $\dot{\theta}$ and $\dot{\phi}$ as was L . Therefore,

$$H' = \frac{\partial L}{\partial \dot{\theta}} \dot{\theta} - L' = \frac{1}{2}ma^2(\dot{\theta}^2 - \omega^2 \sin^2 \theta) + mga \cos \theta$$

is *not* the total energy; the minus sign on the $\sin^2 \theta$ prevents this from happening. But H' is conserved. The equation $dH' = 0$ still holds since L' is independent of time and shows that the kinetic energy $T' = \frac{1}{2}ma^2(\dot{\theta}^2 + \omega^2 \sin^2 \theta)$ obeys the relation

$$dT = mga \sin \theta d\theta + dt \frac{d}{dt}(ma^2 \omega^2 \sin^2 \theta).$$

But, from the equation of motion for ϕ as given by L , we have

$$\lambda + \frac{\partial L}{\partial \phi} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\phi}} \right) = \lambda - \frac{d}{dt}(ma^2 \omega \sin^2 \theta) = 0,$$

which determines λ . Thus dT is given by the same expression as before, since the second term indeed equals $dt\omega\lambda$, the work done by the time-dependent constraint during the actual motion.

We have also remarked that the presence of constraints (time-dependent or not) can upset the conservation of momentum laws. For instance,

since L in (1.85) is independent of ϕ , we might be tempted to conclude that

$$p_\phi = \frac{\partial L}{\partial \dot{\phi}} = ma^2 \sin^2 \theta \dot{\phi}$$

is a constant of motion. This would be wrong: p_ϕ actually changes at a rate $\dot{p}_\phi = \lambda$ according to the equation we found above for λ . The forces of constraint regulate how p_ϕ varies with time. The error in our argument is again the fact that ϕ is not an independent variable for the problem. The structure of the Lagrange function L' from which we have eliminated all constraints shows that the only constant of motion for the problem is H' .

Our example illustrates a general property. If a system with a time-independent Lagrange function L is subjected to time-dependent, holonomic constraints $F_\alpha = 0$, then, from (1.83)

$$dH + dt \sum \lambda_\alpha \frac{\partial F_\alpha}{\partial t} = 0 \quad (1.86)$$

holds if these constraints are not eliminated, and from (1.87)

$$dH' + dt \frac{\partial L'}{\partial t} = 0 \quad (1.87)$$

holds if they are (we use primes as a reminder that H and L are now functions of independent coordinates). The effect of the constraints in the latter equation is made up for partly in the functional dependence of H' on the new coordinates and partly by the time-dependence induced in L' . The difference between energy conservation and the conservation of H is also brought out by (1.86) and (1.87) if L' happens to stay independent of time when the constraints are eliminated, i.e. if $\partial L'/\partial t = 0$. If H is the total energy of the system, this changes at a rate $-\sum_\alpha \lambda_\alpha \partial F_\alpha / \partial t$ according to (1.86). But $dH' = 0$ if $\partial L'/\partial t = 0$ or H' stays constant during motion.

1-8 Other Variational Principles

The principle of least action that is discussed in Sec. 1-4. is but one of a whole host of variational formulations that are available in mechanics, that are grouped under the general title of *least action principles*. The emphasis here on *least* is historical rather than factual. All such principles only require that the action in question attains a *stationary* value for the actual motion. No statement about the value of the action itself is involved. We have seen one example of this in Sec. 1-4. We will see another one in the principle of Maupertuis that we are about to derive.

Maupertuis' principle only concerns itself with systems that conserve energy. The Lagrange function for such systems can be written as

$$L = \sum p_k \dot{q}_k - H, \quad (1.88)$$

where $H = E$ is the total energy, and the action S then becomes

$$S = \int_{t_1}^{t_2} \sum_k p_k \dot{q}_k dt - E(t_2 - t_1). \quad (1.89)$$

We can take advantage of this breakup of S into two parts by designing another type of variation. As before, we induce variations in S by replacing $q_k(t)$ by new functions $s_k(t) + \delta q_k(t)$, but now also allow the *transit time* to vary during the virtual motion. *However, we maintain the same total energy for all virtual motions.* Then,

$$\delta' S = \delta' \int_{t_1}^{t_2} \sum_k p_k \dot{q}_k dt - E(\delta t_2 - \delta t_1) \quad (1.90)$$

follows from (1.89) where δt_1 and δt_2 are the variations in time at the endpoints of the motion. This new type of variation, symbolized by δ' , is clearly different from the δ -variation visualized in Sec. 1-4. There, we required that the transit time for all motions be the same ($\delta t_1 = \delta t_2 = 0$). Now, we demand that the *energy* be the same ($\delta E = 0$).

To see what consequences this new variation has for S we also compute $\delta' S$ directly from (1.20). Then,

$$\delta' S = \int_{t_1 + \delta t_1}^{t_2 + \delta t_2} L(q + \delta q, \dot{q} + \delta \dot{q}, t) dt - \int_{t_1 + \delta t_1}^{t_2 + \delta t_2} L(q, \dot{q}, t) dt \simeq \int_{t_1}^{t_2} \delta L dt + [L \delta t]_{t_1}^{t_2},$$

if all virtual displacements are considered small. Now, the time integral of δL has already been calculated in (1.38). We introduce that expression in the evaluation of $\delta' S$ and find that

$$\delta' S = \int_{t_1}^{t_2} \sum_k \left(\frac{\partial L}{\partial q_k} - p_k \right) \delta q_k dt + \left[\sum_k p_k (\delta q_k + \dot{q}_k \delta t) \right]_{t_1}^{t_2} - [H \delta t]_{t_1}^{t_2} \quad (1.91)$$

after inserting p_k for $\partial L / \partial \dot{q}_k$ and $\sum_k p_k \dot{q}_k - H$ for L . This expression is true for any system, whether conservative or not. We now specialize it as follows: first only the actual motion of the system is considered. This drops out the first term on the right due to the validity of the Lagrange equations. The second term drops out if we replace our former boundary conditions, $\delta q_k(t_1) = \delta q_k(t_2) = 0$, by

$$\delta q_k(t) + \dot{q}_k(t) \delta t = 0, \quad \text{for } t_1 \text{ and } t_2. \quad (1.92)$$

This simply means that all virtual motions are made to pass through the same endpoints, even if they do so at different times. Finally we assume H is constant and equal to the total energy E . Then $\delta' S = E(\delta t_2 - \delta t_1)$. Inserting this result into (1.90) we learn

$$\delta' \int_{t_1}^{t_2} \sum_k p_k \dot{q}_k dt = 0 \quad (1.93)$$

under the stated variations and boundary conditions. This is Maupertuis' principle. The action in this case is represented by

$$S_0 = \int_{t_1}^{t_2} \sum_k p_k \dot{q}_k dt \quad (1.94)$$

and is often called the *reduced action* to distinguish it from S .

The contents of this principle can be visualized as follows: represent the system by plotting its coordinates as a point in the n -dimensional space they define. As the system moves this point moves too, tracing out a curve in n dimensions. To apply the variational principle we run the representative point along all paths joining A and B in Fig. 1.7 for which the energy of the system remains the same. Then, according to (1.93) and (1.94) the system moves from A to B in such a way that the integral S_0 is a minimum¹³ when integrated over whatever transit time that is required to go from A to B while keeping the energy constant.

¹³ Actually an extremum.

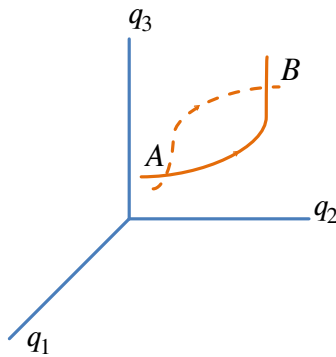


Figure 1.7: Illustration of two possible paths of the representative point for a system with three degrees of freedom.

We have thus shown that (1.93) is true if the Lagrange equations describe the motion. To complete the demonstration of Maupertuis' principle we must also show that it leads to the correct equation of motion. Carrying out the variation of the action S_0 we get a sum of three terms:

$$\delta' S_0 = \int_{t_1}^{t_2} \sum_k \delta p_k \dot{q}_k dt + \int_{t_1}^{t_2} \sum_k p_k \delta \dot{q}_k dt + [\sum_k p_k \dot{q}_k \delta t]_{t_1}^{t_2}. \quad (1.95)$$

We know from previous work that $\delta \dot{q}_k = [(d/dt)(\delta q_k)]$. But what is δp_k ? The condition that all virtual paths are to have the same energy enters at this point. We always have

$$\delta H = \delta(\sum_k p_k \dot{q}_k) - \delta L = \sum_k (p_k - \frac{\partial L}{\partial \dot{q}_k}) \delta \dot{q}_k + \sum_k (\delta p_k \dot{q}_k - \frac{\partial L}{\partial q_k} \delta q_k).$$

But $p_k = \partial L / \partial \dot{q}_k$ by definition and $\delta H = 0$ by design. Therefore,

$$\sum_k \delta p_k \dot{q}_k = \sum_k \frac{\partial L}{\partial q_k} \delta q_k$$

and one finds that

$$\delta' S_0 = \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial q_k} - \dot{p}_k \right) \delta q_k dt + \left[\sum p_k (\delta q_k + \dot{q}_k \delta t) \right]_{t_1}^{t_2} \quad (1.96)$$

after interchanging δ and $\frac{d}{dt}$ as before in the second term on the right of (1.95) and integrating by parts. The boundary conditions at the end-points get rid of the last term and

$$\delta' S_0 = \int_{t_1}^{t_2} \sum_k \left(\frac{\partial L}{\partial q_k} - \dot{p}_k \right) \delta q_k dt = 0$$

results, a *condition* determining the equations of motion. It obviously leads to the Lagrange equations again! It also goes without saying that the question of independence of the δq_k must be raised again if there are constraints on the motion. The reader will have an opportunity to worry about such matters in Probs. 1-4. and 1-5.

We realize from the preceding discussion that the only important point about H is that it must be constant. It need not be the total energy. However, if H is the total energy, one can cast (1.93) into a number of equivalent forms, each having its own special significance. Thus, if $H = E = T + V$, then T must be a homogeneous quadratic function of the velocities and V independent of them. Therefore, $\sum_k p_k \dot{q}_k = 2T$ and

$$\delta' \int_{t_1}^{t_2} 2T dt = 0 \quad (1.97)$$

is another rendering of (1.91) in terms of the total kinetic energy of the system. In particular, if the motion takes place under no forces, then T is also constant, and

$$\delta' \int_{t_1}^{t_2} dt = \delta(t_2 - t_1) = 0.$$

This is the *principle of least time* for motion under no forces. It first gained recognition as a principle governing the propagation of light (Fermat). For such applications it is useful to introduce the optical path length $n ds$ (n = the index of refraction) of the light ray and write

$$\delta' \int_{s_1}^{s_2} n ds = 0,$$

since $dt = ds/v$, where $v = c/n$ is the velocity of light in the medium, c its velocity *in vacuo*. This transformation then gives a variational principle for determining the *path* of a light ray from s_1 to s_2 in a refractive medium. The time variable has disappeared from the scene completely.

A similar transformation is possible in dynamics. We go back to (1.97) and apply it to the motion of a particle of mass m . Then

$$2T dt = 2T \frac{ds}{v} = mv ds$$

where ds is the arc length travelled in time dt with speed v . Consequently¹⁴,

$$\delta' \int_{s_1}^{s_2} mv ds = 0 \tag{1.98}$$

determines how the particle moves from s_1 to s_2 . However, we still have to express v in terms of the position of the particle in order to clear (1.98) of all time variables. This is simple to do using the energy equation: if the particle moves in a potential field V , then $mv = \sqrt{2m(E - V)}$, and

$$\delta' \int_{s_1}^{s_2} \sqrt{E - V} ds = 0. \tag{1.99}$$

This form is due to Jacobi (1842). It is in complete analogy with Fermat's principle. The potential field throughout which the particle moves acts on the particle like a heterogeneous medium with refractive index $\sim \sqrt{E - V}$, and (1.99) is a variational principle for determining the *path* of the particle. As a corollary, we see that a free particle moves along the shortest path from s_1 to s_2 ,

$$\delta' \int_{s_1}^{s_2} ds = \delta(s_2 - s_1) = 0$$

since mv is constant in this case and factors out of (1.98) or (1.99).

Let us apply Jacobi's form to the simple case where a free particle enters a region of space where the potential suddenly decreases from zero to a constant value $-V_0$ (the particle is suddenly accelerated at O as it crosses the boundary between the two regions, see Fig. 1.8). The

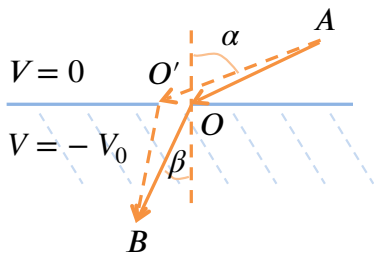


Figure 1.8: Application of Fermat's principle to the motion of a particle.

particle starts at A with energy E and has to reach B . We try the path AOB and calculate the reduced action:

$$S_0 = \sqrt{2mE}(AO) + \sqrt{2m(E + V)}(OB).$$

S_0 has to be stationary for the actual path. In order to minimize S_0 we try various paths by moving the point O where the particle strikes the boundary, and so determine the relation between the angles α and β that holds for the actual motion. Moving O to O' , i.e. going along the

dotted path, changes AO and OB by $OO' \sin \alpha$ and $-OO' \sin \beta$. Hence, S_0 changes by

$$\delta' S_0 \sim (\sqrt{E} \sin \alpha - \sqrt{E + V_0} \sin \beta) OO'.$$

This is zero for arbitrary displacements OO' if

$$\frac{\sin \alpha}{\sin \beta} = \sqrt{1 + \frac{V_0}{E}}.$$

Thus, the particle is *refracted* at the interface between the two potential regions: the actual path is *not* simply the straight line joining A and B . If we call $\sqrt{1 + \frac{V_0}{E}}$ the relative index of refraction of the region of negative potential to free space, then this result is analogous to Snell's law of light refraction. Physically, the origin of this effect is obvious: the particle receives an impulse normal to the interface that kinks its path.

It should be realized, however, that Jacobi's form holds for a much wider class of systems than that of a single particle moving in a potential field. The only requirement is that the Lagrange function be expressible in the form

$$L = \frac{1}{2} \sum_{kl} a_{kl} \dot{q}_k \dot{q}_l - V(q_1, q_2, \dots).$$

The a_{kl} only depend on the coordinates and not their time derivatives. Then, the element of action $2T dt$ can be re-expressed as

$$2T dt = \sqrt{2T} \sqrt{\sum_{kl} a_{kl} \dot{q}_k \dot{q}_l} dt = \sqrt{2T} ds$$

to *define* the arc length

$$ds = \sqrt{\sum_{kl} a_{kl} dq_k dq_l} dt \quad (1.100)$$

travelled by the system point in time dt with "speed" $\sqrt{2T}$. Using this extended meaning of ds , (1.99) is unaltered in form, but its meaning is different. It now gives the "path" of the representative point of the system in n -dimensional space. This only coincides with the path of the system in ordinary space if the degrees of freedom are 3 or less and refer to the position of the system in space.

Differential equations determining this path may readily be derived from (1.99). We carry out the derivation for the special, but most usual case of a particle moving in a potential V in two dimensions. Let (x, y) be the cartesian coordinates of the particle. Following the tradition for curvilinear coordinates in two dimensions, we introduce new coordinates $q_1 = u$ and $q_2 = v$, where the grid of curves

$$\begin{aligned} u(x, y) &= \text{constant} \\ v(x, y) &= \text{constant} \end{aligned}$$

defines these coordinates. The line element ds is written as

$$(ds)^2 = E_u(du)^2 + 2F_{uv} du dv + G_v(dv)^2, \quad (1.101)$$

where E_u , F_{uv} and G_v are the *first differential parameters* introduced by Gauss. They are functions of u and v in general.

Since the variation in (1.99) means considering different paths passing through the same endpoints, the line element ds must also be varied. We consider variations in u only and find

$$\delta' \int_{s_1}^{s_2} \sqrt{E - V} ds = \int_{s_1}^{s_2} (-) \frac{\partial V}{\partial u} \frac{\delta u}{2\sqrt{E - V}} ds + \int_{s_1}^{s_2} \sqrt{E - V} \delta(ds) = 0.$$

Now,

$$ds \delta(ds) = (E_u du + F_{uv} dv) \delta(du) + \frac{1}{2} \left[\frac{\partial E_u}{\partial u} (du)^2 + 2 \frac{\partial F_{uv}}{\partial u} du dv + \frac{\partial G_v}{\partial u} (dv)^2 \right] \delta u.$$

The term multiplying (du) in this expression gives the contribution

$$- \int_{s_1}^{s_2} \frac{d}{ds} \left\{ \sqrt{E - V} \left(E_u \frac{du}{ds} + F_{uv} \frac{dv}{ds} \right) \right\} \delta u ds,$$

after integrating by parts and enforcing the boundary conditions $\delta u(s_1) = \delta u(s_2) = 0$. Therefore (1.99) is satisfied in the present case if

$$\begin{aligned} \frac{1}{\sqrt{E - V}} \frac{d}{ds} \left\{ \sqrt{E - V} \left(E_u \frac{du}{ds} + F_{uv} \frac{dv}{ds} \right) \right\} &= - \frac{1}{2} \frac{1}{E - V} \frac{\partial V}{\partial u} \\ &+ \frac{1}{2} \left\{ \frac{\partial E_u}{\partial u} \left(\frac{du}{ds} \right)^2 + 2 \frac{\partial F_{uv}}{\partial u} \frac{du}{ds} \frac{dv}{ds} + \frac{\partial G_v}{\partial u} \left(\frac{dv}{ds} \right)^2 \right\}. \end{aligned} \quad (1.102)$$

The equation for v follows by interchanging u and v , E_u and G_v . We have thus obtained differential equations for the path, or *orbit*, of a particle moving in two dimensions in an arbitrary potential field $V(u, v)$. These equations contain as a special case the differential equations for the orbit of a particle in a central potential $V(r)$, $r = \sqrt{x^2 + y^2}$, as will be shown in Chap. 2. Finally, if the particle is moving under no forces, (1.102) becomes

$$\frac{d}{ds} \left(E_u \frac{du}{ds} + F_{uv} \frac{dv}{ds} \right) = \frac{1}{2} \left\{ \frac{\partial E_u}{\partial u} \left(\frac{du}{ds} \right)^2 + 2 \frac{\partial F_{uv}}{\partial u} \frac{du}{ds} \frac{dv}{ds} + \frac{\partial G_v}{\partial u} \left(\frac{dv}{ds} \right)^2 \right\}.$$

This is one of the differential equations determining the shortest path between s_1 and s_2 . Such a path is called a *geodesic*. A particle moving freely in the u, v space thus follows one of the geodesics in that space.

1-9 Properties of the Action Functions

We remarked at the beginning of this section that the adjective *least* attached to the various variational principles can be misleading; we

could equally well have a *greatest* action as far as the equations of motion are concerned. We did not ever have to raise the issue of whether S is a maximum or a minimum in finding the equations of motion. Typically, physical systems in motion are only prepared to admit to the stationary character of the action. Just as typically, however, the various principles of least action have been made the basis of many a philosophical foray into the "deeper" meaning of the laws of dynamics. The trouble always seems to be the appearance of a finite time interval in the calculation of say the action S . This gives the impression that the system already "knows" at time t_1 where it is going to be at t_2 , i.e. it endows mechanics with a *teleological*¹⁵ character. Actually this impression is false. As we have seen, only the stationary character of the action and not its value determines the equations of motion.

This does not mean, however, that the action functions themselves are devoid of meaning. For consider (1.91) and (1.96); up to now we have always used these expressions to find the equations of motion by stipulating boundary conditions such that the contribution at the "boundaries" t_1 and t_2 vanishes. Now, we turn the procedure around and evaluate the variations in S and S_0 due to changes in the coordinates and the time at these boundaries *for the actual motion*. We set $\delta q_k(t_1)$ and δt_1 equal to zero and call it the time at the other endpoint of the action. Then,

$$\delta' S = \sum_k p_k (\delta q_k + \dot{q}_k \delta t) - H \delta t \quad (1.103)$$

expresses the variation in S as the coordinates and the time at the other endpoint are allowed to vary. The circumstance of this variation should be clearly distinguished from the variations employed so far. We are now assuming that Lagrange equations hold along all paths that are considered. Therefore, (1.103) shows how S depends on the coordinates of the motion at time t . In fact, by setting $\delta t = 0$, we obtain the *partial* derivatives of S with respect to the q_k :

$$\delta' S = \sum_k \frac{\partial S}{\partial q_k} \delta q_k = \sum_k p_k \delta q_k \quad \text{for } \delta t = 0,$$

or

$$\frac{\partial S}{\partial q_k} = p_k. \quad (1.104)$$

On the other hand, setting all the δq_k equal to zero does not define the partial derivative with respect to time. This requires in addition that the position of the endpoint not vary when t is varied. Therefore, not δq_k but $\delta q_k + \dot{q}_k \delta t$ must be set equal to zero at time t . Then

$$\delta' S = -H \delta t$$

defines the partial derivative

$$\frac{\partial S}{\partial t} = -H. \quad (1.105)$$

¹⁵ Teleology, *n.*, (from Greek *telos* "end" and *logos* "reason"), explanation by reference to some purpose, end, goal, or function. Traditionally, it was also described as final causality, in contrast with explanation solely in terms of efficient causes (the origin of a change or a state of rest in something). *Encyclopaedia Britannica*, 2023.

Thus, the partial time derivative of S gives H ; by contrast its total time derivative gives L :

$$\frac{dS}{dt} = L. \quad (1.106)$$

The latter statement follows from (1.20) with $t_2 = t$.

The corresponding partial derivatives of S_0 follow from (1.96) in a similar fashion. From

$$\delta' S_0 = \sum_k p_k (\delta q_k + \dot{q}_k \delta t) \quad (1.107)$$

we conclude that

$$\frac{\partial S_0}{\partial q_k} = p_k, \quad \text{and} \quad \frac{\partial S_0}{\partial t} = 0, \quad (1.108)$$

provided of course that the energy is conserved (otherwise (1.107) is not valid!). Because of this requirement S_0 contains the energy E as a parameter. The derivative $\partial S/\partial E$ has a surprising significance. We go back to (1.90) and also vary the energy from path to path. Then,

$$\delta' S = \delta' S_0 - E \delta t - \delta E (t - t_1).$$

But, according to (1.106), $\delta' S$ equals $-E \delta t$ if $H = E$. Therefore,

$$\delta' S_0 - \delta E (t - t_1) = 0$$

defines the partial derivative we seek:

$$\frac{\partial S_0}{\partial E} = t - t_1. \quad (1.109)$$

Thus, the change of S_0 with energy gives the transit time from t_1 to t . This result allows one to express the coordinates as a function of time by inserting the value for S_0 and differentiating:

$$\frac{\partial}{\partial E} \int_{s_1}^{s_2} \sqrt{2(E - V)} ds = \int_{s_1}^{s_2} \frac{ds}{\sqrt{2(E - V)}} = t - t_1. \quad (1.110)$$

Here $s - s_1$ is the distance travelled in time $t - t_1$. Together with the equation of the path this relation completely determines the motion.

The relations (1.106) through (1.109) will appear again in Chap. 7 from a different point of view. However, it is useful and interesting to indicate one basic feature of (1.106) at this point. From its definition H on the right of that equation depends on q , \dot{q} and t . On the other hand, \dot{q} may be eliminated in favour of p by inverting the relations $p_k = \partial L/\partial \dot{q}_k$; then p may be re-expressed as in (1.104) and we get

$$\frac{\partial S}{\partial t} + H\left(q, \frac{\partial S}{\partial q}, t\right) = 0. \quad (1.111)$$

This is the famous *Hamilton-Jacobi partial differential equation* for determining S . It will emerge again in Chap. 7 - ostensibly from an entirely

different point of view. If H is conserved and also happens to be the total energy, then the Hamilton-Jacobi equation turns into a partial differential equation for S_0 ,

$$H\left(q, \frac{\partial S_0}{\partial q}\right) - E = 0. \quad (1.112)$$

The meaning and practical significance of both these equations are discussed at length in Chap. 7.

1-10 D'Alembert's Principle

The action principles we have discussed so far are *integral* principles of mechanics in this respect as they concern themselves with properties of the motion over a finite time interval $t_2 - t_1$. *Differential* principles are also available which concern themselves with the motion at a given instant t and small variations about this motion. Perhaps the most important principle of the latter type is due to d'Alembert (1758)¹⁶. His principle simply states that the inertial forces are in equilibrium with the applied forces on any system. To appreciate this statement we go back to Newton I in (1.1) and notice that all bodies remain at rest or in uniform rectilinear motion if no forces act. A measure of their resistance to change, or *inertia*, therefore is \dot{p} . D'Alembert calls $\tilde{\mathbf{F}} = -\dot{\mathbf{p}}$ the *inertial force*. If k again labels a typical particle, then (1.23) gives the equation of motion for this particle in the form

$$F_k + \tilde{F}_k = 0.$$

Multiplying both sides by the virtual displacements δx_k and summing on all k , one has

$$\sum_k (F_k + \tilde{F}_k) \delta x_k = 0. \quad (1.113)$$

This all seems rather trivial until we realize that the forces of constraint drop out of the virtual work $\sum_k F_k \delta x_k$. Therefore, this equation only contains the applied forces. It states that the applied forces are in equilibrium with the inertial forces (the total virtual work is zero).

Equation (1.113) is the mathematical realization of d'Alembert's principle. It is of course identical with our former equation, (1.28),

$$\sum_k (F_k + \tilde{F}_k) \delta x_k = \sum_k F_k \delta x_k + \delta T - \frac{d}{dt} \left(\sum_k m_k \dot{x}_k \delta x_k \right) = 0 \quad (1.114)$$

if we restore the definition of \tilde{F}_k . We can also use this equation to determine the equations of motion. Take the case first where there are no constraints on the system. Then $F_k - \dot{p}_k = 0$ must follow if (1.114) is to hold for arbitrary displacements δx_k . But if constraints are present, this result does *not* follow. A transformation to independent coordinates

¹⁶ Jean-le-Rond d'Alembert (1717 - 1783), a famous French mathematician and philosopher.

is necessary before any conclusions can be drawn. But we can introduce such transformations since (1.114) is a *scalar* relation and therefore valid in any set of coordinates. Accordingly, we pass to new coordinates $x_k \rightarrow f_k(q_1, q_2, \dots)$ and observe the various parts of (1.114) transform:

$$\sum_k F_k \delta x_k = \sum_k Q_k \delta q_k$$

from the two forms (1.29) and (1.41) for the virtual work,

$$\delta T = \delta \left(\sum_k \frac{1}{2} m_k \dot{f}_k^2 \right) = \sum_k \frac{\partial T}{\partial q_k} \delta q_k + \sum_k \frac{\partial T}{\partial \dot{q}_k} \delta \dot{q}_k$$

from the functional form of T in the new coordinates, and

$$\sum_k m_k \dot{x}_k \delta x_k = \sum_{k,l} m_k \dot{f}_k \frac{\partial f_k}{\partial q_l} \delta q_l$$

from the definition of the virtual displacements δx . But

$$\frac{\partial f_k}{\partial q_l} = \frac{\partial \dot{x}_k}{\partial \dot{q}_l} = \frac{\partial \dot{f}_k}{\partial \dot{q}_l}$$

from (1.36). Therefore,

$$\sum_k m_k \dot{x}_k \delta x_k = \sum_{k,l} m_k \dot{f}_k \frac{\partial f_k}{\partial \dot{q}_k} \delta \dot{q}_k = \sum_k \frac{\partial}{\partial \dot{q}_k} \left(\sum_l \frac{1}{2} m_l \dot{f}_l \right) \delta q_k = \sum_k \frac{\partial T}{\partial \dot{q}_k} \delta q_k. \quad (1.115)$$

This is the crucial relation. Knowing it, we can write (1.114) as

$$\sum_k \left[Q_k + \frac{\partial T}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_k} \right) \right] \delta q_k = 0. \quad (1.116)$$

Now, assume that there are no constraints on the motion, or if there are, of the holonomic variety. Then, all the δq_k can be chosen independently so that

$$Q_k + \frac{\partial T}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_k} \right) = 0.$$

These are the, by now, familiar Lagrange equations for holonomic systems.

In the form (1.114) d'Alembert's principle has an interesting consequence. If a system is in *equilibrium* under applied forces, then $T = 0$ and all $\dot{x}_k = 0$ so that

$$\delta W = 0. \quad (1.117)$$

The virtual work of the applied forces must be zero for equilibrium. This succinct statement summarizes the entire subject of *statics* (systems in equilibrium).

Two further differential principles, Gauss' principle of least constraint, and Hertz's principle of least curvature (which is a special case of Gauss'

principle) are also available, but do not introduce any new insight that is not already contained in d'Alembert's principle. In summary, we remark: the differential principles only require a knowledge of differentiation to derive the equations of motion. On the other hand, the integral principles (which require the calculus of variations) have perhaps a more dependable "feel", since their implementation does not depend on spotting relations like (1.116) and (1.113) ahead of time. Their dependence on a single scalar function L makes it manifest that the equations they lead to, *must* have the same form in all sets of coordinates. This is also true of course for the differential principles, but perhaps less obvious to the uninitiated (see Prob. 1-6).

These comments bring us to the end of our discussion of the general principles of mechanics. Our next task (it will occupy the remainder of the book!) is to apply these principles to the motion of particles and systems of particles that are of interest in physics.

Problems

- 1-1. Find out how Newton's law of motion, (1.2), is modified when referred to a frame of reference that moves with constant acceleration.
- 1-2. Describe the sequence of events that allow a cat to reorient while falling from an inverted position.
- 1-3. Verify the calculation of the action $S[\bar{f}]$ given in the text for a particle falling under gravity. Guess at some other forms $\bar{f}(t)$ for the virtual motion and verify that $S[\bar{f}]$ is stationary for the actual motion in each case.
- 1-4. Discuss the validity of Maupertuis' principle in the presence of holonomic constraints. Pay particular attention to the case where such constraints may be time-dependent.
- 1-5. Discuss the validity of Maupertuis' principle in the presence of non-holonomic constraints of the type given in (1.52). Pay particular attention to the case where such constraints may be time-dependent.
- 1-6. Show by a *direct* transformation of coordinates $q_k = f_k(q'_1, q'_2, \dots)$ that the Lagrange equations are unaltered in form. This property is referred to as the *covariance* of these equations.
- 1-7. Find the geodesics joining any two points of a sphere. Can you always find the shortest distance between the points in question? Explain.

Chapter 2 Particle Dynamics

2-1 Introduction

This chapter is devoted to a discussion of the motion of a single particle under specified forces. Its purpose is twofold. Firstly, the simplicity of such problems will allow us to become acquainted with the methods of chapter 1 in specific applications, without undue complication. Secondly, the physical discussion of the two most important particle systems in nature - the celestial system and the atomic system - require for their building blocks an understanding of single particle motion in a force field (the planets around the sun for the former, the electrons around the nucleus in the latter). Also, most of the information on atomic, nuclear and nucleon structure comes from devising *scattering experiments* in which the system under study is bombarded with a probe particle. The full interpretation of the results of such experiments lies outside the capabilities of classical mechanics; one has to appeal to quantum mechanics. However, most of the concepts and parameters of the classical description appear unscathed in the quantum mechanical version and it is therefore of utmost importance to study the classical one-body scattering problem in detail. This is done in Sec. 2-6.

One of many practical aspects of applied classical mechanics appears in everyday life in the form of satellite communications and weather systems, which depend for their successful operation on knowing quite precisely how such a satellite will orbit the earth. The actual calculation of satellite orbits is much more complex than the standard orbit problem we discuss in Sec. 2-4. But the complexity is in degree of sophistication and not in the principles involved. References are provided there to bring the curious reader up to date on what these complexities are. The classic problem of planetary motion around the sun falls in this category also and is of course indelibly woven into the historical development of mechanics.

2-2 Systems with One Degree of Freedom

Consider the motion of a particle, mass m , in a conservative force field $V(x)$, where x measures the displacement from some standard position.

The Lagrange function reads

$$L = \frac{1}{2}m\dot{x}^2 - V(x). \quad (2.1)$$

The equation of motion for x is

$$\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} = -\frac{\partial V}{\partial x} - m\ddot{x} = 0,$$

or

$$m\ddot{x} = F(x), \quad (2.2)$$

where $F(x) = -\partial V/\partial x$ is the force on the particle at position x . This equation is special only to the extent that $F(x)$ was derivable from a potential. Otherwise, we would use the general Lagrange equations (1.43); then F could have any form, e.g. $F = F(x, \dot{x}, t)$, that is not necessarily derivable from a potential function. However, if F is derivable from a potential as supposed above, we can integrate (2.2) immediately. Multiply by \dot{x} . Then,

$$m\dot{x}\ddot{x} = \frac{d}{dt} \left(\frac{1}{2}m\dot{x}^2 \right) = F(x)\dot{x},$$

or

$$\left[\frac{1}{2}m\dot{x}^2 \right]_{x_0}^x = \int_{x_0}^x f(x) dx = -V(x) + V(x_0),$$

after integrating both sides from x_0 to x . The last statement says

$$\frac{1}{2}m\dot{x}^2 + V(x) = \frac{1}{2}m\dot{x}_0^2 + V(x_0) = \text{a constant } E, \quad (2.3)$$

which just reproduces the law of energy conservation. We could have realized this directly from the form of L in (2.1): since L does not depend on time, the quantity

$$H = \frac{\partial L}{\partial \dot{x}} \dot{x} - L = \frac{1}{2}m\dot{x}^2 + V(x) = E \quad (2.4)$$

is conserved and equals the total energy, according to (1.77) and (1.81). A second integration of the energy equation gives the position at time t :

$$t - t_0 = \sqrt{\frac{m}{2}} \int_{x_0}^x \frac{dx}{\sqrt{E - V(x)}}, \quad (2.5)$$

where x_0 is the position of the particle at time t_0 . We have chosen the positive square root, implying that the particle is moving from x_0 to x . Notice that *two* boundary conditions are required before this solution becomes unique: for example, the position and velocity at time t_0 are required to determine E in (2.5).

If the force F in (2.2) depends on the velocity of the particle, or the time, then the integral (2.5) is invalidated, and one has to seek other methods of integration. The variety of such problems is endless, depending on the form of F , and so the integration of various special cases

seems pointless here. Rather we discuss some standard problems that appear frequently in physical applications, and that will furthermore serve as "guinea pigs" for the more exotic solution techniques that are developed in subsequent chapters.

(i) *Free fall under gravity*

Here the accelerating force is the gravitational pull of the earth. Near the earth's surface this is a constant force mg . The potential is therefore $-mgx$ if we measure x from the point of release. Then, the time to fall a distance x is given by (2.5):

$$t - t_0 = \sqrt{\frac{m}{2}} \int_0^x \frac{dx}{\sqrt{E + mgx}} = \frac{1}{g} \left[\sqrt{\frac{2}{m}(E + mgx)} - \sqrt{\frac{2E}{m}} \right]. \quad (2.6)$$

The value of E reflects with what velocity the particle was released at $t = t_0$. In free fall from rest, $E = 0$, and

$$t - t_0 = \sqrt{\frac{2x}{g}}, \quad \text{or} \quad x = \frac{1}{2}g(t - t_0)^2$$

gives the distance fallen in time $t - t_0$.

(ii) *Harmonic oscillator in one dimension*

A particle attracted to a fixed point by a force proportional to its displacement from that point performs an oscillatory motion. For if $F = -m\omega^2x$, calling the constant of proportionality $m\omega^2$, then the potential is $V = \frac{1}{2}m\omega^2x^2$ and the equation of motion and equation of energy read

$$\ddot{x} + \omega^2x = 0 \quad (2.7)$$

and

$$\dot{x}^2 + \omega^2x^2 = \frac{2E}{m}, \quad (2.8)$$

respectively.

Since \dot{x} must be real, we find from the second equation that the particle is not found further from the origin than $\pm a = \pm\sqrt{2E/m\omega^2}$; the velocity vanishes at this value of x but not the force. The particle is pulled back towards the origin, so that the sense of its motion reverses. The positions $x = \pm a$ are called *turning points* of the motion. Obviously, the particle oscillates back and forth between a and $-a$ an infinite number of times. The quantity a is called the *amplitude* of oscillation. It is determined by the energy of the system, for a given force constant $m\omega^2$. Therefore, instead of specifying the energy at which the motion occurs, we can prescribe its amplitude. The position at any time t is then given by

$$t - t_0 = \frac{1}{\omega} \int_0^x \frac{dx}{\sqrt{a^2 - x^2}} = \frac{1}{\omega} \sin^{-1}\left(\frac{x}{a}\right),$$

or¹⁷

$$x = a \sin(\omega t - \omega t_0) \quad (2.9)$$

if the particle passes the origin at $t = t_0$. The constant ωt_0 is called the phase¹⁸. The solution we have obtained for x confirms our picture of the motion. The particle oscillates between $+a$ and $-a$; the time taken for one complete oscillation (called the *period*) is

$$T = \frac{2\pi}{\omega}. \quad (2.10)$$

It is inversely proportional to ω , i.e. to the square root of the coupling strength per unit mass.

The problem we have just solved is an example of a *bounded* motion in mechanics. The particle never gets further than $\pm a$ from the origin. It is possible to decide quite generally from the energy equation when such motion occurs. Writing

$$\dot{x} = \pm \sqrt{\frac{2}{m}(E - V(x))}, \quad (2.11)$$

we see that \dot{x} must vanish if $V(x)$ equals E so that the particle is turned back by the potential $V(x)$ at values of x that are roots of

$$V(x) = E. \quad (2.12)$$

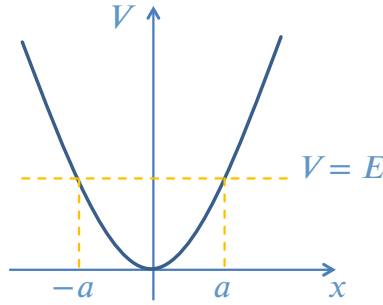
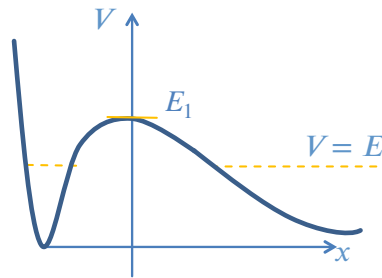
If this equation has two *adjacent* roots at $x = a$ and $x = b$ respectively, (it may have more than two roots of course), we can get an oscillatory motion between a and b , the particle shuttling back and forth an infinite number of times. The motion is therefore bounded by the turning points a and b . Furthermore, (2.11) shows (the plus and minus sign) that the velocity is the same whether the particle is "coming or going" at x . Therefore, the transit time from a to b is the same as the transit time from b to a and stays constant for every subsequent traversal (provided the system does not lose energy). Hence, we can define a period T of the motion; T is taken as twice the time required to go from a to b . Obviously, the latter time equals the time taken for the particle to return to any position x on its path.

On the other hand, a single root means the particle just turns around once and proceeds back to infinity from whence it came. Hence, the shape of the potential $V(x)$ and the energy determine the type of motion. In Fig. 2.1 oscillatory motion occurs for all positive values of the energy, $E > 0$; in Fig. 2.2 oscillatory motion occurs for $E_1 > E > 0$, if the particle is located on the left of the origin, unbounded motion if the particle is located on the right. The "hump" in the potential in Fig. 2.2 is termed a *potential barrier*. For $E < E_1$ the particle cannot¹⁹ penetrate such a barrier since that would require a negative kinetic energy. On the other hand, if

¹⁷ Or $x = a \cos(\omega t + \alpha)$, where α is some other phase. Either form is a linear combination of the fundamental pair of solutions $x_{\pm} = \exp(\pm i\omega t)$ of (2.7) with a different boundary condition.

¹⁸ Sometimes the negative of this quantity is called the phase.

¹⁹ This restriction is lifted in quantum mechanics.


 Figure 2.1: The potential V of a harmonic oscillator.

 Figure 2.2: The potential V with a barrier.

$E > E_1$ the particle passes over the hump, and the motion is no longer oscillatory.

The period of the motion is easily determined in the general case. By definition, the period T is the time taken to go "there and back", starting at a particular turning point a , going to a further one b and returning to a . Thus,

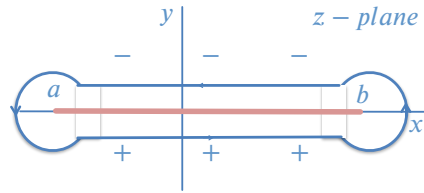
$$T = \int_a^b \frac{dx}{\sqrt{\frac{2}{m}(E - V)}} - \int_b^a \frac{dx}{\sqrt{\frac{2}{m}(E - V)}},$$

using the minus sign on the second radical for the return journey. Mathematically speaking, both branches of the multivalued function $\sqrt{E - V}$ appear in T , if $V(x)$ is analytic in a suitable region in the complex plane, this facet allows one to express T as a contour integral²⁰ around the branch points (at a and b) of this function, in the complex plane $z = x + iy$,

$$T = \int_C \frac{dz}{\sqrt{\frac{2}{m}(E - V)}} = \frac{\partial}{\partial E} \int_C \sqrt{2m(E - V)} dz. \quad (2.13)$$

The "dumbbell" contour C is shown in Fig. 2.3. We have to cut the z -plane from a to b along the real axis to make the integrand in (2.13) single-valued, after choosing the positive square root just below the cut, the negative square root just above. A word of caution is in order: since the turning points a and b are zeros of the integrand in (2.13), the first integral for T is actually an improper one. Physically, we want the oscillation to have a *finite* period. Therefore, we also stipulate that

²⁰ A reader unfamiliar with the elements of complex integration may consult H. Margenau and G.M. Murphy, *The Mathematics of Physics and Chemistry*, D. van Nostrand Company, Inc., New York, 1943.

Figure 2.3: The contour of integration C .

the improper integral should converge. This is the case, provided the singularities are not too severe (see Prob. 2-1). Incidentally, the second form for T in (2.13) shows a premise of things to come.

Since

$$\int_C \sqrt{2m(E - V)} dx = S_0$$

is just the reduced action for one period of oscillation, one has $T = \partial S_0 / \partial E$ as a special case of (1.109). If we can find the reduced action function S_0 over one period in a periodic motion by some means or other, the period of this motion is just the energy derivative of that S_0 .

As an example of the application of (2.13), take the harmonic oscillator problem. Then $E = \frac{1}{2}m\omega^2 a^2$, $V(x) = \frac{1}{2}m\omega^2 x^2$, and

$$\omega T = \int_C \frac{dz}{\sqrt{a^2 - z^2}}.$$

We can deform the contour C into a large circle of radius R surrounding the origin without crossing any singularities of the integrand. Therefore

$$\omega T = -i \int_T \frac{dz}{z} = -i(2\pi i) = 2\pi$$

by Cauchy's theorem, in agreement with the result (2.10). Notice in passing that the period of oscillations in an oscillator potential is *independent* of the amplitude of oscillation. This is not generally true. Any deviation from the parabolic potential shape introduces an amplitude dependence into T . However, many potential fields in physics can be usefully replaced by a quadratic approximation near the equilibrium position of the system ($x = 0$ in Fig. 2.1) and therefore retain this simple property approximately. The entire subject of small oscillations that is fundamental to the understanding of molecular and crystal lattice vibrations is based on such a premise. We return to such matters in Chap. 4.

2-3 Systems with Two or More Degrees of Freedom

An important example of a system with two degrees of freedom consists of a mass m moving in a potential field V . The Lagrange function in any system of coordinates $q_1 = u$, $q_2 = v$ is

$$L = \frac{1}{2}m\left(\frac{ds}{dt}\right)^2 - V(u, v) \quad (2.14)$$

where ds is the line element given by (1.101). The general problem posed by (2.14) does not have either u or v as cyclic coordinates. Therefore there are no conserved canonical momenta. However, the energy E is conserved,

$$E = \frac{1}{2}m\left(\frac{ds}{dt}\right)^2 + V(u, v), \quad (2.15)$$

since L does not contain the time explicitly. We discuss two special cases of (2.14):

(i) *Projectile motion near the surface of the earth*

We position the projectile in cartesian coordinates $u = x, v = y$ with respect to the point of projection. If the x -axis points vertically downwards, the potential energy is just a function of x and is given by $V = -mgx$, where as before g is the acceleration due to gravity. The line element is, trivially,

$$(ds)^2 = (dx)^2 + (dy)^2$$

making $E_u = 1, F_{uv} = 0$ and $G_V = 1$. Therefore

$$L = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) + mgx.$$

Since the coordinate y is cyclic, the linear momentum

$$p_y = \frac{\partial L}{\partial \dot{y}} = m\dot{y}$$

is conserved. This relation allows us to pass from derivatives in t to derivatives in y :

$$\frac{d}{dt} = \dot{y} \frac{d}{dy} = \frac{p_y}{m} \frac{d}{dy},$$

and, together with the energy equation,

$$E = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) - mgx = \frac{1}{2}m\dot{x}^2 + \frac{p_y^2}{2m} - mgx$$

determines the motion completely. We write \dot{x} in the alternative forms

$$\dot{x} = \pm \sqrt{\frac{2}{m}\left(E - \frac{p_y^2}{2m} + mgx\right)} = \frac{p_y}{m} \frac{dx}{dy}$$

and find

$$t = \int_0^x \frac{dx}{\sqrt{\frac{2}{m}\left(E - \frac{p_y^2}{2m} + mgx\right)}}; \quad y = \int_0^x \frac{p_y dx}{\sqrt{2m\left(E - \frac{p_y^2}{2m} + mgx\right)}}$$

if the projectile leaves the origin at $t = 0$. The integration is elementary and we find

$$x = v_x t + \frac{1}{2}gt^2; \quad \text{or} \quad x = \frac{v_x}{v_y}y + \frac{g}{2v_y^2}y^2$$

for the motion in time, and the path (a parabola). Here, v_x and v_y are the velocity components at the projection point. The ambiguity in sign of the radicals is settled by knowing the direction of v_x . Alternatively we could have found the path directly from the variational principle [(1.102) and its sister equation for v].

(ii) *Motion in a central field*

A problem of particular physical interest arises in (2.14) when V is spherically symmetric, i.e. a *central potential*.

This means that V only depends on the distance $r = OP$ of the particle P from some fixed point O (the force center, see Fig. 2.4). The force $F(r)$ that this potential gives rise to points along the radius vector OP , i.e. it is a *central force*.

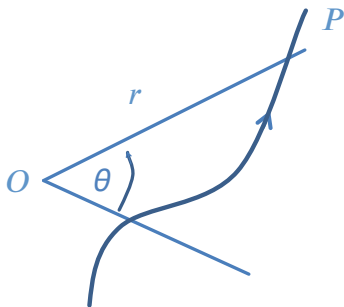


Figure 2.4: A particle moving in a trajectory measured from the fixed point O .

Polar coordinates are indicated: setting $u = r$ and $v = \theta$ the line element becomes

$$(ds)^2 = (dr)^2 + r^2(d\theta)^2,$$

making $E_u = 1$, $F_{uv} = 0$ and $G_v = r^2$. The Lagrange function is therefore

$$L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - V(r). \quad (2.16)$$

The angle θ is cyclic. Therefore, the canonical momentum

$$p_\theta = \frac{\partial l}{\partial \dot{\theta}} = mr^2\dot{\theta} \quad (2.17)$$

is constant in a central field. Physically, p_θ is the angular momentum of the particle about O . It is conserved because the radial force $F(r)$ has no moment about O . As before, (2.17) allows us to pass from derivatives in t to derivatives in θ :

$$\frac{d}{dt} = \dot{\theta} \frac{d}{d\theta} = \frac{p_\theta}{mr^2} \frac{d}{d\theta}. \quad (2.18)$$

Taken together with the angular momentum equation, the equation of energy

$$E = \frac{1}{2}m\dot{r}^2 + \frac{1}{2}mr^2\dot{\theta}^2 + V(r) = \frac{1}{2}m\dot{r}^2 + \frac{p_\theta^2}{2mr^2} + V(r)$$

provides us with a complete picture of the motion again. Introduce the *effective potential*

$$U(r) = \frac{p_\theta^2}{2mr^2} + V(r) \quad (2.19)$$

consisting of the original potential V and the *centrifugal potential*²¹ $p_\theta^2/2mr^2$. A knowledge of U and E determines the motion completely. To see this, write the energy equation as

²¹ See Sec. 2-9.

$$\dot{r} = \pm \sqrt{\frac{2}{m}(E - U)}. \quad (2.20)$$

This is now a one-dimensional problem as in Sec. 2-2. We ask: Where does the radial motion cease? This is determined by the roots of

$$U(r) = E.$$

A single root means that the motion is unbounded. The particle comes in from infinity, is turned around at the *closest distance of approach* to 0 and is pushed out to infinity again. On the other hand, two roots at r_{min} and r_{max} ,

$$U(r_{min}) = U(r_{max}) = E$$

mean that \dot{r} vanishes twice. If r_{min} and r_{max} are separated by a region of potential that is energetically accessible to the particle, then the motion can be bounded: r oscillates between circles of radii r_{min} and r_{max} . This does not mean, however, that the motion ceases at these points or that the particle returns to its original position when r returns to its original value. To find out what angle OP swings through when r changes, i.e. to find the *orbit*, we write (2.20) in the form

$$\frac{dr}{dt} = \pm \sqrt{\frac{2}{m}(E - U)} = \frac{p_\theta}{mr^2} \frac{dr}{d\theta}, \quad (2.21)$$

and calculate the change in θ from

$$d\theta = \pm \frac{1}{r^2} \frac{p_\theta dr}{\sqrt{2m(E - U)}}.$$

A single integration gives

$$\theta - \theta_0 = \int_{r_{min}}^r \frac{p_\theta}{\sqrt{2m(E - U)}} \frac{dr}{r^2}, \quad (2.22)$$

if we set $\theta = \theta_0$ when $r = r_{min}$. Since r must increase with increasing θ after this point the sign on the radical is determined. The time taken for OP to reach the position θ is given by the first form in (2.21):

$$t - t_0 = \int_{r_{min}}^r \frac{dr}{\sqrt{\frac{2}{m}(E - U)}}, \quad (2.23)$$

where t_0 is the time at $\theta = \theta_0$. As OP changes from r_{min} to r_{max} and back, θ changes by

$$\Delta\theta = 2 \int_{r_{min}}^{r_{max}} \frac{p_\theta}{\sqrt{2m(E-U)}} \frac{dr}{r^2} = -\frac{\partial}{\partial p_\theta} \int_C \sqrt{2m(E-U)} dz \quad (2.24)$$

in time

$$\Delta t = 2 \int_{r_{min}}^{r_{max}} \frac{dr}{\sqrt{2m(E-U)}} = \frac{\partial}{\partial E} \int_C \sqrt{2m(E-U)} dz \quad (2.25)$$

if we use the same trick as before to express the integrals on r as contour integrals in the complex plane $z = r + i\zeta$. The contour C is the same as that depicted in Fig. 2.3 after replacing a and b by r_{min} and r_{max} . The surprising symmetry in the expressions for $\Delta\theta$ and Δt (and θ and t for that matter) as derivatives of the same integral with respect to the constants of motion p_θ and E is no accident. Its appreciation must, however, wait for the developments of Chap. 6. Equation (2.25) is once more a special case of (1.109) with the reduced action calculated over a full period of the variable r .

Equations (2.22) and (2.23) determine the motion completely. As far as the orbit is concerned, we realize from the plus and minus signs on the radial velocity in (2.21) that the turning points can be approached from either side with the same radial velocity, i.e. the orbit is *symmetric* about the line $\theta = \theta_0$. The same is true at the other turning point. Therefore, the entire orbit can be constructed once the segment from r_{min} to r_{max} is known, by successively "folding over" this segment about the lines $\theta = \theta_0$ and $\theta_0 + \Delta\theta/2$, respectively.

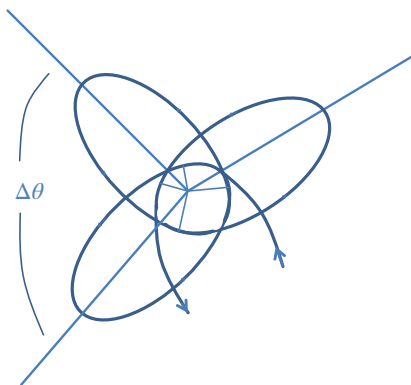


Figure 2.5: A possible central orbit.

A typical orbit is shown in Fig. 2.5. This figure illustrates an important fact: the orbit will not be a *closed* one unless $\Delta\theta$ is some rational fraction of 2π . Otherwise, the particle never returns to its original position for any θ and the orbit will eventually fill out the entire annulus between the two circles $r = r_{min}$ and $r = r_{max}$ in Fig. 2.5. It is also clear that the orbit

in a central field always lies in a plane, the orientation of this plane being determined by the direction of the angular momentum \mathbf{p}_θ . Vectorially, we have

$$\mathbf{p}_\theta = \mathbf{r} \times \mathbf{p},$$

and since the field $V(r)$ is independent of the orientation of \mathbf{r} in space, \mathbf{p}_θ will clearly maintain its original *direction* also (we have already seen that its magnitude is constant). This means that the orbit lies in the plane perpendicular to \mathbf{p}_θ , passing through the center of force.

We stress the fact: all the information gained up to this point (and it is a great deal) has not required us to specify $V(r)$. Of course the actual shape of the orbit does depend on $V(r)$ and we have to know it before the orbit can be obtained. We will do such a calculation in the next section for the important cases of inverse ($\sim 1/r$) and quadratic ($\sim r^2$) central fields. However, two general comments on the orbit problem are in order here: often the problem is not to find the orbit but to find the force (or potential) giving rise to a given orbit, i.e. to turn (2.22) "inside out" for the function U . This is easily accomplished by obtaining instead a *differential equation* for the orbit. Instead of integrating (2.21) to find the orbit, we differentiate it to find the force (no, $dU/d\theta$ is *not* zero, because a total derivative is being formed):

$$p_\theta \frac{d}{d\theta} \left(\frac{1}{r^2} \frac{dr}{d\theta} \right) = \frac{-\frac{dU}{d\theta}}{\sqrt{\frac{2}{m}(E-U)}} = \frac{-\frac{\partial V}{\partial r} \frac{dr}{d\theta}}{\frac{p_\theta}{mr^2} \frac{dr}{d\theta}} = -\frac{mr^2}{p_\theta} \frac{\partial U}{\partial r},$$

or

$$\frac{d^2}{d\theta^2} \left(\frac{1}{r} \right) = \frac{mr^2}{p_\theta^2} \frac{\partial U}{\partial r}.$$

Entering the value of U from (2.19), writing $F = -\partial V/\partial r$ and setting $u = 1/r$, we find the more usual form of this equation (see also Prob. 2-3),

$$\frac{d^2 u}{d\theta^2} + u = -\frac{m}{p_\theta^2 u^2} F\left(\frac{1}{u}\right). \quad (2.26)$$

Knowing the orbits, the force can therefore be determined up to a constant (m/p_θ^2). No further information is necessary.

The other comment concerns calculating the orbits for a given force law. The construction of (2.22) has reduced this problem "to quadratures"; mathematically the problem has been solved. The integral on the right of this equation can often be expressed in terms of elementary functions. More often not so elementary functions appear (elliptic integrals) depending on the form of $V(r)$. Whittaker ("Analytical Dynamics", p81, fourth edition, Cambridge University Press, London, 1936) gives a rather complete list of integrable potentials in this sense. Of course there is nothing to prevent a numerical evaluation of the integral if it converges. Furthermore, as we have seen, a knowledge of the segment of the path

from r_{min} to r_{max} determines the entire orbit, reducing such numerical work considerably.

2-4 Kepler's Problem

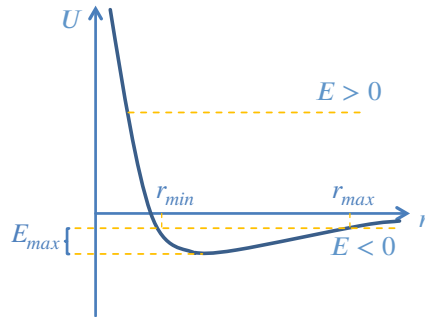
A problem of fundamental significance, both historically and scientifically²², concerns the motion of a particle in an attractive inverse square central force field²³. This force field is generated by the potential ($\alpha > 0$)

$$V = -\frac{\alpha}{r}, \tag{2.27}$$

which at the same time represents the potential energy of a planet around the sun, or an electron around an atomic nucleus. The effective potential for a particle of angular momentum p_θ is

$$U = \frac{p_\theta^2}{2mr^2} - \frac{\alpha}{r} \tag{2.28}$$

in this field. The function U is seen to approach $+\infty$ at the origin, and zero from the negative side as $r \rightarrow \infty$. Therefore, two turning points are possible if $E < 0$, but only one for $E \geq 0$. Hence the orbits in the potential (2.27) are uniquely classified as bounded or unbounded accordingly, as $E < 0$ or $E \geq 0$, see Figs. 2.6 and 2.7.



²² The science of mechanics may be said to have begun seriously with the explanation of Kepler's laws of planetary motion by Newton (1687).

²³ For the reader curious enough to want to know how planetary (or satellite) orbits are actually calculated we recommend D. Brouwer and G.M. Clemence, "Methods of Celestial Mechanics", Academic Press Inc., London and New York, 1961.

Figure 2.6: The potential U showing positive and negative energies.

We first assume E is negative, $E = -|E|$. The orbit is a bound one and its shape is given by (2.22):

$$\begin{aligned} \theta - \theta_0 &= \int_{r_{min}}^r \frac{p_\theta}{\sqrt{2m(-|E| - \frac{p_\theta^2}{2mr^2} + \frac{\alpha}{r})}} \frac{dr}{r^2} \\ &= \left[\cos^{-1} \left[\frac{\frac{p_\theta^2}{m\alpha} \frac{1}{r} - 1}{\sqrt{1 - \frac{2|E|p_\theta^2}{m\alpha^2}}} \right] \right]_{r_{min}}^r, \end{aligned}$$

or (if we measure θ from the point of closest approach r_{min}),

$$\frac{l}{r} = 1 + e \cos \theta \tag{2.29}$$

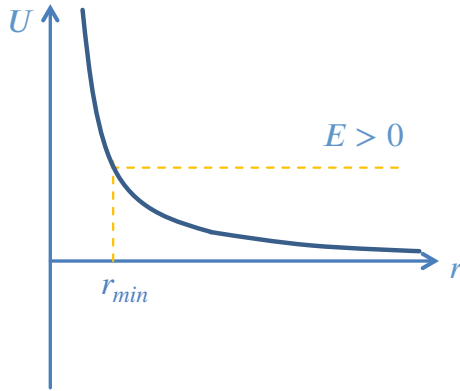


Figure 2.7: A potential $U \sim 1/r^2$ that allows for a positive energy only.

since the ratio in brackets of the cosine bracket is unity when $r = r_{min}$. The constants l and e are

$$l = \frac{p_\theta^2}{m\alpha}; \quad e = \sqrt{1 - \frac{2p_\theta^2}{m\alpha^2}|E|}. \quad (2.30)$$

The reader will recognise (2.29) as the equation for a conic section with semi-latus rectum l and eccentricity e , referred to polar coordinates with origin at one focus. The constants l and e are called the *geometrical* constants of the orbit as opposed to the *dynamical* constants E and p_θ . Clearly $e < 1$ from (2.30) so the bound orbit is an (*ellipse*²⁴). Its semi-major and -minor axes are

$$a = \frac{l}{1 - e^2} = \frac{\alpha}{2|E|}; \quad b = \frac{l}{\sqrt{1 - e^2}} = \frac{p_\theta}{\sqrt{2m|E|}} \quad \text{elliptic orbits.} \quad (2.31)$$

The geometrical significance of these various constants is shown in Fig. 2.8. The force center is at S , the one focus of the ellipse (we use

²⁴ There are two limiting cases: if $|E|$ attains its largest physically allowable value $E_{max} = p_\theta^2/2mr^2$, the radii r_{min} and r_{max} coalesce into a single radius r_0 and a circular orbit of this radius is performed. The other limiting case, $|E| = 0$ gives rise to a *parabola*.

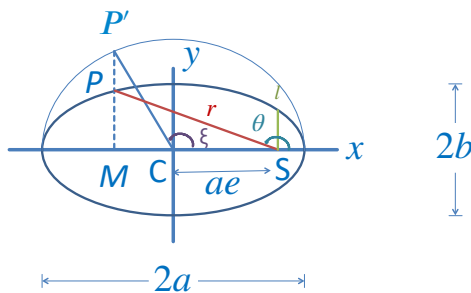


Figure 2.8: The particle of mass m moving in a vertical circular path, and acted on by gravity.

S instead of 0 in Fig. 2.8 to recall the role of the sun as the center of attraction for the planetary system). The orbit is a closed one, the angle $\Delta\theta$ of (2.24) equalling 2π in this case. The least and greatest distances

from S are

$$r_{min} = \frac{l}{1+e} = a(1-e); \quad r_{max} = \frac{l}{1-e} = a(1+e). \quad (2.32)$$

The time required to reach any point in the orbit is given by (2.23). In the present example

$$t - t_0 = \sqrt{\frac{m}{2|E|}} \int_{r_{min}}^r \frac{r \, dr}{\sqrt{-\frac{p_\theta^2}{2m|E|} + \frac{\alpha}{|E|}r - r^2}}.$$

But

$$-\frac{p_\theta^2}{2m|E|} + \frac{\alpha}{|E|}r - r^2 = -la + 2ar - r^2 = (ae)^2 - (a-r)^2,$$

using the relations (2.30) and (2.31) to introduce the geometrical constants of the elliptic orbit. The integral can be evaluated by introducing the angular parameter ξ , where

$$a - r = ae \cos \xi. \quad (2.33)$$

[The value $\xi = 0$ puts r at its minimum value $r_{min} = a(1 - e)$]. Then,

$$t = \sqrt{\frac{m}{2|E|}} \int_0^\xi a(1 - e \cos \xi) \, d\xi = a\sqrt{\frac{m}{2|E|}} (\xi - e \sin \xi)$$

if t is also measured from the closest distance of approach. Thus,

$$r = a(1 - e \cos \xi); \quad t = a\sqrt{\frac{m}{2|E|}} (\xi - e \sin \xi) \quad (2.34)$$

provide parametric equations for the radial position at time t . Finally, (2.29) gives the corresponding angular position and thus provides us with a complete picture of the motion.

In astronomy, the angles ξ and θ denote the *eccentric anomaly* and *true anomaly* of a planet when, as here, they are measured from the closest point of approach (the *perihelion*) to the sun. The geometrical meaning of θ is clear but what is ξ ?²⁵ Its meaning can be ascertained by inspection from Fig. 2.8. If we draw a circle of radius a centered at the center C of the ellipse and extend the perpendicular PM through P on the major axis to intersect this circle at P' , then the position (a, ξ) of P' is related to the position (r, θ) of P by

$$ae + a \cos(\pi - \xi) = r \cos(\pi - \theta).$$

This equation expresses the length SM in two ways. Substituting for $\cos \theta$ from the orbit equation (2.29) one has

$$ae - a \cos \xi = -\frac{l}{e} + \frac{r}{e} = -\frac{a}{e}(1 - e^2) + \frac{r}{e} \quad (2.37)$$

²⁵ The inverse problem of finding ξ at a given time is interesting in that it gave rise to the development of what we now know as Bessel functions. Since $\xi(t)$ must be an odd function of nt in the interval $(-\pi, \pi)$ it will bear expansion in a Fourier sine series. Consequently, its derivative $d\xi/d(nt)$ has a cosine expansion

$$\begin{aligned} \frac{d\xi}{d(nt)} &= \frac{1}{2}b_0 \\ &+ \sum_{m=1}^{\infty} b_m \cos mt, \end{aligned} \quad (2.35)$$

with $b_m =$

$$\begin{aligned} &\frac{1}{\pi} \int_{-\pi}^{\pi} \frac{d\xi}{d(nt)} \\ &\times \cos mnt \, d(nt) \\ &= 2J_m(me), \end{aligned}$$

where $J_m(z)$ is a Bessel function of order m (see for example W. Magnus and F. Oberhettinger, *Functions of Mathematical Physics*, Chelsea Publishing Company, New York, 1949). Therefore, upon integration,

$$\begin{aligned} \xi(nt) &= nt \\ &+ 2 \sum_{m=1}^{\infty} \frac{1}{m} J_m(me) \\ &\times \sin mnt. \end{aligned} \quad (2.36)$$

This expression is originally due to Lagrange.

or $r = a(1 - e \cos \zeta)$, which confirms the first member of (2.34). The second member of (2.34) shows that the time taken to complete one orbit is

$$T = 2\pi \sqrt{\frac{ma^3}{\alpha}} = \pi\alpha \sqrt{\frac{m}{2|E|^3}}, \quad (2.38)$$

that is, the periodic time only depends on the major axis (equivalently, total energy) of the orbit. The proportionality $T^2 \sim a^3$ is one of the laws of planetary motion discovered empirically by Kepler. Actually Kepler's work went a lot further than just this statement. Since the ratio $2\pi/T = n$ is the mean angular velocity of the planet over one orbit, the second member of (2.34) is equivalent to

$$nt = \zeta - e \sin \zeta. \quad (2.39)$$

This is Kepler's equation expressing the *mean anomaly* nt in terms of the eccentric anomaly. ²⁶ To express nt in terms of the true anomaly θ , we multiply r as given by the first of equations (2.34) together with (2.29) to find

$$(1 - e \cos \zeta)(1 + e \cos \theta) = 1 - e^2, \quad \text{or} \quad \sin \zeta = \sqrt{1 - e^2} \frac{\sin \theta}{1 + e \cos \theta}. \quad (2.40)$$

Consequently, the mean anomaly is connected to the true anomaly through

$$nt = \sin^{-1} \left\{ \sqrt{1 - e^2} \frac{\sin \theta}{1 + e \cos \theta} \right\} - e \sqrt{1 - e^2} \frac{\sin \theta}{1 + e \cos \theta}. \quad (2.41)$$

This equation gives the time required to reach a specified angle θ in the elliptic orbit. Knowing ζ and its relation to θ also allows us to write down the cartesian coordinates (x, y) of the particle with respect to the center C of the orbit:

$$\begin{aligned} x = a \cos \zeta &= a \frac{e + \cos \theta}{1 + e \cos \theta} \\ y = a \sqrt{1 - e^2} \sin \zeta &= a(1 - e^2) \frac{\sin \theta}{1 + e \cos \theta} \end{aligned} \quad (2.42)$$

so that

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1; \quad b = a \sqrt{1 - e^2} \quad (2.43)$$

as expected.

Now, suppose the energy is positive, $E > 0$. Then, the orbit becomes unbounded. The shape is given by (2.29) but with E replacing $-|E|$ in the formula (2.30) for the eccentricity e . Thus e becomes greater than unity: *hyperbolic* orbits are described for positive energies. The relation between the semi-major axis, the semi-latus rectum l and the energy E is changed to

$$a = \frac{l}{e^2 - 1} = \frac{\alpha}{2E} \quad \text{hyperbolic orbits} \quad (2.44)$$

²⁶ Kepler published his laws of planetary motion in 1609.

instead of the first member of (2.31); the minor axis b becomes imaginary as it should and the closest distance of approach is

$$r_{min} = \frac{l}{e+1} = a(e-1). \tag{2.45}$$

The relations (2.34) are also untenable. Instead, the time to reach a given radial distance r is now given by exactly similar methods as

$$r = (e \cosh \zeta - 1); \quad t = a \sqrt{\frac{m}{2E}} (e \sinh \zeta - \zeta), \tag{2.46}$$

where the parameter ζ now varies from $-\infty$ to $+\infty$. The orbit itself is shown in Fig. 2.9.

Finally, we consider the repulsive potential $V = \alpha/r$. Then, the effective potential U is always positive, and only positive values of E are physically permissible: the motion is always unbounded, see Fig. 2.7. The shape of the orbit is hyperbolic,

$$\frac{l}{r} = -1 + e \cos \theta, \tag{2.47}$$

where l and e are again given by (2.30) with E replacing $-|E|$. The time-dependence is given by the parametric equations

$$r = a(a \cosh \zeta + 1), \quad t = a \sqrt{\frac{m}{2E}} (e \sinh \zeta + \zeta) \tag{2.48}$$

instead of (2.46).

From (2.47) we see that the closest distance of approach to the force center is

$$r_{min} = \frac{l}{e-1} = a(e+1), \tag{2.49}$$

instead of the value (2.45). The orbit is a hyperbola with the force center at the *external* focus, contrary to the case of an attractive field [(2.29) for $e > 1$], where the force center lies at the *internal* focus, cf. Fig. 2.9.

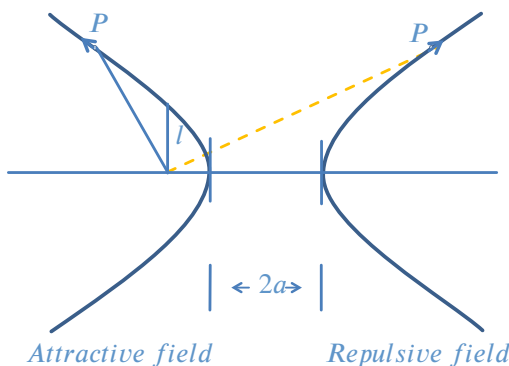


Figure 2.9: Attractive and repulsive motion.

2-5 The Space Oscillator

A particle moving subject to the potential

$$V(r) = \frac{1}{2}m\omega^2 r^2 \quad (2.50)$$

($\omega =$ a constant) is called an *isotropic space oscillator*. As before the potential is a central one, so the orbit always lies in a plane. But contrary to our previous example, all orbits are bounded, the potential V increasing without limit as $r \rightarrow \infty$.

The effective potential is sketched in Fig. 2.10. Any energy $> E_{min}$ gives a bound orbit. Its shape can be computed from (2.22). However, it is much easier to proceed as follows: since $V(r)$ can be written as

$$V(r) = \frac{1}{2}m\omega^2(x^2 + y^2) \quad (2.51)$$

in terms of cartesian coordinates (x, y) through the center of force, we simply obtain two one-dimensional oscillators along x and y

$$\begin{aligned} \ddot{x} + \omega^2 x &= 0 \\ \ddot{y} + \omega^2 y &= 0, \end{aligned} \quad (2.52)$$

that are not coupled to each other.

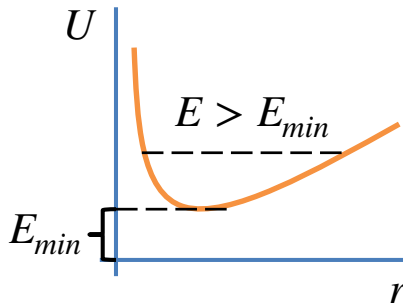


Figure 2.10: Bound state in the effective potential.

Therefore, the motion in time is given essentially by

$$x = a \cos \omega t, \quad y = b \sin \omega t, \quad (2.53)$$

where a and b are the maximum amplitudes of oscillation in the x and y directions. Consequently, the orbit is the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (2.54)$$

about the center of force which lies at the center of this ellipse (not the focus). As in the case of the $1/r$ field, the orbit is also closed. The

connection between the geometrical constants of this orbit and the dynamical constants p_θ and E follow from (2.52):

$$p_\theta = m(x\dot{y} - \dot{x}y) = m\omega(ab)$$

$$E = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) + \frac{1}{2}m\omega^2(x^2 + y^2) = \frac{1}{2}m\omega^2(a^2 + b^2).$$

Therefore, both axes determine the energy (or vice versa) in this case, contrary to the situation in the $1/r$ field. However, the period of motion in the oscillator potential only depends on the force constant $T = 2\pi/\omega$, independent of the size of the orbit.

2-6 Scattering by a Central Force

We mentioned the importance of scattering experiments in physics at the beginning of this chapter. Let us now study in detail what such an experiment entails within the domain of classical mechanics.

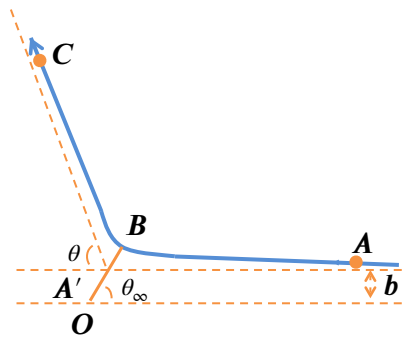


Figure 2.11: The geometry of a scattering problem.

A beam of particles is incident on a scattering center at O , which repels each particle according to some force law $F(r)$. A particular incident particle at A that is initially aimed to miss the scattering center by a distance b will be deflected along the orbit ABC , and be turned through an angle ϑ from its initial direction (Fig. 2.11). This angle is called the *scattering angle*. We know from our previous work that the orbit is symmetrical about the closest distance of approach, line OB . Consequently, the scattering angle is related to the total angle turned through by the radius vector OP as the particle moves in from infinity to its closest distance of approach. Calling this angle θ_∞ , we have

$$\vartheta = |\pi - 2\theta_\infty|. \quad (2.55)$$

(The modulus sign is necessary since ϑ is considered to lie between 0 and π by convention). Now, θ_∞ is given by (2.22) if we set the upper limit $r \rightarrow \infty$. Before using this equation, however, let us insert for p_θ its value in terms of the *impact parameter* b :

$$p_\theta = \sqrt{2mEb},$$

for a particle of energy E .

The parameters E and b are the natural parameters for describing the scattering of the particle. In terms of them we have

$$\theta_\infty = \int_{r_{\min}}^{\infty} \frac{b}{\sqrt{1 - \frac{b^2}{r^2} - \frac{V(r)}{E}}} \frac{dr}{r^2}, \quad (2.56)$$

where $V(r)$ is the potential energy. Equations (2.55) and (2.56) uniquely determine the angle through which a particle of energy E and impact parameter b is deflected. However, this is not what we want. Under normal experimental conditions, a uniform beam of n particles per unit area impinges on the scattering center with energy E , but with *different* impact parameters b . Therefore, the interesting quantity is the number of particles that are scattered per second into the solid angle $d\Omega = \sin\vartheta d\vartheta d\varphi$ subtended at (ϑ, φ) on a large sphere centered at the scattering center O (see Fig. 2.12).

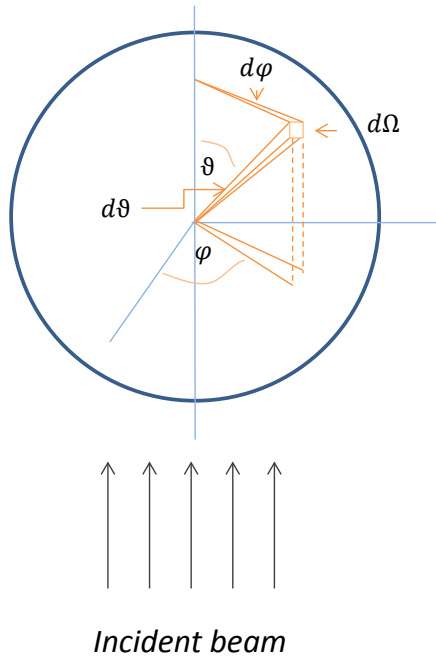


Figure 2.12: Geometry of a scattering center.

The number of particles dN passing into $d\Omega$ per second (the particle *flux*) clearly equals the number passing through the area segment $b db d\varphi$ between circles b and $b + db$ on a plane held perpendicular to the incident beam at infinity. Hence, $dN = nb db d\varphi$ where n is the incident flux. Since dN is proportional to the incident flux n , not dN but the ratio $d\sigma = dN/n$ gives a physically useful description of the effect of the scattering center; this ratio is called the *differential cross section*. We have

$$d\sigma = \frac{dN}{n} = b db d\varphi = \frac{b}{\sin\vartheta} \frac{db}{d\vartheta} \sin\vartheta d\vartheta d\varphi,$$

or

$$\frac{d\sigma}{d\Omega} = \frac{b}{\sin\vartheta} \left| \frac{db}{d\vartheta} \right|, \quad (2.57)$$

if we introduce the solid angle element $d\Omega = \sin\vartheta d\vartheta d\varphi$. The modulus keeps the intrinsically positive quantity $d\sigma/d\Omega$ positive. The derivation leading up to this formula has assumed a one-to-one correspondence between a given impact parameter b and angle of scattering ϑ . If this is not so, i.e. if the function $b = b(\vartheta)$ is multivalued, the formula (2.57) must include a sum over all these values. Finally, if convergent, the integral of (2.57) over all angles gives the *total cross section*

$$\sigma = \int \frac{d\sigma}{d\Omega} = \int \frac{b}{\sin\vartheta} \left| \frac{db}{d\vartheta} \right| d\Omega.$$

The scattering cross section is thus determined classically by knowing b as a function of ϑ (or θ_∞). This information is supplied by (2.56) for any central field $V(r)$. We calculate $d\sigma/d\Omega$ for the important case that $V(r) = \alpha/r$. Then

$$\theta_\infty = \cos^{-1} \left\{ \frac{1}{\sqrt{1 + 4E^2 b^2 / \alpha^2}} \right\}$$

by elementary integration (or simply setting $r \rightarrow \infty$ and $p_\theta = \sqrt{2mEb}$ in (2.47)). Inverting, we find

$$1 + \frac{4E^2 b^2}{\alpha^2} = \frac{1}{\sin^2(\frac{\vartheta}{2})}; \quad \frac{8E^2}{\alpha^2} b \frac{db}{d\vartheta} = -\frac{\cos(\frac{\vartheta}{2})}{\sin^2(\frac{\vartheta}{2})}$$

and

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{8E^2} \frac{\cos(\frac{\vartheta}{2})}{\sin^2(\frac{\vartheta}{2})} \frac{1}{2 \sin \frac{\vartheta}{2} \cos(\frac{\vartheta}{2})} = \left(\frac{\alpha}{4E} \right)^2 \frac{1}{\sin^4(\frac{\vartheta}{2})}. \quad (2.58)$$

This is the famous *Rutherford scattering law*²⁷ obeyed by charged particles scattered by a center of charge. This formula played a fundamental role in paving the way to what we now know about the structure of atoms. Notice that this formula is insensitive to the *sign* of the interaction. An attractive inverse field gives rise to exactly the same scattering cross section. Notice also that the cross sections are infinite in this case (why?).

A second example of scattering by a central force that is in many ways the opposite of the Rutherford scattering law is provided by the square well potential,

$$\begin{aligned} V &= -V_0, & r < a \\ &= 0, & r \geq a \end{aligned}$$

of range a and depth V_0 . This potential has a finite range and consequently gives rise to a finite total cross section, contrary to the case of the inverse field that has an infinite range.

²⁷ Phil. Mag., bf 21, 669 (1911).

The particle receives an inward radial impulse as it crosses the spherical surface $r = a$; we saw in Chap.1, Sec. 1-8, that a particle of energy E is refracted at such a potential discontinuity according to the relation

$$\frac{\sin \alpha}{\sin \beta} = n = \sqrt{1 + \frac{V_0}{E}}, \quad (2.59)$$

where α and β are the angle of incidence and angle of refraction respectively. Applying this relation at the spherical surface $r = a$ where the particle enters and leaves, we get

$$\frac{1}{2}\theta = \alpha - \beta$$

from the geometry of Fig. 2.13.

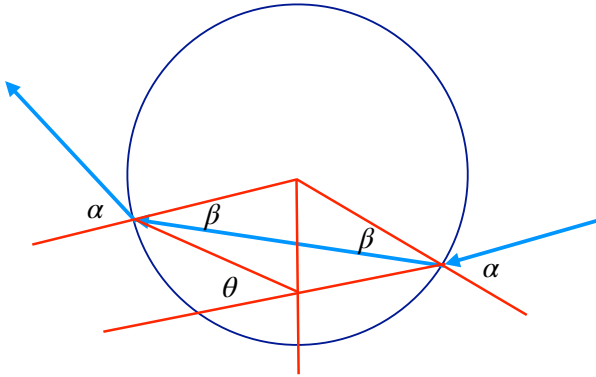


Figure 2.13: Geometry for scattering by a square well.

More formally, we can proceed from (2.56) after establishing that the effective U/E potential and distance of closest approach are given by

$$\frac{U}{E} = \frac{b^2}{r^2} - n^2 + 1 \quad \text{for } r < a, \quad \frac{b^2}{r^2} + 1 \quad \text{for } r \geq a$$

and $r_{min} = b/n$. Then, for $b \leq a$ (the particle must hit the potential to scatter),

$$\begin{aligned} \theta_{\infty} &= \int_{b/a}^a \frac{b}{\sqrt{n^2 - b^2/r^2}} \frac{dr}{r^2} + \int_a^{\infty} \frac{b}{\sqrt{1 - b^2/r^2}} \frac{dr}{r^2} \\ &= \cos^{-1}\left(\frac{b}{na}\right) + \frac{\pi}{2} - \cos^{-1}\left(\frac{b}{a}\right) \\ &= \left(\frac{\pi}{2} - \beta\right) + \frac{\pi}{2} - \left(\frac{\pi}{2} - \alpha\right) \end{aligned} \quad (2.60)$$

using Fig. 2.13 again to identify the angles $\cos^{-1}(b/na)$ and $\cos^{-1}(b/a)$. Finally, $\theta = (\pi - 2\theta_{\infty})$, so we regain (2.59). Writing the refraction law as

$$\frac{\sin \beta}{\sin \alpha} = \frac{\sin(\alpha - \theta/2)}{\sin \alpha} = \frac{1}{n},$$

and noting that $b/a = \sin \alpha$, one finds the impact parameter b as a function of ϑ . The differential cross-section follows immediately from (2.57):

$$\begin{aligned} \frac{b^2}{a^2} &= \frac{\sin^2 \frac{\vartheta}{2}}{1 - \frac{2}{n} \cos \frac{\vartheta}{2} + \frac{1}{n^2}} \quad (b \leq a) \\ \frac{d\sigma}{d\Omega} &= \frac{a^2}{4 \cos \frac{\vartheta}{2}} \frac{(\cos \frac{\vartheta}{2} - \frac{1}{n})(1 - \frac{1}{n} \cos \frac{\vartheta}{2})}{(1 - \frac{2}{n} \cos \frac{\vartheta}{2} + \frac{1}{n^2})^2}, \quad (\frac{1}{n} \leq \cos \frac{\vartheta}{2} \leq n). \end{aligned} \tag{2.61}$$

The total cross-section is finite: $\sigma = \pi a^2$.

2-7 The Two-Body Problem

So far, we have pretended that the center of force in both the bound and scattering problems of the previous two sections was fixed. This is not the case in the physical world. The naturally occurring central force systems are interactions *between* two masses (the planet-sun system in astronomy, the electron-proton system in atomic hydrogen, the neutron-proton system in the deuteron). We are dealing with a *two-body* problem, and, since the forces acting on each are equal and opposite by Newton’s third law, the motion of one influences the motion of the other. We now show how this problem can be solved exactly in terms of the theory of central motion developed in Sec. 2-3.

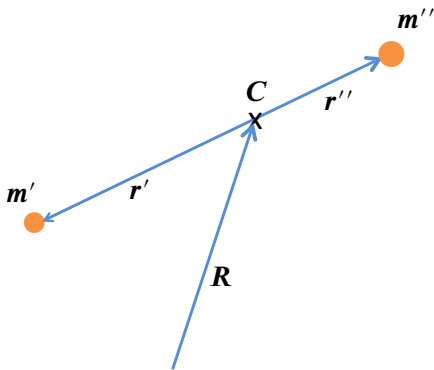


Figure 2.14: Relative and center-of-mass coordinates for the two-body problem.

Suppose then that m' and m'' are two interacting particles. We locate them at \mathbf{r}' and \mathbf{r}'' with respect to their center of mass C which lies at \mathbf{R} say, see Fig. 2.14. Then, the Lagrange function is

$$L = \frac{1}{2}(m' + m'')\dot{R}^2 + \frac{1}{2}m'\dot{r}'^2 + \frac{1}{2}m''\dot{r}''^2 - V(|\mathbf{r}' - \mathbf{r}''|), \tag{2.62}$$

since the interaction potential can only depend on the relative separation of the two particles. We see immediately that \mathbf{R} is cyclic so that the

center of mass moves with a constant momentum

$$\mathbf{P} = \nabla_{\mathbf{R}} L = (m' + m'')\dot{\mathbf{R}}.$$

This result is also obvious from (1.7), for our system has no external forces acting on it. Our problem is therefore to determine the motion of the two particles relative to their center of mass, i.e. in the center-of-mass system. We can also do this from L in (2.62). But there is one difficulty: the vectors \mathbf{r}' and \mathbf{r}'' are not independent and the system is overspecified. Indeed,

$$m'\mathbf{r}' + m''\mathbf{r}'' = 0 \quad (2.63)$$

just gives the center-of-mass position in the center-of-mass system and is therefore identically zero. So we can use either \mathbf{r}' or \mathbf{r}'' as the other vector to describe the motion. However, the form of the potential energy in L suggests that neither \mathbf{r}' nor \mathbf{r}'' , but rather their difference

$$\mathbf{r} = \mathbf{r}' - \mathbf{r}''$$

is the appropriate coordinate to use. We have

$$\mathbf{r}' = \frac{m''}{m' + m''}\mathbf{r}; \quad \mathbf{r}'' = -\frac{m'}{m' + m''}\mathbf{r}, \quad (2.64)$$

so that L reads

$$L = \frac{1}{2}(m' + m'')\dot{\mathbf{R}}^2 + \frac{1}{2}\frac{m'm''}{m' + m''}\dot{\mathbf{r}}^2 - V(r), \quad (2.65)$$

in the independent position coordinates \mathbf{R} and \mathbf{r} . The motion of the center-of-mass is given by the first term as we have already seen. The remaining part

$$L' = \frac{1}{2}\mu\left(\frac{d\mathbf{r}}{dt}\right)^2 - V(r); \quad \mu = \frac{m'm''}{m' + m''} \quad (2.66)$$

governs the relative motion. It is *identical* with the problem of a particle of *reduced mass* μ moving around a *fixed* force center with velocity $d\mathbf{r}/dt$, as comparison with (2.14) will show. Thus, the two-body problem can be solved completely by considering an appropriate *one-body* problem. [It was one of the great regrets and headaches of the founders of analytic mechanics²⁸ that systems of three or more particles do not submit to such simplification]. Once the solutions for \mathbf{r} are known we can calculate the motion of the individual masses from the relations (2.64). Notice that the shapes of their orbits are identical (since \mathbf{r}' and \mathbf{r}'' are proportional to \mathbf{r}) apart from a scaling factor due to the mass difference. Thus, if the orbit for relative motion is an ellipse, the orbits of m' and m'' will likewise be ellipses described about their common center of mass, and these ellipses share a common focus (see Prob. 2-14).

²⁸ The history of the Problem of Three Bodies is given by A. Gautier, *Essai historique sur le problème des trois corps* (Paris, 1817).

Then B moves onto the circle at C (Fig. 2.16), the length AB now equalling $|\mathbf{p}'|$. The scattering angle χ of the projectile relative to its initial direction \mathbf{p}' is evidently given by

$$\tan \chi = \frac{C'M}{AM} = \frac{OC' \sin \vartheta}{AO + OC' \cos \vartheta} = \frac{\sin \vartheta}{\frac{m'}{m''} + \cos \vartheta}, \quad (2.68)$$

since $OA/OC' = m'/m''$ in this case. This result goes over into $\tan \chi = \tan \vartheta$, or $\chi = \vartheta$, if the target is infinitely heavy ($m'' \rightarrow \infty$) as is to be expected. Since the target particle is at rest, the angle χ' is now undefined. However, we can measure the direction of travel of this particle relative to the direction of \mathbf{p}' , i.e. by the angle ABC' ; this is given by $(\pi - \vartheta)/2$.

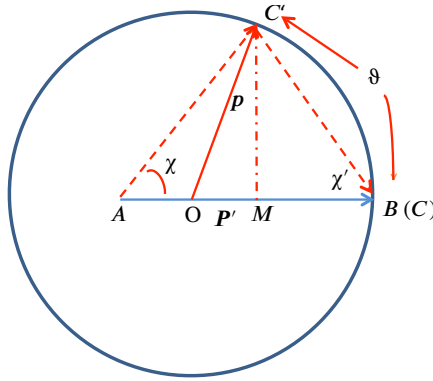


Figure 2.16: Vectors in the center-of-mass system: B moves to C.

The scattering cross-section in the laboratory system can now be calculated by using (2.68) to express the solid angle $d\omega$ in the laboratory system as

$$\begin{aligned} d\omega &= \sin \chi d\chi d\varphi = (\sin \vartheta d\vartheta d\varphi) \frac{\sin \chi d\chi}{\sin \vartheta d\vartheta} \\ &= d\Omega \frac{|1 + \frac{m'}{m''} \cos \vartheta|}{[1 + \frac{2m'}{m''} \cos \vartheta + (\frac{m'}{m''})^2]^{\frac{3}{2}}}, \end{aligned}$$

so that

$$\frac{d\sigma}{d\Omega} = \frac{d\sigma'}{d\omega} \frac{d\omega}{d\Omega} = \frac{|1 + \frac{m'}{m''} \cos \vartheta|}{[1 + \frac{2m'}{m''} \cos \vartheta + (\frac{m'}{m''})^2]^{\frac{3}{2}}} \frac{d\sigma'}{d\omega'}, \quad (2.69)$$

where $d\sigma'/d\omega$ is the differential cross-section as measured in the laboratory. This formula is rather complicated. On the other hand, the scattering cross-section of the target particle that was initially at rest is especially simple. Its scattering angle in the laboratory is just $\chi'' = (\pi - \vartheta)/2$, as we have just seen. Therefore, the differential cross-section is

$$\frac{d\sigma''}{d\omega''} = \frac{d\sigma}{d\Omega} (\vartheta = \pi - 2\chi'') \cdot (4 \cos \chi'')$$

for scattering into the solid angle $d\omega'' = \sin \chi'' d\chi'' d\phi$. For example, the recoiling particle scatters according to the law

$$\frac{d\sigma''}{d\omega''} = \left(\frac{\alpha}{2E}\right)^2 \frac{1}{\cos^3 \chi''}$$

for Rutherford scattering of the incident particle, (2.58). Here E is still the total energy in the center-of-mass system.

2-8 Motion in an Electromagnetic Field

We have already discussed the Rutherford scattering law that has important applications in atomic and nuclear physics. Another important example of single particle motion in physics is the motion of a charged particle in an electromagnetic field. Incidentally, this problem will afford us with our first example where the Lagrange function is *not* simply $T - V$.

Consider then a particle of mass m and charge e moving in an electromagnetic field characterized by electric and magnetic vectors \mathbf{E} and \mathbf{B} . The interaction of a point charge with such a field is given by the famous *Lorentz force law*

$$\mathbf{F} = e\mathbf{E} + e(\mathbf{v} \times \mathbf{B}), \quad (2.70)$$

as determined by experiment; \mathbf{v} is the velocity of the charge. How to fit this into a Lagrangian formulation for the motion of e ? We *seek* a suitable Lagrange function, guided by the following observations: since the action integral S in (1.20) is a scalar under point transformations of coordinates, the only permissible forms of L in its integrand must likewise be scalars. Furthermore, the part of L (call it L_{int}) which refers to the interaction between the charge and the electromagnetic field must have "one foot on the charge and the other in the electromagnetic field", otherwise change in the latter could not affect the motion of the former. We know that the electromagnetic field can be conveniently described³⁰ in terms of its *scalar* and *vector potentials* $\phi(\mathbf{x}, t)$ and $\mathbf{A}(\mathbf{x}, t)$, where

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}; \quad \mathbf{B} = \nabla \times \mathbf{A}. \quad (2.71)$$

Therefore, at least two scalars are available that refer jointly to the charge and the field. They are³¹

$$e\phi(\mathbf{x}, t) \quad \text{and} \quad e\mathbf{A}(\mathbf{x}, t) \cdot \mathbf{v}, \quad (2.72)$$

where \mathbf{v} is the velocity of the charge. Notice the fact that the particle *is* charged must always appear through the *coupling constant* e in these expressions, because this is how the electromagnetic field distinguishes the attribute "charge" of the particle.

³⁰ See, for example L. Landau and E. Lifshitz, *The Classical Theory of Fields*, Addison-Wesley Press Inc., Cambridge, Mass., 1951 (translated from the Russian by M. Hamermesh).

³¹ The reader may well wonder why one does not form scalar products directly from \mathbf{E} and \mathbf{B} and some vector associated with the particle. The reason for not doing so will become clear in Chap. 6.

The other ingredient we need for our Lagrange function is L_0 , describing the free motion of the charge under no forces (since it has mass). This is just given by the kinetic energy

$$L_0 = T = \frac{1}{2}mv^2.$$

We now show that the combination (the units are all MKS)

$$L = L_0 + L_{int} = \frac{1}{2}mv^2 - e\phi + e\mathbf{A} \cdot \mathbf{v} \quad (2.73)$$

leads to the correct equation of motion, i.e. gives the Lorentz force (2.70) correctly. We work in cartesian coordinates $x_k = (x, y, z)$ for the particle. Then,

$$\begin{aligned} L &= \sum_l \frac{1}{2}m\dot{x}_l^2 - e\phi(\mathbf{x}, t) + e \sum_l A_l(\mathbf{x}, t)\dot{x}_l \\ \frac{\partial L}{\partial x_k} &= -e \frac{\partial \phi}{\partial x_k} + \sum_l e \frac{\partial A_l}{\partial x_k} \dot{x}_l \\ \frac{\partial L}{\partial \dot{x}_k} &= m\dot{x}_k + eA_k = p_k, \end{aligned} \quad (2.74)$$

so that the Lagrange equation of motion for x_k is

$$\dot{p}_k = \frac{d}{dt}(m\dot{x}_k + eA_k) = -e \frac{\partial \phi}{\partial x_k} + \sum_l e \frac{\partial A_l}{\partial x_k} \dot{x}_l. \quad (2.75)$$

The canonical momentum p_k is *not* the mechanical momentum $m\dot{x}_k$ in this case; it is supplemented by a contribution eA_k from the field. However, Newton's law of motion refers only to the rate of change of the mechanical momentum. Therefore, we recast (2.75) as

$$\frac{d}{dt}(m\dot{x}_k) = -e \frac{\partial \phi}{\partial x_k} + \sum_l e \left(\frac{\partial A_l}{\partial x_k} - \frac{\partial A_k}{\partial x_l} \right) \dot{x}_l - e \frac{\partial A_k}{\partial t} \quad (2.76)$$

after evaluating the total time derivative of $A_k(\mathbf{x}, t)$ explicitly. Now

$$\begin{aligned} \sum_l \left(\frac{\partial A_l}{\partial x_k} - \frac{\partial A_k}{\partial x_l} \right) \dot{x}_l &= \left(\frac{\partial A_x}{\partial x} - \frac{\partial A_x}{\partial x} \right) \dot{x} + \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) \dot{y} + \left(\frac{\partial A_z}{\partial x} - \frac{\partial A_x}{\partial z} \right) \dot{z} \\ &= \dot{y}(\nabla \times \mathbf{A})_z - \dot{z}(\nabla \times \mathbf{A})_y \\ &= [\mathbf{v} \times \mathbf{A}]_x \end{aligned}$$

for the x -component ($k = 1$) of this sum. By symmetry the y - and z -components will lead to the corresponding components of $\mathbf{v} \times (\nabla \times \mathbf{A})$. Thus, we obtain the single vector equation

$$\frac{d}{dt}(m\mathbf{v}) = -e(\nabla\phi + \frac{\partial \mathbf{A}}{\partial t}) + e\mathbf{v} \times (\nabla \times \mathbf{A}),$$

or

$$\frac{d}{dt}(m\mathbf{v}) = e\mathbf{E} + e(\mathbf{v} \times \mathbf{B}), \quad (2.77)$$

on account of (2.71). The force on the right side of this equation therefore agrees with (2.70) with the choice (2.73) for L . Observe that for (2.77) to be an acceptable vector equation, the quantity \mathbf{B} must *not* change sign under reflections of the coordinate directions (since \mathbf{v} does), i.e. \mathbf{B} must be a *pseudovector*. It is. This can be seen directly from its construction in terms of \mathbf{A} in (2.71).

The details of the motion described by (2.77) of course depends on how \mathbf{E} and \mathbf{B} behave. Nevertheless, we can obtain some general information by going back to L and examining it for (i) cyclic coordinates, (ii) time-dependence and (iii) uniqueness. Item (i) is a specific detail influenced by how the electromagnetic fields vary in space. Without this information, we cannot say anything more. For item (ii) we calculate

$$dH = -dt\left(-e\frac{\partial\phi}{\partial t} + \sum_k e\frac{\partial A_k}{\partial t}\dot{x}_k\right),$$

where

$$H = \sum_k (m\dot{x}_k + eA_k)\dot{x}_k - \sum_k \frac{1}{2}m\dot{x}_k^2 + e\phi - \sum_k eA_k\dot{x}_k = T + e\phi \quad (2.78)$$

according to the instructions of (1.77) and (1.78). Therefore, the change in kinetic energy, dT , is

$$dT = -d(e\phi) + e\frac{\partial\phi}{\partial t}dt - \sum_k e\frac{\partial A_k}{\partial t}dx_k = \sum_k e\left(\frac{\partial\phi}{\partial x_k} + \frac{\partial A_k}{\partial t}\right)dx_k$$

or

$$dT = \sum_k eE_k dx_k = e\mathbf{E} \cdot d\mathbf{r}. \quad (2.79)$$

Change in particle energy is wrought by the electric field \mathbf{E} only; the magnetic vector \mathbf{B} does no work on the charge in motion and consequently does not contribute to dT . This can be seen directly from the form of the Lorentz force (2.70): the magnetic interaction $\mathbf{v} \times \mathbf{B}$ is always perpendicular to the displacement $d\mathbf{r} = \mathbf{v}dt$ of the particle in time dt . Therefore, this contribution to the Lorentz force is workless. (iii) It is important to observe that only the field strengths \mathbf{E} and \mathbf{B} appear in dynamical relations like (2.77) or (2.79). The scalar or vector potentials ϕ or \mathbf{A} do not appear. Indeed, they cannot, since one knows that these potentials are not unique. A transformation (called a *gauge transformation* in electrodynamics),

$$\begin{aligned} \phi &\rightarrow \phi' = \phi - \frac{\partial f}{\partial t} \\ \mathbf{A} &\rightarrow \mathbf{A}' = \mathbf{A} + \nabla f, \end{aligned} \quad (2.80)$$

where f is an arbitrary function of position and time, renders \mathbf{E} and \mathbf{B} in (2.71) unchanged in value. The equation of motion (2.77), should

likewise be immune to a gauge transformation on the Lagrange function. This is indeed so, since L goes into

$$\begin{aligned} L \rightarrow L' &= \frac{1}{2}mv^2 - e\left(\phi - \frac{\partial f}{\partial t}\right) + e(\mathbf{A} + \nabla f) \cdot \mathbf{v} \\ &= \frac{1}{2}mv^2 - e\phi + e\mathbf{A} \cdot \mathbf{v} + e\frac{df}{dt}, \end{aligned} \quad (2.81)$$

and therefore only differs from L in (2.73) by the total time derivative of f . We saw in Chap. 1, Sec. 1-4, that the addition of such a term to the Lagrange function does not alter the equations of motion. The relation (2.79) is also left unchanged, but the value of the function H itself is changed of course under a gauge transformation:

$$H \rightarrow H' = T + e\left(\phi - \frac{\partial f}{\partial t}\right)$$

instead of the value (2.78).

The Lagrange function (2.73) has many applications in pure and applied physics, governing at the same time the motion of an electron in an atom and the motion of a charged particle in a cyclotron³². Indeed, the motion of charged particles in electromagnetic fields constitute the only actual example of "particle" motion in Newtonian mechanics. We discuss two examples:

(i) Motion in a constant magnetic field. In this case

$$\phi = 0, \quad \text{and} \quad \mathbf{A} = \frac{1}{2}(\mathbf{B} \times \mathbf{r}),$$

so that

$$L = \frac{1}{2}mv^2 + \frac{1}{2}e(\mathbf{B} \times \mathbf{r}) \cdot \mathbf{v} = \frac{1}{2}mv^2 + \frac{e}{2m}(\mathbf{B} \cdot \mathbf{L}). \quad (2.82)$$

$\mathbf{L} = \mathbf{r} \times \mathbf{p}$ is the angular momentum of the particle. We see that the charge in motion interacts with external field \mathbf{B} as though it possesses a *magnetic moment*³³ $\mu = e\mathbf{L}/2m$ in its interaction with the magnetic field. The Lorentz force arising from a pure magnetic field is seen from (2.70) to be perpendicular to both \mathbf{B} and the velocity \mathbf{v} . Furthermore, the kinetic energy $\frac{1}{2}mv^2$ of the motion is constant, since \mathbf{B} does no work while moving the charge [see (2.79)]. Therefore, only the direction of \mathbf{v} is changed by the magnetic field, resulting in a *circular* orbit about the direction of \mathbf{B} (counterclockwise if the charge is positive, clockwise if the charge is negative).

The reaction force needed to keep a particle having speed v moving in a circle of radius ρ was calculated in Chap. 1, Sec. 1-6, to be $R = mv^2/\rho$. In the present circumstance, $R = eBv_{\perp}$, (v_{\perp} is the velocity component in the plane perpendicular to \mathbf{B}), so that

$$B\rho = \frac{mv_{\perp}}{e}.$$

³² A device for producing high energy beams of charged particles.

³³ See, for example, L. Landau and E. Lifschitz, *ibid.*

The product $B\rho$ is sometimes called the magnetic rigidity of a charged particle. It measures the mechanical momentum to charge ratio of the particle.

The calculation of this orbit goes as follows: choose \mathbf{B} along the z axis of a cartesian coordinate system and locate the particle at (x, y, z) . Then

$$L = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) + \frac{eB}{2}(x\dot{y} - \dot{x}y)$$

so that

$$\dot{p}_x = \frac{eB}{2}\dot{y}; \quad \dot{p}_y = -\frac{eB}{2}\dot{x}, \quad \dot{p}_z = 0,$$

with the canonical momenta themselves given by

$$p_x = m\dot{x} - \frac{eB}{2}y, \quad p_y = m\dot{y} + \frac{eB}{2}x, \quad p_z = m\dot{z}. \quad (2.83)$$

Since $\dot{p}_z = 0$ the motion along the direction of \mathbf{B} is always uniform. The projection of motion in a plane perpendicular to \mathbf{B} (the $x - y$ plane in our example) is given by

$$\dot{x} = \frac{eB}{m}\dot{y}, \quad \dot{y} = -\frac{eB}{m}\dot{x}$$

or

$$\ddot{\zeta} + i\frac{eB}{m}\dot{\zeta} = 0, \quad (2.84)$$

where $\zeta = x + iy$. Trying an oscillatory solution of the form $\zeta \sim \exp -i\omega t$, one finds that

$$\omega(\omega - \frac{eB}{m}) = 0, \quad \text{or } \omega = 0, \quad \text{or } \omega = \frac{eB}{m}.$$

Therefore,

$$\zeta = A + A'e^{-i\frac{eB}{m}t} \quad (2.85)$$

is the general solution, where A and A' are (complex) constants. They are connected to the position $\zeta_0 = x_0 + iy_0$ and velocity $\dot{\zeta}_0 = v_x + iv_y$ at $t = 0$ by

$$A = \zeta_0 - i\frac{m\dot{\zeta}_0}{eB}; \quad A' = i\frac{m\dot{\zeta}_0}{eB}. \quad (2.86)$$

Consequently, the charge travels in a *circle* of radius

$$|A'| = \frac{m|\dot{\zeta}_0|}{eB} = \frac{mv_{\perp}}{eB},$$

with angular frequency eB/m . The coordinates of the center of this circle are given by the real and imaginary parts of A in (2.86). The speed $v_{\perp} = |v_x + iv_y|$ in the $x - y$ plane remains constant of course: from (2.85) and (2.86)

$$\dot{\zeta} = -i\frac{eB}{m}A'e^{-i\frac{eB}{m}t} = \dot{\zeta}_0 e^{-i\frac{eB}{m}t}$$

or

$$|\dot{\zeta}| = |\dot{\zeta}_0| = v_{\perp}, \quad \text{a constant.}$$

Thus, we confirm our previous picture of the motion in a constant magnetic field.

The Lagrange function (2.82) is of special interest in the case that it contains in addition a restoring potential $\frac{1}{2}\omega_0^2(|\zeta|^2 + z^2)$ that binds the charge to some center of force:

$$L = \frac{1}{2}mv^2 - \frac{1}{2}m\omega_0^2(x^2 + y^2 + z^2) + \frac{e}{2m}\mathbf{B} \cdot \mathbf{l}.$$

This is the *Lorentz oscillator* model of a light source in an external magnetic field. The equations of motion for a charge under such circumstances are

$$\ddot{\zeta} + \omega_0^2\zeta + i\frac{eB}{m}\dot{\zeta} = 0; \quad \ddot{z} + \omega_0^2z = 0. \quad (2.87)$$

We insert $\zeta \sim e^{-i\omega t}$ once more and find the allowed frequencies in the $x - y$ plane to be

$$\omega^2 - 2\left(\frac{eB}{2m}\right)\omega - \omega_0^2 = 0, \quad (2.88)$$

or

$$\begin{aligned} \omega_1 &= \Omega + \sqrt{\omega_0^2 + \Omega^2} \simeq \Omega + \omega_0 \\ \omega_2 &= \Omega - \sqrt{\omega_0^2 + \Omega^2} \simeq \Omega - \omega_0 \end{aligned} \quad (2.89)$$

where $\Omega = eB/2m$. The approximate expressions of ω_1 and ω_2 are valid if $\Omega \ll \omega_0$ (weak field case). The solutions of (2.87) are

$$x + iy = \zeta = Ae^{-i\omega_1 t} + A'e^{-i\omega_2 t}; \quad z = A''(\cos \omega_0 t + \alpha), \quad (2.90)$$

where the A 's and α are constants. Thus, the motion of the charge in space can be broken down into a linear oscillation along z of angular frequency ω_0 , and a *superposition* of circular motions $\zeta_1 = Ae^{-i\omega_1 t}$ and $\zeta_2 = A'e^{-i\omega_2 t}$ with angular frequencies ω_1 and ω_2 . For *weak* fields ($\Omega \ll \omega_0$) the motion in the $x - y$ plane is especially simple to describe. We insert the approximate values of ω_1 and ω_2 in the equation for $\zeta = x + iy$ and find

$$\zeta \simeq Ae^{-i(\omega_0 + \Omega)t} + A'e^{i(\omega_0 - \Omega)t} = \zeta' e^{-i\Omega t}, \quad (2.91)$$

where

$$\zeta' = Ae^{-i\omega_0 t} + A'e^{i\omega_0 t} \quad (2.92)$$

is the motion in the *absence* of the magnetic field. Thus, all a weak field does is to *rotate* the pattern of motion (2.92) around the direction of \mathbf{B} (in a negative sense for a positive charge, a positive sense for a negative charge) without changing the amplitudes A and A' . This is called *Larmor's theorem*, Ω the Larmor (angular) frequency. We discuss this theorem generally in a moment. But before doing so, let us trespass into the classical theory of radiation from moving charges³⁴ and ask, with

³⁴ L. Landau and E. Lifshitz, *The Classical Theory of Fields*, Addison-Wesley Press Inc., Cambridge, Mass.; 1951 (translated from the Russian by M. Hamermesh).

Lorentz, how the pattern of radiation from the oscillator (2.92) changes in an external magnetic field (classical theory of the *Zeeman effect*).

Now, the motion (2.92) consists of a superposition of two circular motions in the $x - y$ plane with the same angular frequency ω_0 . When viewed along the z -axis (\mathbf{B} points at us) these motions give rise to separate radiation components of angular frequency ω_0 , but which are circularly polarized in opposite senses and have different amplitudes. These two components add coherently and the resultant radiation is elliptically polarized in general. Actually, the radiation from a collection of such oscillators is unpolarized since there is no preferred direction. The magnetic field changes this picture. From (2.91) we see that the positive frequency component speeds up to $(\omega_0 + |\Omega|)$ ($\Omega = -|\Omega|$) in the magnetic field. Therefore, circularly polarized radiation of *two* frequencies is now expected along the z -axis:

$$\begin{aligned} \omega_0 + \left| \frac{eB}{2m} \right| & \text{ left circular polarized along } \mathbf{B} \\ \omega_0 - \left| \frac{eB}{2m} \right| & \text{ right circular polarized along } \mathbf{B}. \end{aligned}$$

No radiation with frequency ω_0 will be observed along the z axis since the z motion is not "visible" in this direction. However, radiation in the $x - y$ plane perpendicular to \mathbf{B} will contain three frequencies

$$\omega_0, \quad \omega_0 \pm \left| \frac{eB}{2m} \right|$$

all linearly polarized (the first along \mathbf{B} , the other two perpendicular to \mathbf{B}). These predictions are in full accord with experiment and constitute the *normal Zeeman effect* in atomic spectra. Actually, the normal Zeeman effect is less normal than the so-called *anomalous Zeeman effect* which does not show the simple separation of a spectral line into only three components. The explanation of the anomalous Zeeman effect lies outside the domain of classical mechanics.

(ii) Larmor's Theorem. We saw in the previous problem (2.91) in particular, that a weak magnetic field simply rotates the unperturbed pattern of motion around the direction of the field. One may obtain this result directly from the Lagrange function as we now show for any atomic system. Suppose an atom containing Z electrons is placed in an external magnetic field \mathbf{B} . Introduce cartesian coordinate axes with z pointing along B and the origin at the atomic nucleus. Then, if (x_k, y_k, z_k) locates the k^{th} electron,

$$L = \sum_k \frac{1}{2} m (\dot{x}_k^2 + \dot{y}_k^2 + \dot{z}_k^2) - V(x_1, y_1, z_1; x_2, y_2, z_2; \dots) - |\Omega| \sum_k l_k,$$

where

$$l_k = m(x_k \dot{y}_k - y_k \dot{x}_k)$$

is the angular momentum of the k^{th} electron along B ; V is the total electrostatic interaction potential. Now, introduce the *cylindrical polar coordinates* (r_k, φ_k, z_k) for the k^{th} electron [see Fig. 2.17].

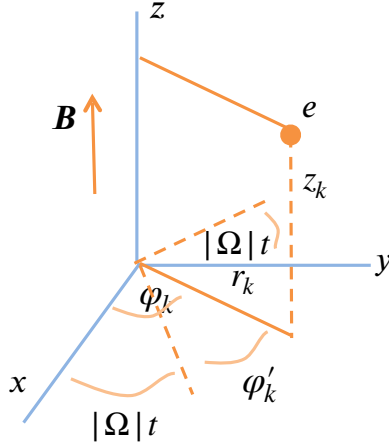


Figure 2.17: Cylindrical coordinate system with external magnetic field \mathbf{B} indicated.

Then, the expression for $\sum l_k = \sum_k m r_k^2 \dot{\varphi}_k$, so that

$$L = \sum_k \frac{1}{2} m (\dot{r}_k^2 + r_k^2 [(\dot{\varphi}_k - |\Omega|)^2 - \Omega^2] + \dot{z}_k^2) - V \quad (2.93)$$

after regrouping the terms involving the $\dot{\varphi}_k$. We now drop the term $\sim \Omega^2$ and introduce a moving coordinate system that rotates about z in Fig. 2.17 in the positive sense with angular frequency $|\Omega|$, then

$$\varphi_k = |\Omega|t + \varphi'_k \quad (2.94)$$

while r_k and z_k are not affected. Then, L reads (note that V does not depend on φ_k)

$$L = \sum_k \frac{1}{2} m (\dot{r}_k^2 + r_k^2 \dot{\varphi}'_k{}^2 + \dot{z}_k^2) - V$$

in the moving frame, which is identical with the Lagrange function in the fixed frame before the magnetic field was turned on. Therefore, the unperturbed pattern of motion of the electrons is rotated bodily by the magnetic field, so that their total angular momentum vector

$$\mathbf{I} = \sum_k (\mathbf{r}_k \times \mathbf{p}_k) \quad (2.95)$$

precesses about the direction of \mathbf{B} with angular frequency $|\Omega|$ (see Fig. 2.18).

It is clear that the result (2.91) is a special case of this theorem, since a rotation of coordinates about the z axis through the angle $|\Omega|t = -\Omega t$ is expressed by

$$x + iy = (x' \cos \omega t + y' \sin \omega t) + i(y' \cos \Omega t - x' \sin \Omega t) = (x' + iy') e^{-i\Omega t}. \quad (2.96)$$

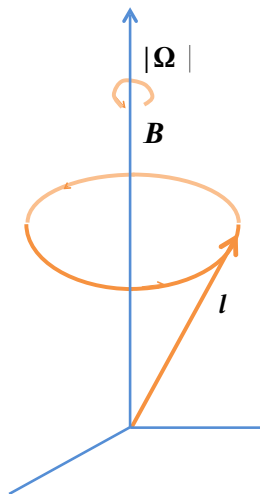


Figure 2.18: The angular momentum vector \mathbf{l} precesses about the direction of the field \mathbf{B} .

2-9 Non-Inertial Frames

We saw in the discussion of Larmor’s theorem that it can be advantageous to introduce moving frames of reference for certain problems. Such moving frames occur naturally in the description of the motion of a particle near the earth’s surface (which is rotating) and for the motion of rigid bodies (Chap. 3). We study their properties in this section. From the discussion in Chap. 1, Sec. 1-2., we know that a frame Σ' moving in an arbitrary manner with respect to an inertial frame Σ is *not* an inertial frame. Suppose however, that we deem a description of the motion referred to Σ' as convenient. Consider a particle P moving in Σ ; its position vector and velocity in Σ are \mathbf{r} and $\mathbf{v} = d\mathbf{r}/dt$, so that

$$L = \frac{1}{2}mv^2 - V(\mathbf{r}), \tag{2.97}$$

where V is the potential energy. Now suppose Σ' shares a common origin with Σ but rotates with angular velocity $\boldsymbol{\Omega}(t)$ with respect to Σ . The position vector of P as seen from Σ' and Σ is the same, $\mathbf{r}' = \mathbf{r}$; however, the velocities differ. If P has no velocity in Σ' , then it is still displaced to P' in time dt , where (see Fig. 2.19)

$$PP' = (\boldsymbol{\Omega} dt)(r \sin \vartheta) = |\boldsymbol{\Omega} \times \mathbf{r}|dt$$

and therefore moves with velocity $\boldsymbol{\Omega} \times \mathbf{r}$ in Σ . If P also has a velocity \mathbf{v}' in Σ' the total velocity as seen from Σ is

$$\mathbf{v} = \mathbf{v}' + (\boldsymbol{\Omega} \times \mathbf{r}'). \tag{2.98}$$

Now, the equations of motion in Σ' are still given by the Lagrange equations (1.39), since the action principle is independent of the frame of

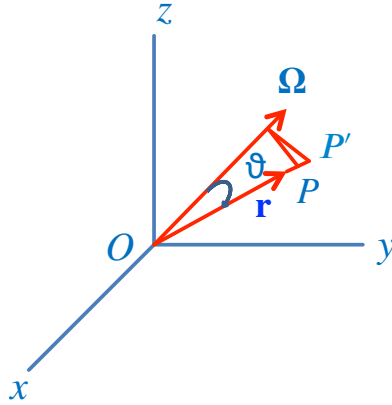


Figure 2.19: Coordinate vectors.

reference. However, L is not given by (2.97); we have to find it. But the value of L is known in the inertial frame from (2.97).

Therefore we only have to express this equation in the moving system: using (2.98) and the fact that $\mathbf{r} = \mathbf{r}'$; we have

$$L' = \frac{1}{2}mv'^2 - V(\mathbf{r}') + \boldsymbol{\Omega} \cdot \mathbf{l}' + \frac{1}{2}m(\mathbf{r}' \times \boldsymbol{\Omega})^2 \quad (2.99)$$

in the moving frame, where $\mathbf{l}' = \mathbf{r}' \times m\mathbf{v}'$. This is our second example of a Lagrange function that is not simply $T - V$. The modifications to L in the moving frame are seen to depend both on the angular velocity of the frame and the velocity (more precisely: angular momentum) of the particle in that frame.

The construction of the Lagrange equations requires the derivatives $\partial L' / \partial x'$, $\partial L' / \partial \dot{x}'$, etc. We can get these all at once by subjecting L' to a virtual displacement in the moving frame, $x' \rightarrow x' + \delta x'$, $\dot{x}' \rightarrow \dot{x}' + \delta \dot{x}'$, etc. Then

$$\begin{aligned} \delta L' &= \nabla L' \cdot \delta \mathbf{r}' + \nabla_{\mathbf{v}'} L' \cdot \mathbf{v}' \\ &= (-\nabla V + (\boldsymbol{\Omega} \times m\mathbf{v}') + \boldsymbol{\Omega} \times (m\mathbf{r}' \times \boldsymbol{\Omega})) \cdot \delta \mathbf{r}' \\ &\quad + (m\mathbf{v}' + (\boldsymbol{\Omega} \times m\mathbf{r}')) \cdot \delta \mathbf{r}', \end{aligned}$$

in an obvious notation. The partial derivatives $\nabla L'$ and $\nabla_{\mathbf{v}'} L'$ can be read off directly from this display, leading to the Lagrange equations of motion

$$\frac{d}{dt}(m\mathbf{v}') = \mathbf{F}' + 2(m\mathbf{v}' \times \boldsymbol{\Omega}) + \boldsymbol{\Omega} \times (m\mathbf{r}' \times \boldsymbol{\Omega}) + (m\mathbf{r}' \times \dot{\boldsymbol{\Omega}}) \quad (2.100)$$

in the rotating frame; $\mathbf{F}' = -\nabla V$ is the force on the particle in this frame. Equation (2.100) shows how Newton's second law of motion has to be amended *if we insist* on retaining the form $\dot{\mathbf{p}} = \mathbf{F}$ in a non-inertial frame also. We see that the actual force \mathbf{F}' has to be supplemented by *fictitious forces*

$$2(m\mathbf{v}' \times \boldsymbol{\Omega}), \quad \boldsymbol{\Omega} \times (m\mathbf{r}' \times \boldsymbol{\Omega}) \text{ and } (m\mathbf{r}' \times \dot{\boldsymbol{\Omega}}). \quad (2.101)$$

The first and second expressions are referred to as the *Coriolis*³⁵ and *centrifugal forces* respectively. The third one is only present if the rotation is non-uniform, and is known as the Euler force.³⁶ They are all proportional to the mass of the particle and betray their origin in the non-inertial nature of Σ' by depending on Ω . Much of the confusion caused by naming the last three members in (2.100) forces (without the qualifying term fictitious) can be avoided by recalling their origin and consequently the fact that they can be transformed away by returning to an inertial frame.

Returning to the Lagrange function L in the inertial frame, we know that the total energy $E = \frac{1}{2}mv^2 + V$ is conserved. The corresponding conservation law in the rotating frame is found by calculating the function H of (1.78):

$$\begin{aligned} H' &= (m\mathbf{v}' + \Omega \times m\mathbf{r}') \cdot \mathbf{v}' - L' \\ &= \frac{1}{2}m\mathbf{v}'^2 + V(\mathbf{r}') - \frac{1}{2}m(\mathbf{r}' \times \Omega)^2. \end{aligned} \tag{2.102}$$

H' will be conserved if L' does not have an explicit time-dependence, which means the rotation must be *uniform*. The rotation thus adds a term proportional to Ω^2 to the "energy" $\frac{1}{2}mv'^2 + V(\mathbf{r}')$ in the rotating frame. However, the latter quantity is not constant, only the sum H' . The term $-\frac{1}{2}m(\mathbf{r}' \times \Omega)^2$ is called the *centrifugal potential*, because it gives rise to the centrifugal force term in (2.78). The value of the constant H' is

$$H' = m\mathbf{v} \cdot \mathbf{v}' - L = E - \Omega \cdot \mathbf{l}$$

(using (2.98) and (2.97)), where $\mathbf{l} = \mathbf{r} \times m\mathbf{v}$ is the angular momentum of the particle in *either* frame, since $\mathbf{l} = \mathbf{l}'$.

(i) Again, let us look at the motion of a charged particle in a constant magnetic field. Let us view the motion of the charge in a *slowly* rotating reference frame. Then, this Lagrange function is given by

$$L' = \frac{1}{2}mv'^2 - V(\mathbf{r}') + \left(\frac{e}{2m}\mathbf{B} + \Omega\right) \cdot \mathbf{l}'$$

in this frame. The choice

$$\Omega = -\frac{e}{2m}\mathbf{B}$$

eliminates the magnetic field and gives an almost trivial proof of Larmor's theorem. Alternatively, we may interpret Larmor's theorem to be the consequence of choosing the rotational angular velocity of the moving frame so that the Coriolis force just cancels the contribution from the magnetic part of the Lorentz force (2.70).

(ii) Motion relative to a rotating earth. In Sec. 2.3 we considered the motion of a projectile near the surface of the earth. Our treatment involved the tacit assumption that axes attached to the earth form an inertial frame. This is not exactly true of course, due to the rotation of

³⁵ Introduced by G. Coriolis, J. de l'école Polytechnique, Cahier 24, 142 (1835).

³⁶ For the purists, we admit that for forces proportional to the mass of the particle, one cannot in fact distinguish between the "actual" and "fictitious" attributes insofar as the action of such forces can always be ascribed to the non-inertial nature of the coordinate system of the observer. In fact, this is one of the theses (*the principle of equivalence*) upon which Einstein's theory of gravitational attraction is based. Such matters lie outside the scope of these notes. The interested reader should consult L. Landau and E. Lifshitz, *The Classical Theory of Fields*, Addison-Wesley Press Inc., Cambridge, Mass. 1951, Chap. 11.

the earth about its own axis and its motion around the sun. Let us study the effect of this rotation.

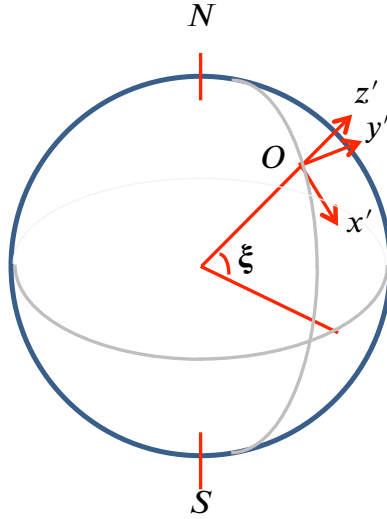


Figure 2.20: Coordinate axes x', y', z' for describing motion relative to the earth's surface.

Suppose $O_{x',y',z'}$ in (2.21) represents a rectangular frame of reference rigidly attached to the surface of the earth at O , see Fig. 2.20. We choose z' "vertical" and x' in the plane of the great circle passing through the poles. The point O is located by its angle of *latitude* ξ relative to the equatorial plane that lies at $\xi = 0$. Strictly speaking, the theory developed in this section does not apply to the $O_{x',y',z'}$. These axes have an additional acceleration due to the motion of the point O in space, and the Lagrange function (2.99) should be modified accordingly (see Prob. 2-15). However, this acceleration is proportional to Ω^2 , where Ω is the angular velocity of rotation of the earth and, together with the centrifugal potential, will be neglected in this discussion. Then, the Lagrange function describing the motion of a particle near the surface of a rotating earth is given by (2.99).

$$L' = \frac{1}{2}mv'^2 - V(\mathbf{r}') + \boldsymbol{\Omega} \cdot \mathbf{l}',$$

where $V(\mathbf{r}') = V(x', y', z')$ is the potential energy of the particle of mass m under study. Since $\boldsymbol{\Omega}$ lies along the North-South axis of the earth, we have

$$L' = \frac{1}{2}(\dot{x}'^2 + \dot{y}'^2 + \dot{z}'^2) - V(x', y', z') + \Omega_{x'}l_{x'} + \Omega_{z'}l_{z'}. \quad (2.103)$$

The rotation of the earth about its own axis is revealed by the last two terms; indeed, one can use this equation to demonstrate the earth's rotation. Consider a simple pendulum $O'P$ of length l , suspended directly above O from a point O' on the z' axis (Fig. 2.21). For small amplitude oscillations the motion of the pendulum is effectively restricted to the

plane $z' = l$ parallel to the $x' - y'$ plane. The restoring force components parallel to x' and y' are $-m\omega_0^2 x'$ and $-m\omega_0^2 y'$ respectively, where $\omega_0 = \sqrt{g/l}$ is the pendulum frequency relative to a fixed earth. Therefore, the potential energy in (2.103) is $V = \frac{1}{2}m\omega_0^2(x'^2 + y'^2)$ and L' becomes

$$L' = \frac{1}{2}m(\dot{x}'^2 + \dot{y}'^2) - \frac{1}{2}m\omega_0^2(x'^2 + y'^2) + \Omega_{z'}l_{z'}. \quad (2.104)$$

Introducing polar coordinates r', θ' in the $x' - y'$ plane, we find

$$\begin{aligned} L' &= \frac{1}{2}m(\dot{r}'^2 + r'^2\dot{\theta}'^2) - \frac{1}{2}m\omega_0^2 r'^2 + \Omega_{z'}mr'^2\dot{\theta}', \\ &\simeq \frac{1}{2}m\dot{r}'^2 + \frac{1}{2}mr'^2(\dot{\theta}' + \Omega_{z'})^2 - \frac{1}{2}m\omega_0 r'^2 \end{aligned}$$

to first order in $\Omega_{z'}$. The situation is the reverse of the case of a

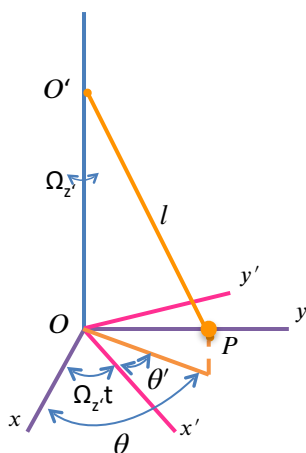


Figure 2.21: Foucault's pendulum.

charge in a magnetic field: here the effect of the motion of the earth is removed by going into an *inertial* frame: $\theta = \theta' + \Omega_{z'}t$. The analysis is in all respects similar to that leading up to (2.90); we simply give the answer

$$(x' + iy') = e^{-i\Omega_{z'}t}(x + iy),$$

where $x + iy$ gives the motion of the pendulum with respect to a fixed earth. Fig. 2.21 illustrates the motion.

The pendulum oscillates in the plane OPO' that retains its orientation in space, the "drift" of the pendulum away from a straight line in the $x' - y'$ plane being due, quite literally, to the earth turning out from under the pendulum. But, as observers on the earth, we turn with it and so record the path in the $x' - y'$ plane. Suppose then that the pendulum is pulled aside a distance a along the x' axis and released from rest. The motion we see as earth-dwellers is

$$x' + iy' = e^{-i\Omega_{z'}t} \left[a \cos \omega_0 t + \frac{a\Omega_{z'}}{\omega_0} \sin \omega_0 t \right],$$

since the initial velocity $\dot{x}' + i\dot{y}'$ is zero. The velocity at later times is

$$\dot{x}' + i\dot{y}' = -e^{-i\Omega_z t} a\omega_0 \left(1 - \frac{\Omega_z^2}{\omega_0^2}\right) \sin \omega_0 t.$$

The sine factor shows that the bob repeatedly comes to rest at times $\omega_0 t = 0, \pi, 2\pi$, etc., and is at the same distance $|x' + iy'| = a$ from the origin as initially at these times. How then does the projection of the bob on the $x' - y'$ plane move? The positions at which $\dot{x}' + i\dot{y}' = 0$ are clearly *cusps* of the path. The bob moves from one cusp to the next in a path that continually bends out of the $Ox'z'$ plane, the plane of oscillation rotating clockwise as seen from above (see Fig. 2.22). The cusp positions advance through an angle $2\pi\Omega_z/\omega_0 = 2\pi(\Omega/\omega_0) \cos \zeta$ along the circumference of a circle radius a per oscillation period of the pendulum. (The Kirchhoff Institute of Physics in Heidelberg has a beautiful example of such a pendulum on display). The device is known as *Foucault's pendulum* after G. Foucault, who used it in 1851 to demonstrate the rotation of the earth³⁷.

³⁷ It was fortunate that Foucault lived in the 19th century. Demonstration of such facts was not to be recommended during the 17th century (Galileo's century), being contrary to the theological beliefs of the time.

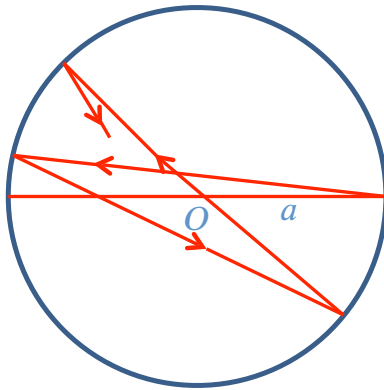


Figure 2.22: Projected path of the bob of Foucault's pendulum.

2-10 Systems of Interacting Particles

We finally consider a system consisting of N *interacting* particles. Such a system constitutes the fundamental many-body problem that is of primary interest in fields as far apart as celestial mechanics and atomic physics. As before, we assign cartesian coordinates x_k to each particle. The Lagrange function is then given by (1.37), i.e.

$$L = \sum_k \frac{1}{2} m_k \dot{x}_k^2 - V(x_1, x_2, \dots). \quad (2.105)$$

We will be primarily interested in systems where the potential field $V(x_1, x_2, \dots)$ is due to mutual interactions between the particles only. Such systems are termed *closed systems*. Notice in passing that the mere

introduction of a potential function $V(x_1, x_2, \dots)$ in the Lagrange function (2.105) implies that the change in position of any particle in the system *instantaneously* affects the motion of all other particles, i.e. the interaction between particles propagates with infinite velocity. Thus, (2.105) violates one of the basic tenets of the Special Theory of Relativity (see Chap. 6).

For N interacting particles, L in (2.105) leads to $3N$ coupled Lagrange equations of motion. To progress anywhere beyond the formality of just writing down this set of equations, one usually has to introduce specific additional assumptions about the many-body system. Such special many-body systems are the subject of the next two chapters. Nevertheless, we can make some general statements about the system described by (2.105) by examining the invariance properties of L itself.

We are assuming the system is a closed one. Therefore, $V(x_1, x_2, \dots)$ is independent of the coordinate origin chosen for the x_k , i.e. the system is translationally invariant. In symbols, this means for example that displacing the x coordinates of all particles by a common amount Δx along the x axis³⁸ does not change V :

$$\Delta V = \left(\frac{\partial V}{\partial x} + \frac{\partial V}{\partial x'} + \dots \right) \Delta x = 0, \quad (2.106)$$

[we have reverted to $(x_1, x_2, x_3) = (x, y, z)$, $(x_4, x_5, x_6) = (x', y', z')$, etc. for greater clarity]. On the other hand, the equations of motion satisfied by the x, x' , etc. are

$$\frac{\partial V}{\partial x} - \dot{p}_x = 0, \quad \frac{\partial V}{\partial x'} - \dot{p}_{x'} = 0, \quad \dots \quad (2.107)$$

Adding these equations together and using (2.106) we conclude that

$$\frac{d}{dt}(p_x + p_{x'} + \dots) = 0$$

or that the total linear momentum, $P_x = p_x + p_{x'} + \dots$ in the x direction is conserved. Likewise, the components P_y and P_z are conserved. Thus, the conservation of the total linear momentum reflects the *homogeneity* of the space in which the many-body system is moving.

If L is invariant under rotations we get another conservation law. A rotation of the coordinate axes around the z direction by a small angle $\Delta\varphi$ replaces the x and y coordinates of each particle by

$$x \rightarrow x + y\Delta\varphi; \quad y \rightarrow y - x\Delta\varphi,$$

[see (3.31)]. The invariance of $V(x_1, x_2, \dots)$ under such rotations means that the change

$$\Delta V = V(x + y\Delta\varphi, y - x\Delta\varphi, z; x' + y'\Delta\varphi, y' - x'\Delta\varphi, z'; \dots) - V(x, y, z; x', y', z'; \dots)$$

³⁸ Observe again that the translation Δx does not constitute a virtual displacement, see remarks after (1.75).

is zero, or that

$$\left(y \frac{\partial V}{\partial x} - x \frac{\partial V}{\partial y} + y' \frac{\partial V}{\partial x'} - x' \frac{\partial V}{\partial y'} + \dots\right) \Delta\varphi = 0. \quad (2.108)$$

Substituting for the space derivatives of V from (2.107) in this equation, we find that

$$\frac{d}{dt} \{ (xp_y - yp_x) + (x'p_{y'} - y'p_{x'}) + \dots \} = 0,$$

or that the total angular momentum along the z direction, $L_z = (xp_y - yp_x) + (x'p_{y'} - y'p_{x'}) + \dots$ is conserved. Invariance under rotations about x and y likewise show that L_x and L_y are conserved. Thus, the *isotropy* of space and angular momentum conservation go together. Finally, the lack of any explicit time dependence in L (homogeneity in time) shows that the total energy of the system is conserved.

Two further remarks about (2.105) are important. First, the lack of dependence of V on the coordinate origin can be made explicit by going over into center of mass and relative coordinates for the system. Changing notation once more, we define the position vectors $\mathbf{r}_1 = (x_1, x_2, x_3)$, $\mathbf{r}_2 = (x_4, x_5, x_6)$ for each particle as well as their center of mass coordinate

$$\mathbf{MR} = \sum_{i=1}^N m_i \mathbf{r}_i, \quad M = \sum_{i=1}^N m_i. \quad (2.109)$$

Now, the translational and rotational invariance of $V(x_1, x_2, \dots)$ shows it can only depend on the coordinate differences $(\mathbf{r}_i - \mathbf{r}_j)$, or only on the coordinate differences $(\mathbf{r}'_i - \mathbf{r}'_j)$, where the coordinate \mathbf{r}'_i are measured relative to the center-of-mass according to

$$\mathbf{r}'_i = \mathbf{r}_i - \mathbf{R}. \quad (2.110)$$

Introduction of these coordinates into L in (2.105) leads to a Lagrange function of the form

$$L = \sum_i \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 + \frac{1}{2} M \dot{\mathbf{R}}^2 - V(\mathbf{r}'_1, \mathbf{r}'_2, \dots) \quad (2.111)$$

The center-of-mass coordinate \mathbf{R} is explicitly cyclic; thus its canonical momentum

$$\mathbf{P} = \nabla_{\mathbf{R}} L = M \dot{\mathbf{R}}$$

is conserved. This is exactly the conservation of total momentum again as can be seen by differentiating the first member of (2.109) with respect to time to identify $M \dot{\mathbf{R}}$.

The other remark concerns the structure of $V(\mathbf{r}'_1, \mathbf{r}'_2, \dots)$ itself. As it stands in (2.111) V gives rise to many-body forces acting between particles (interaction between any pair of particles depends also on where

all the other particles are). This is not the case if V can be displayed as a sum of two-body potentials $V_{ij}(|\mathbf{r}'_i - \mathbf{r}'_j|)$:

$$V_{ij}(|\mathbf{r}'_i - \mathbf{r}'_j|) = \sum_{i < j} V_{ij}(|\mathbf{r}'_i - \mathbf{r}'_j|). \quad (2.112)$$

Then, the interaction between the i, j^{th} pair only depends on their relative separation, and is quite independent of where the remaining particles are. The forces acting in atomic systems are of this type for example.

The remaining task is to discuss the motion of a many-body system described by the Lagrange function (2.105). The mathematical complexities of such a program are enormous in general and one has to lean heavily on specific properties of the system as an aid to its solution. For example, in *rigid bodies* the interparticle forces are strong enough to effectively "freeze" the constituent particles into a rigid lattice. The interparticle spacing therefore does not change during the motion of such a system, so the potential energy-term in (2.105) remains constant and we can simply drop it. Thus

$$L = \sum_i \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 + \frac{1}{2} M \dot{\mathbf{R}}^2 \quad (2.113)$$

for a rigid body moving under no external forces. However, the conditions of rigidity introduce constraints between the coordinates r'_i , leaving the system with only six degrees of freedom (3 translational plus 3 rotational), instead of $3N$. Such systems are discussed in detail in the next chapter.

Relaxing the condition of perfect rigidity slightly allows the constituent particles of the system to indulge in small displacements about the equilibrium positions they held in the completely rigid body. Now, the potential term in (2.105) enters once more, but this complication is offset by the assumption of "small displacements". Perturbation methods can be applied to the system to study the oscillatory motion of all the \mathbf{r}'_i about their equilibrium values. A systematic study of such small oscillations will be found in Chapter 4.

Finally we may relax the rigidity conditions completely and "let the system take care of itself". This involves us in the full complications of the many-body problem as it presents itself in atomic and nuclear physics. Furthermore, the number of degrees of freedom become infinite in the limit of a large number of interacting particles, $N \rightarrow \infty$. The "particle character" of the system is effectively overwhelmed by their mere number in this limit and we find ourselves discussing a "fluid", i.e. the dynamics of a continuous (on a macroscopic) medium. The dynamics of continuous media is taken up briefly in Chapter 5.

Problems

2-1. Discuss the convergence of the integral expression given in (2.13) for the period of a one-dimensional oscillation. Evaluate T for the periodic time of a planet [see (2.25)] by the method of contour integration.

2-2 Obtain the path of a projectile moving in a constant gravitational field directly from Jacobi's principle, (1.99).

2-3 Obtain the differential equation (2.26) for the orbit in a central field directly from Jacobi's principle, (1.99).

2-4 Show that the true anomaly of a planet may be expressed as the series

$$\theta = nt + 2e \sin nt + \frac{5}{4}e^2 \sin 2nt + \dots$$

in terms of the mean anomaly nt (e is the eccentricity of the elliptic orbit).

2-5 Find the equation expressing the time t in terms of the angular position θ for hyperbolic motion of a planetary object under an inverse square force law.

2-6 Show that

$$\sqrt{\frac{\alpha}{2p^3}}t = \tan \frac{\theta}{2} + \frac{1}{3} \tan^3 \frac{\theta}{3}$$

for parabolic motion in the potential field (2.27); p is the distance from the focus to the vertex of the parabola.

2-7 Two space capsules are in common circular orbit of radius r_0 around the earth, and have an angular separation β . The rear space capsule is desirous of transmitting a *small* metal object to the lead capsule. In what direction and with what velocity must this object be ejected? In particular, explain what sort of orbits the object follows when ejected (i) tangentially forwards, (ii) tangentially backwards, to the circular orbit.

2-8 A space capsule is in a polar orbit around the earth that is a perfect circle. Why are retro-rockets necessary to bring the capsule down again? If the capsule has to be brought down so that it is moving tangential to the earth's surface at the equator, calculate the energy change, ΔE , that the rockets have to cause. Assume that the rockets fire when the capsule is a height h above one pole and that they fire instantaneously. Must the capsule lose or gain energy in order to return to the earth? Is the direction the rockets fire in important?

2-9 Investigate the motion of a satellite in the gravitational field of a deformed earth. Assume that the earth has the shape of an oblate spheroid, i.e. flattening at the poles and bulging at the equator. Calculate the gravitational potential of such a mass distribution assuming the deviation from a sphere to be small. Hence calculate approximately the

main effects of such deviations in the gravitational potential away from a central one on satellite motion. (A systematic way of handling such problems is discussed in Chapter 7.)

2-10 Show that the radius vector r to the particle P in Fig. 2.4 sweeps out equal areas in equal times in central motion.

2-11 Calculate the differential cross-section for the scattering of particles by a repulsive square well potential [replace $-V_0$ by $+V_0$ in (2.58)].

2-12 Obtain the relation (2.59) of the text by making use of conservation laws only.

2-13 Calculate the equation for the orbit and the differential scattering cross-section for particle motion in the central field

$$V(r) = \frac{\alpha}{r} + \frac{\beta}{r^2},$$

where α and β are constants. Do the relative and/or absolute signs of α and β matter? Comment.

2-14 Verify the statements made below (2.66) in the text by calculating explicitly the orbits *in space* for two particles m' and m'' interacting via an attractive inverse square force.

2-15 Construct the Lagrange function for a free particle moving relative to a frame of reference that performs an arbitrary motion.

Chapter 3 Rigid Body Dynamics

3-1 Introduction

We start our discussion of specialized many-body systems by considering the motion of a *rigid body*. By this term one understands a system of N interacting particles in which the relative separations of all particles remain constant even when the system is acted upon by external forces. The use of N discrete particles is no restriction on the validity of the results that we derive in the following sections. In the limit where the mass distribution can be regarded as continuous rather than discrete ($N \rightarrow \infty$ in such a way that the mass density ρ at every point in the body remains finite) we simply replace summations over all particles by integrals over the mass density, i.e. $\sum_i \dots m_i \rightarrow \int \dots d^3r$, where d^3r is the volume element at position \mathbf{r} in the body.

Perfectly rigid bodies do not exist in nature. However, the idealization we have just described is completely adequate for discussing the motion of bodies that are rigid enough so that small distortions induced by external forces do not matter. In the questions that relate to the motion of a body in space, the idealization of a perfectly rigid body will be adhered to. It is well to realise, however, that situations can arise in the laboratory in which the internal structure of the body becomes important. A simple illustration is provided by a rapidly rotating flywheel, which behaves like a perfectly rigid body until angular velocities are reached where the accelerations that are required to force each element of the flywheel to move in a circle cannot be provided by the internal interactions holding the various parts of the flywheel together and it simply flies apart³⁹.

The subject of rigid body dynamics is an extensive one and it is not the purpose of this chapter to provide a complete treatment of the field. Rather the aim is to illustrate the methods and the difficulties that arise where one uses Newton's equations of motion to study the motion of a rigid body. The reader who is interested in a much more detailed treatment of the subject would find the classic, if somewhat old-fashioned, treatment in Routh's *Rigid Body Dynamics*⁴⁰ delightful, and well worth the time spent in sorting out a notation that was in vogue at the turn of the century.

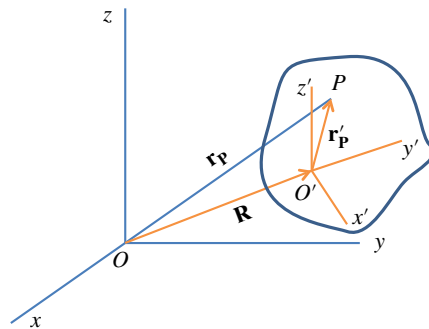
³⁹ This phenomenon actually happens in high speed centrifuges.

⁴⁰ E.J. Routh, *Elementary Rigid Dynamics*, Seventh Edition, Macmillan and Co., Ltd., London, 1905; *Advanced Rigid Dynamics*, Sixth Edition, MacMillan and Co., Ltd., London, 1905.

3-2 Frames of Reference

In order to describe the motion of a rigid body it is necessary to introduce a suitable set of axes to serve as a frame of reference from which to view the motion. Two systems of axes occur quite naturally in this subject. The first system is the *space-fixed* or laboratory system of axes. Call this set O_{xyz} ; it forms an inertial frame of reference. The directions of x, y, z are fixed in space, as is the origin O . Newton's laws apply in this frame. The second system of axes is the *body-fixed* system $O'_{x'y'z'}$ whose origin is attached to a fixed point in the body, and whose axes $x'y'z'$ are *frozen* into the rigid body and rotate when the body rotates⁴¹. The body-fixed system is *not* an inertial system since when the body accelerates it does also, and so Newton's laws as written in an inertial frame do not apply. Much of the complexity of describing the motion of a rigid body in fact arises just because, while quantities like the total angular momentum or kinetic energy of rotation of a rigid body assume very simple forms in the body-fixed system of axes, the equations of motion become much more complicated.

The origin O' of the body-fixed system can be attached to any point that moves with the rigid body. Most commonly however, O' is either a fixed point about which the rigid body is moving (hence O' is also fixed in space), or is taken at the center-of-mass of the rigid body. The two systems of axes are displayed in Fig. 3.1.



⁴¹ An olive with three toothpicks stuck into it to form an orthogonal set of axes, and frozen into an ice cube, brings this situation to mind very graphically.

Figure 3.1: Coordinates of body-fixed system relative to the space-fixed system.

Consider the particle P of mass m_0 located at \mathbf{r}_p with respect to O and \mathbf{r}'_p with respect to O' . The separation of O and O' is \mathbf{R} . Naturally, $\mathbf{r}_p, \mathbf{r}'_p$ and \mathbf{R} in general vary with time. An infinitesimal displacement of the particle P at \mathbf{r}_p can be accomplished by displacing O' by an amount $\delta\mathbf{R}$ with respect to O and then displacing P relative to O' by an amount $\delta\mathbf{r}'_p$, so that the displacement $\delta\mathbf{r}_p$ of the particle at P appears as

$$\delta\mathbf{r}_p = \delta\mathbf{R} + \delta\mathbf{r}'_p. \tag{3.1}$$

The result (3.1) is obvious, but not particularly useful until we use the fact that the particle at P forms part of a rigid body. For then it is clear

that the displacement $\delta\mathbf{r}'_p$ of the particle at P located with respect to O' can only come about from a change in orientation of $\mathbf{r}'_p = O'P$, since $|\mathbf{r}'_p| = \text{constant}$. Moreover, the distances between all particles are fixed. If we rotate the body-fixed axes $O'x'y'z'$ through an angle $\delta\phi$ about a direction \mathbf{n} passing through O' the displacement of any particle is $\delta\mathbf{r}'_p = \delta\phi\mathbf{n} \times \mathbf{r}'_p$ since all particles undergo a common angular displacement $\delta\phi\mathbf{n}$. Then we have $\delta\mathbf{r}_p = \delta\mathbf{R} + \delta\phi\mathbf{n} \times \mathbf{r}'_p$ and if this displacement occurs in time δt the velocity of the particle at P is given as

$$\mathbf{v}_p = \lim_{\delta t \rightarrow 0} \frac{\delta\mathbf{r}_p}{\delta t} = \mathbf{V} + \boldsymbol{\omega} \times \mathbf{r}'_p, \quad (3.2)$$

where $\mathbf{V} = \lim_{\delta t \rightarrow 0} \delta\mathbf{R}/\delta t$ is the velocity of O' as seen from O . The vector $\boldsymbol{\omega}$ is called the angular velocity of the rigid body. Obviously,

$$\boldsymbol{\omega} = \lim_{\delta t \rightarrow 0} \frac{\delta\phi}{\delta t} \mathbf{n}. \quad (3.3)$$

If we consider the point O' to be instantaneously at rest we see from (3.2) that the velocity of all particles is perpendicular to $\boldsymbol{\omega}$ at this instant. Furthermore, there exists a set of positions \mathbf{r}_p lying on a straight line passing through O' along which the particle velocities are zero at the instant considered (see Prob. 3-1). This line is called the instantaneous axis of rotation. We see then that the most general motion of a rigid body moving about a fixed point can be described as a pure rotation of the body about the instantaneous axis passing through that point (Euler's theorem). It is clear therefore that the *most general* displacement of a rigid body can be reduced to a translation followed by a rotation about some point.

We remark in passing that $\boldsymbol{\omega}$ in (3.3) is a vector by construction. Furthermore, two simultaneous infinitesimal rotations, $\delta\boldsymbol{\phi}_1 = \delta\phi_1\mathbf{n}_1$ and $\delta\boldsymbol{\phi}_2 = \delta\phi_2\mathbf{n}_2$ say, behave like vectors in the sense that the quantity $\delta\boldsymbol{\phi} = \delta\phi_1\mathbf{n}_1 + \delta\phi_2\mathbf{n}_2$, combined vectorially, represents the result of carrying out the infinitesimal rotations $\delta\boldsymbol{\phi}_1$ and $\delta\boldsymbol{\phi}_2$. For if \mathbf{r}'_p positions the particle P with respect to O' , then the first rotation moves P to $\mathbf{r}'_p + \delta\phi_1\mathbf{n}_1 \times \mathbf{r}'_p$; the second rotation moves P to

$$\begin{aligned} & \mathbf{r}'_p + \delta\phi_1\mathbf{n}_1 \times \mathbf{r}'_p + \delta\phi_2\mathbf{n}_2 \times (\mathbf{r}'_p + \delta\phi_1\mathbf{n}_1 \times \mathbf{r}'_p) \\ & \simeq \mathbf{r}'_p + (\delta\phi_1\mathbf{n}_1 + \delta\phi_2\mathbf{n}_2) \times \mathbf{r}'_p, \end{aligned}$$

since $\delta\phi_1$ and $\delta\phi_2$ are infinitesimal, which proves the assertion. However, the appearance of $\boldsymbol{\omega}$ in the combination $\boldsymbol{\omega} \times \mathbf{r}'_p$ in Eq. (3.2) demonstrates that it is not an ordinary vector, but rather a pseudo- or axial-vector that does not change sign under coordinate reflections as do "ordinary" vectors. This property of $\boldsymbol{\omega}$ guarantees that the vector \mathbf{v}_p in (3.2) will change sign properly under coordinate reflections.

3-3 The Inertia Tensor

Our next task is to construct the kinetic energy and the angular momentum of a moving rigid body and it is in so doing that the advantages of the body-fixed frame of reference will become apparent. Let us therefore examine the form of the kinetic energy T when looked at from the space- and body-fixed systems. In the space-fixed system we simply have $T = \sum_{p=1}^N \frac{1}{2} m_p v_p^2$; if we substitute for \mathbf{v}_p its equivalent in terms of \mathbf{V} and $\boldsymbol{\omega}$ we have

$$T = \sum_p \frac{1}{2} m_p v_p^2 = \frac{1}{2} \left(\sum_p m_p \right) V^2 + \sum_p \frac{1}{2} m_p (\boldsymbol{\omega} \times \mathbf{r}'_p)^2 [\boldsymbol{\omega} \times \left(\sum_p m_p \mathbf{r}'_p \right)] \cdot \mathbf{V}. \quad (3.4)$$

We have mentioned previously that O' is most commonly taken to be either a fixed point in space about which the body is rotating, or the center of mass of the body. In the former case, both the first and last terms in (3.4) are zero and we are left with the kinetic energy of rotational motion about O' :

$$T_{rot} = \sum_p \frac{1}{2} m_p (\boldsymbol{\omega} \times \mathbf{r}'_p)^2. \quad (3.5)$$

In the second case where O' is at the center of mass of the moving body, we have $\sum_p m_p \mathbf{r}'_p = 0$. Then, the last term in (3.4) still drops out while the first term $\frac{1}{2} (\sum_p m_p) V^2 = \frac{1}{2} M V^2$, where $M = \sum_p m_p$ is the total mass, simply adds the kinetic energy of translational motion to T_{rot} . In either case the interesting quantity is T_{rot} . Writing out the square of the vector product in (3.5) we have

$$\begin{aligned} T_{rot} &= \sum_p \frac{1}{2} m_p [\omega^2 (\mathbf{r}'_p \cdot \mathbf{r}'_p) - (\boldsymbol{\omega} \cdot \mathbf{r}'_p)^2] \\ &= \sum_{i,j=1}^3 \sum_{p=1}^N \frac{1}{2} m_p \omega_i (r_p^2 \delta_{i,j} - x_{p,i} x_{p,j}) \omega_j, \end{aligned}$$

if we introduce the components ω_i and $x_{p,i}$, $i = (1, 2, 3)$ of the vectors $\boldsymbol{\omega}$ and \mathbf{r}_p along the body-fixed system O'_{xyz} . Since the ω_i are the same for all particles they can be taken outside of the p summation in (3.6). Then, the set of quantities

$$I_{i,j} = \sum_p m_p (r_p^2 \delta_{i,j} - x_{p,i} x_{p,j}) \quad (3.6)$$

can be calculated once and for all for a given rigid body, and the kinetic energy of rotation becomes

$$T_{rot} = \sum_{i,j} \frac{1}{2} \omega_i I_{ij} \omega_j. \quad (3.7)$$

The quantity I_{ij} is called the inertia tensor. It is a quantity which characterizes the distribution of mass of a rigid body about the point O' rather

than the mass itself. We see from (3.7) that rigid bodies of the same total mass can possess quite different kinetic energies of rotational motion depending on how the mass in the body is distributed. Thus, the I_{ij} play the role of a set of inertial parameters which determine the energy content of the rotational motion in the same way that the total mass M determines the energy of translational motion of the body as a whole. From the definition (3.6) we see that I_{ij} is real and symmetric, $I_{ij} = I_{ji}$. We may therefore display the inertia tensor as a real 3×3 symmetric matrix (writing now x, y, z for x'_1, x'_2, x'_3),

$$I = (I_{ij}) = \begin{bmatrix} \sum_p m_p (y_p^2 + z_p^2) & -\sum m_p x_p y_p & -\sum m_p x_p z_p \\ -\sum m_p x_p y_p & \sum m_p (z_p^2 + x_p^2) & -\sum m_p y_p z_p \\ -\sum m_p x_p z_p & -\sum m_p y_p z_p & \sum m_p (x_p^2 + y_p^2) \end{bmatrix}.$$

It is customary to refer to the diagonal elements of this matrix as the moments of inertia, and the off-diagonal elements as the products of inertia of the system, all taken with respect to the prescribed set of axes O'_{xyz} .

Like all real, symmetric matrices, I can be transformed into a diagonal form

$$M = \begin{bmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{bmatrix}, \quad (3.8)$$

where the I_1, I_2, I_3 are called the principal moments of inertia. The inertia tensor only assumes the simple form (3.8) with respect to a special set of body-fixed axes, called the *principal axes* of the rigid body. Clearly, it is to our advantage to exploit this simplicity by expressing the kinetic energy T_{rot} in terms of the components of the angular velocity ω taken along the principal axes, rather than along an arbitrary set of body-fixed axes. Then, (3.7) reduces to a sum of squares

$$T_{rot} = \frac{1}{2} I_1 \omega_1^2 + \frac{1}{2} I_2 \omega_2^2 + \frac{1}{2} I_3 \omega_3^2, \quad (3.9)$$

where $\omega_1, \omega_2, \omega_3$ are now the components of ω taken along the principal axes. The reduction of the expression for T_{rot} in (3.7) to a sum of squares is a familiar procedure from our discussion of the theory of small oscillations. We need therefore only remark that if we were unlucky enough to choose our original set of body-fixed axes not to be simultaneously also principal axes, we would have to diagonalize the matrix I . This procedure obviously also provides us with the necessary information of how the principal axes are oriented with respect to the original body-fixed set we started out with. For simple rigid bodies, one can usually determine the principal axes by inspection, using the symmetries that the rigid body exhibits. The general case, however, requires that we construct the inertia tensor I_{ij} in a particular frame of reference and then carry out the

diagonalization to obtain the principal moments of inertia and principal axes.

The angular momentum of the rigid body about the point O' can also be expressed in terms of the inertia tensor. Let us again regard O' as either a fixed point in space about which the rigid body rotates, or its center-of-mass. The total angular momentum of the rigid body about O' is

$$\begin{aligned}
 L &= \sum_p (\mathbf{r}'_p \times m_p \mathbf{v}'_p) \\
 &= \sum_p m_p \mathbf{r}'_p \times (\boldsymbol{\omega} \times \mathbf{r}'_p) \\
 &= \sum_p m_p [(\mathbf{r}'_p \cdot \mathbf{r}'_p) \boldsymbol{\omega} - (\boldsymbol{\omega} \cdot \mathbf{r}'_p) \mathbf{r}'_p].
 \end{aligned} \tag{3.10}$$

Therefore, the component of \mathbf{L} along the body-fixed axis i is

$$\begin{aligned}
 L_i &= \sum_p m_p [r_p'^2 \omega_i - (\sum_j \omega_j x_{p,j}) x_{p,i}] \\
 &= \sum_j \sum_p m_p (r_p'^2 \delta_{i,j} - x_{p,i} x_{p,j}) \omega_j \\
 &= \sum_j I_{ij} \omega_j,
 \end{aligned} \tag{3.11}$$

or

$$L_1 = I_1 \omega_1, \quad L_2 = I_2 \omega_2, \quad L_3 = I_3 \omega_3 \tag{3.12}$$

if we project $\boldsymbol{\omega}$ along the principal axes of the rigid body.

Equations (3.9) and (3.12) provide us with simple expressions for the kinetic energy and angular momentum of a rotating rigid body in terms of its angular velocity components along the principal axes.

Our next task is to express $\boldsymbol{\omega}$ and \mathbf{L} in terms of a suitable set of coordinates (and their time derivatives) that describe the motion of the rigid body. We restrict the discussion to a rigid body moving about a fixed point in space. The first question that comes up is how many degrees of freedom are there in this case? This number is simple to determine. If we fix our attention on any point P of the rigid body moving about O' , then P is constrained to move on the surface of a sphere centered at O' and thus has two degrees of freedom. A second point Q of the body can then be uniquely located in terms of a rotation about the line $O'P$. The three points O', P, Q uniquely determine the position of the rigid body as a whole which therefore possesses three degrees of freedom. Therefore, three independent coordinates $q_k(t)$ are necessary to describe the motion. How shall we choose the $q_k(t)$? For a body moving about a fixed point O it is obviously convenient to choose O as a common origin for the space-fixed and body-fixed systems of axes as in Fig. 3.2.

Then, we see that by knowing the orientation of the axes $O_{x'y'z'}$, which move with the body, with respect to the space-fixed axes O_{xyz} at each instant of time, we are provided with a complete picture of the motion of the rigid body itself.

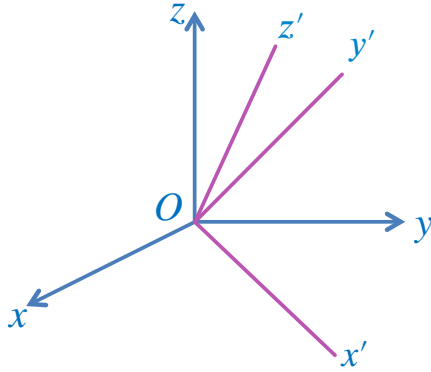


Figure 3.2: The origins of the body-fixed and space-fixed systems coincide, $O = O'$.

We have also remarked that the rotational motion about O can be described at each instant of time as the rotation of the body about some axis passing through O . We may therefore regard the final position of the axes $O_{x'y'z'}$ at time t as having been reached via a succession of infinitesimal rotations of $O_{x'y'z'}$ about O' that were dictated by Newton's equations of motion and the forces acting on the system. The entire problem of determining the motion of a rigid body moving about a fixed point therefore reduces to finding the finite rotation of the axes $O_{x'y'z'}$ that brings them from some initial position at $t = 0$ to occupy their final position in the moving body at time t . Two important questions immediately come up. How shall we parametrize the position of the rotated set $O_{x'y'z'}$ with respect to the space-fixed set O_{xyz} , and how do the components of vector quantities like the angular velocity and angular momentum $\boldsymbol{\omega}$ and \mathbf{L} relate to each other when viewed from the two sets of axes? To answer such questions, we have to study the transformation properties of the components of a vector when the coordinate axes, to which these components refer, are rotated.

3-4 Coordinate Transformations

We begin our discussion of coordinate transformations with a simple example. Suppose the body-fixed axes coincide with O_{xyz} at $t = 0$. Now, rotate them through an angle ϕ about the common z axis to take up the position $O_{x'y'z'}$ shown in Fig. 3.3. If \mathbf{Q} is any vector lying in the $x - y$ plane we see geometrically that the components of \mathbf{Q} along $O_{x'y'z'}$ can be expressed as linear combinations of the components of \mathbf{Q} along O_{xyz}

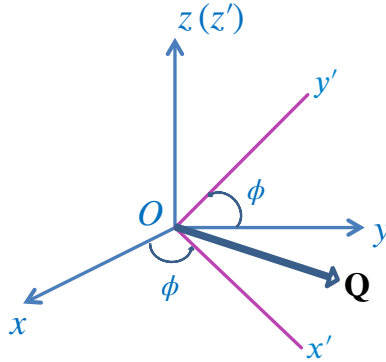


Figure 3.3: The body-fixed and space-fixed origins coincide ($O = O'$). In addition, the z axes are common $z = z'$. The x and y axes are rotated by ϕ to x' and y' .

according to

$$\begin{aligned} Q_{x'} &= \cos \phi Q_x + \sin \phi Q_y \\ Q_{y'} &= -\sin \phi Q_x + \cos \phi Q_y, \end{aligned}$$

where $Q_{x'}$, $Q_{y'}$ are the components of the *same* vector \mathbf{Q} in the rotated coordinate system. A more elegant way of writing (??) is to use a matrix notation and write

$$\begin{bmatrix} Q_{x'} \\ Q_{y'} \end{bmatrix} = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} Q_x \\ Q_y \end{bmatrix},$$

or, more briefly,

$$(\mathbf{Q}') = A(\mathbf{Q}), \tag{3.13}$$

where (\mathbf{Q}') is a column matrix of the components of \mathbf{Q} along the new axes $O_{x'y'z'}$, (\mathbf{Q}) is a column matrix of the components of \mathbf{Q} along the old axes O_{xyz} . A is a square matrix that depends upon the parameters that described the rotation (in this case the angle ϕ). We will speak of A as a rotation matrix inducing a rotation of the axes O_{xyz} to some final position $O_{x'y'z'}$. The length of \mathbf{Q} is unchanged by the rotation of axes "under it". We verify immediately from (??) that this means that

$$AA^T = A^T A = I, \quad \text{and } \text{Det } A = +1. \tag{3.14}$$

We do not consider the inversion of coordinates and so discard $\text{Det } A = -1$ where A^T again denotes the transpose of A , $\text{Det } A$ is the determinant of the matrix A and I is the unit matrix. We now turn this statement around and characterize *all linear transformation* of the type (3.13) where A is a real matrix satisfying the conditions (3.14) as a rotation of axes in the space of n dimensions, where n is the number of linearly independent components of \mathbf{Q} . For our purpose, $n = 2$ or 3. Such rotations are often also referred to as proper orthogonal

transformations, and the rotation matrix A that satisfies $AA^T = A^T A = I$ is called an orthogonal matrix. We note that the inverse rotation, or transformation, also exists because $\text{Det } A \neq 0$ so that A has an inverse $A^{-1} = A^T$. The inverse rotation is therefore

$$(\mathbf{Q}) = A^{-1}(\mathbf{Q}') = A^T(\mathbf{Q}').$$

It will prove convenient to indicate the directions of the rotated and unrotated sets of axes by introducing two sets of orthogonal unit vectors at the origin O . These sets of unit vectors lie along the O_{xyz} and $O_{x'y'z'}$ axes respectively. Let

$$\mathbf{e}_x, \quad \mathbf{e}_y, \quad \mathbf{e}_z \quad (3.15)$$

be unit vectors lying along the coordinate axes of O_{xyz} (the unprimed or unrotated axes) and let

$$\mathbf{f}_{x'}, \quad \mathbf{f}_{y'}, \quad \mathbf{f}_{z'} \quad (3.16)$$

be unit vectors lying along the coordinate axes of $O_{x'y'z'}$ (the primed or rotated axes). Then

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} \quad \text{and} \quad \mathbf{f}_\alpha \cdot \mathbf{f}_\beta = \delta_{\alpha\beta}, \quad (3.17)$$

where $i, j = (x, y, z)$ and $\alpha, \beta = (x', y', z')$ express the orthogonality and normalization properties of the \mathbf{e} 's and \mathbf{f} 's. The \mathbf{e}_i and \mathbf{f}_α constitute a complete set of unit vectors in three dimensions in the sense that an arbitrary vector \mathbf{Q} may be expressed in terms of either set, according to

$$\mathbf{Q} = \sum_i Q_i \mathbf{e}_i = \sum_\alpha Q'_\alpha \mathbf{f}_\alpha, \quad (3.18)$$

where the Q_i or the Q'_α are the components of \mathbf{Q} along the corresponding axes. Equation (3.18) immediately allows us to relate the components of the vector \mathbf{Q} along the primed axes with its components along the unprimed axes. Taking the scalar product of (3.18) with \mathbf{f}_α and using the orthogonality relations (3.17) we get

$$Q'_\alpha = \sum_i (\mathbf{f}_\alpha \cdot \mathbf{e}_i) Q_i = \sum_i A_{\alpha i} Q_i, \quad (3.19)$$

if we call

$$(\mathbf{f}_\alpha \cdot \mathbf{e}_i) = A_{\alpha i}. \quad (3.20)$$

Equation (3.19) is identical with the matrix equation (3.13) written out in detail. We also see from (3.20) that the matrix elements of A in (3.13) are given by the direction cosines of the rotated unit vectors \mathbf{f}_α with respect to the unrotated unit vectors \mathbf{e}_i . The unit vectors \mathbf{f}_α can therefore also be expressed in terms of the unrotated basis according to the equation

$$\mathbf{f}_\alpha = \sum_i A_{\alpha i} \mathbf{e}_i, \quad (3.21)$$

which follows directly from (3.20). We urge the reader to notice the subtle difference between this relation and (3.13). Equation (3.13) is a statement about the components of the *same* vector on both sides of the equation looked at from two different coordinate systems, whereas (3.21) merely relates two *different* vectors to each other, i.e. \mathbf{f}_α with the set \mathbf{e}_i . This difference is most easily appreciated if \mathbf{Q} happens to be the vector \mathbf{f}_α . Then, by contrast with (3.21), equation (3.13) reads

$$(\mathbf{f}'_\alpha) = A(\mathbf{f}_\alpha) \quad (3.22)$$

where (\mathbf{f}'_α) contains the components of \mathbf{f}_α along the rotated axes and (\mathbf{f}_α) contains the components of \mathbf{f}_α along the unrotated axes. Let us ask for the latter components. We invert (3.22) to find

$$(\mathbf{f}_\alpha) = A^T(\mathbf{f}'_\alpha)$$

or (writing $f_{\alpha,i}$ for the component of \mathbf{f}_α along the \mathbf{e}_i , etc.)

$$f_{\alpha,i} = \sum (A^T)_{i\beta} f'_{\alpha,\beta} = (A^T)_{i\alpha} = A_{\alpha i},$$

since $f'_{\alpha,\beta} = \delta_{\alpha\beta}$ because the \mathbf{f}_α happen to be unit vectors along the rotated axes. Thus, we have $\mathbf{f}_\alpha = \sum_i f_{\alpha,i} \mathbf{e}_i = \sum_i A_{\alpha i} \mathbf{e}_i$, which reconstructs (3.21).

Equation (3.13) or, equivalently (3.19), gives the law of change for the components of a vector under rotations of the coordinate system, or basis set of unit vectors \mathbf{e}_i to which the vector was referred. The components of a vector in a particular basis are said to form a *representation* of the vector in that basis. A change of basis according to (3.21) therefore produces a new representation of the same vector. In (3.18) the components Q_i and Q'_α provide an example of two representations of the vector \mathbf{Q} . Equation (3.21) is the connection between these two representations for a given transformation matrix A .

So far, we have talked only about vectors and their representations, a concept which distinguishes between the components of a vector and the vector "itself" so to speak. However, the same concepts extend to a wider class of entities which are called (linear) operators. Consider for example (3.11) relating the angular momentum \mathbf{L} and the angular velocity $\boldsymbol{\omega}$. Since \mathbf{L} and $\boldsymbol{\omega}$ are vectors, it is a useful point of view to regard the inertia tensor I as an *operator* which operates on $\boldsymbol{\omega}$ to produce a *new* vector \mathbf{L} ,

$$I\boldsymbol{\omega} = \mathbf{L}. \quad (3.23)$$

In a representation with respect to a particular basis \mathbf{e}_i , say, $\mathbf{L} = \sum_i \mathbf{e}_i L_i$, $\boldsymbol{\omega} = \sum_j \mathbf{e}_j \omega_j$, (3.23) becomes

$$I \sum_j \mathbf{e}_j \omega_j = \sum_i \mathbf{e}_i L_i,$$

or

$$\sum_j (\mathbf{e}_i \cdot I\mathbf{e}_j)\omega_j = L_i.$$

If we call

$$(\mathbf{e}_i \cdot I\mathbf{e}_j) = I_{ij}, \quad (3.24)$$

we regain (3.11). The quantity $(\mathbf{e}_i \cdot I\mathbf{e}_j)$ does not depend on the vector $\boldsymbol{\omega}$ but only on the operator I and the particular choice of basis vectors \mathbf{e}_i that has been made. The collection of quantities I_{ij} in (3.24) is called a *representation* of the operator I in the *basis* defined by the \mathbf{e}_i . We now ask how this representation changes when the basis is changed. From (3.21) and the definition (3.24) the representation of I in a different basis set \mathbf{f}_α is:

$$\begin{aligned} I'_{\alpha\beta} &= (\mathbf{f}_\alpha \cdot I\mathbf{f}_\beta) = \sum_{i,j} A_{\alpha i} (\mathbf{e}_i \cdot I\mathbf{e}_j) A_{\beta j} \\ &= \sum_{i,j} A_{\alpha i} I_{ij} A_{\beta j}^T, \end{aligned} \quad (3.25)$$

or

$$I' = AIA^T = AIA^{-1} \quad (3.26)$$

in a matrix notation.

In line with the interpretation of I as an operator, another interpretation to (3.13) can be given by regarding the matrix A as an *operator* which acts of \mathbf{Q} and changes it into a new vector \mathbf{q} ,

$$\mathbf{Q} \rightarrow \mathbf{q} = A\mathbf{Q} \quad (3.27)$$

without changing the basis. This is called the active interpretation of the operator A as opposed to the passive interpretation in (3.13) where A changed the basis, but left the vector alone. Equation (3.27) is still a rotation. The effect of A has simply been to point \mathbf{Q} in a new direction without changing its length.

We can also express (3.27) as a relation between components of the vectors \mathbf{q} and \mathbf{Q} referred to the same basis, i.e.

$$\sum_i q_i \mathbf{e}_i = \sum_j A_{ij} \mathbf{e}_j Q_j$$

or

$$q_i = \sum_j A_{ij} Q_j, \quad A_{ij} = (\mathbf{e}_i \cdot A\mathbf{e}_j). \quad (3.28)$$

For a *given* rotation matrix A (3.28) and (3.13) are the same, i.e. the components q_i of the vector \mathbf{q} in the old basis are identical with the components Q'_k of the vector \mathbf{Q} in the *new* basis. The point is that the components of \mathbf{Q} in some basis only depend on the relative orientation of \mathbf{Q} to that basis. We are therefore at liberty to rotate either the *basis* "under the vector", or rotate the vector "over the basis" *in the opposite*

sense and still end up with the same components of the vector that did not move (that moved) with respect to the basis that moved (that did not move). We will adopt either point of view according to our needs. The important link to bear in mind is that A always rotates the vector in the opposite sense to that in which it rotates the basis. Fig. 3.4 will be helpful to the reader who is unaccustomed to the mental gymnastics that are necessary to translate easily from the one point of view to the other.

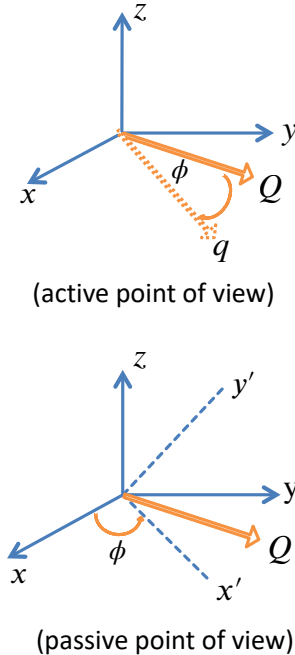


Figure 3.4: Upper panel: Active point of view. $A(\phi)$ rotates \mathbf{Q} closer to the x axis, Lower panel: Passive point of view. $A(\phi)$ rotates the x axis closer to the vector \mathbf{Q} .

One important property of rotations is their *group* property, i.e. the combination of two successive rotations represented by matrices A and B can be described as a single rotation represented by a matrix $C = BA$, where the "law of combination" is matrix multiplication. The proof is simple. If the matrix A rotates O_{xyz} with unit vectors \mathbf{e}_i , to occupy the orientation $O_{x'y'z'}$ with unit vectors \mathbf{f}_α , and B rotates $O_{x'y'z'}$ to occupy the orientation $O_{x''y''z''}$ with unit vectors \mathbf{g}_μ , then we see that

$$\mathbf{g}_\mu = \sum_\alpha B_{\mu\alpha} \mathbf{f}_\alpha = \sum_{\alpha,i} B_{\mu\alpha} A_{\alpha i} \mathbf{e}_i,$$

upon applying (3.21) twice, or that

$$\mathbf{g}_\mu = \sum_i C_{\mu i} \mathbf{e}_i, \quad C_{\mu i} = \sum_\alpha B_{\mu\alpha} A_{\alpha i}, \quad (3.29)$$

so that the matrix C is just the matrix product of A and B in a specified order. The rotational operation is in general a non-commutative one; rotations induced by BA and AB are in general distinct from each other.

3-5 The Euler Angles

We are now in a better position to answer the question of how to parametrize the orientation of the body-fixed axes $O_{x'y'z'}$ with respect to the space-fixed system O_{xyz} , i.e. how to choose the coordinates $q_l(t)$ that describe the motion.

Equation (3.21) connecting the unit vectors along O_{xyz} and $O_{x'y'z'}$ suggests one possibility, that of using the matrix elements $A_{\alpha i}$ as coordinates specifying the orientation of the body-fixed axes. This simply amounts to specifying the direction cosines of each of the \mathbf{f}_α with respect to the fixed axes O_{xyz} . But this gives us 9 parameters for a system having only 3 degrees of freedom. The point is of course that the $A_{\alpha i}$ are not all linearly independent. The condition $AA^T = 1$ in (3.14) provides for 6 linearly dependent relations between the $A_{\alpha i}$, effectively reducing the number of independent parameters to 3. It would obviously be much simpler therefore to choose a set of independent parameters that automatically satisfy these conditions. One such set of parameters is provided by Euler's three angles (ϕ, θ, ψ) which we now proceed to define. The convention followed by Edmonds⁴² will be used in defining ϕ , θ and ψ . However, the reader should beware of the fact that there is no general agreement in the literature of rigid body dynamics (or of quantum mechanics!) on a "best" definition of these angles. In Edmonds' convention the three angles ϕ, θ, ψ have the following significance: initially the body-fixed axes $O_{x'y'z'}$ coincide with the space-fixed axes O_{xyz} . Then the following rotations are performed on the body-fixed axes about the indicated directions (see Fig. 3.5):

⁴² A.R. Edmonds, *Angular Momentum in Quantum Mechanics*, Princeton University Press, Princeton, New Jersey, 1957.

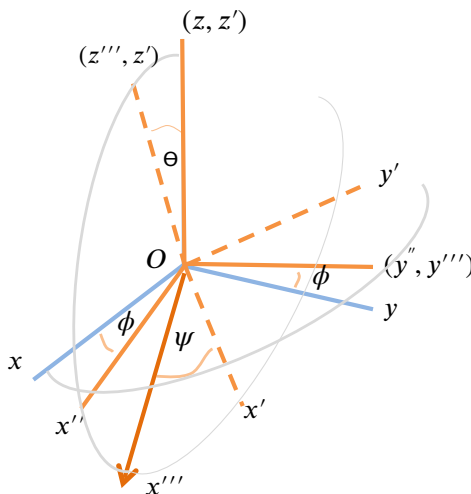


Figure 3.5: Euler angles.

- (i) a rotation ϕ about the z axis, $0 \leq \phi \leq 2\pi$
- (ii) a rotation θ about the *new* y axis (y''), $0 \leq \theta \leq \pi$

(iii) a rotation ψ about the *new* z axis (z'), $0 \leq \psi \leq 2\pi$.

This set of rotations, performed in the indicated manner, brings the body-fixed system to rest in the final orientation $O_{x'y'z'}$ shown in Fig. 3.5. The sense of rotation is such that the angles ϕ, θ, ψ are considered to be positive when the rotation would cause a right handed screw to advance along the axis of rotation.

Each rotation through an Euler angle has a rotation matrix A associated with it. We write

$A_z(\phi)$ for the rotation ϕ about z
 $A_{y''}(\theta)$ for the rotation θ about y''
 $A_{z'}(\psi)$ for the rotation ψ about z' .

From the group property of rotations displayed in (3.29) we know that we can combine these three rotations into a single rotation induced by the matrix

$$R(\phi\theta\psi) = A_{z'}(\psi)A_{y''}(\theta)A_z(\phi). \quad (3.30)$$

$R(\phi\theta\psi)$ is the matrix which will rotate the body-fixed axes directly over to their final orientation shown in Fig. 3.5. The individual rotation matrices in (3.30) are given by

$$A_z(\phi) = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ for a rotation } \phi \text{ about the } z \text{ axis,} \quad (3.31)$$

$$A_{y''} = \begin{pmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{pmatrix} \text{ for a rotation } \theta \text{ about the } y'' \text{ axis,} \quad (3.32)$$

$$A_{z'} = \begin{pmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ for a rotation } \psi \text{ about the } z' \text{ axis.} \quad (3.33)$$

The combined rotation matrix is obtained by multiplying these matrices together in the specified order (3.30). The result is

$$R(\phi\theta\psi) = \begin{pmatrix} \cos \psi \cos \theta \cos \phi - \sin \psi \sin \phi & \cos \psi \cos \theta \sin \phi + \sin \psi \cos \phi & -\cos \psi \sin \theta \\ -\sin \psi \cos \theta \cos \phi - \cos \psi \sin \phi & -\sin \psi \cos \theta \sin \phi + \cos \psi \cos \phi & \sin \psi \sin \theta \\ \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \end{pmatrix}. \quad (3.34)$$

3-6 Infinitesimal Rotations and Euler Angles

The three Euler angles $\theta_\lambda = (\phi, \theta, \psi)$, considered as functions of time, serve as a very convenient set of generalized coordinates $q_k(t)$ to describe the motion of a rigid body. We now ask for expressions for the angular

velocity and angular momentum of a rigid body in these coordinates. To find the angular velocity, consider an infinitesimal rotation $\delta\varphi$ of the rigid body away from its position occupied at time t and given by Euler angles $\vartheta^\lambda(t)$. Writing $\delta\vartheta^\lambda = (\delta\phi, \delta\theta, \delta\psi)$ for the increments in the three Euler angles, we can make up this rotation as follows:

$$\delta\varphi = \sum_{\lambda} \mathbf{g}_{\lambda} \delta\vartheta^{\lambda}, \quad (3.35)$$

where the $\mathbf{g}_{\lambda} \delta\vartheta^{\lambda}$, $\lambda = 1, 2, 3$ represent infinitesimal rotations about the three axes of rotation for the Euler angles. From Fig. 3.5 the unit vectors along these axes are

$$\begin{aligned} \mathbf{g}_{\phi} &= \mathbf{e}_z, & \text{the } z \text{ axis,} \\ \mathbf{g}_{\theta} &= (\mathbf{e}_z \times \mathbf{f}_{z'}) \frac{1}{\sin\theta}, & \text{the new } y \text{ axis } (y''), \\ \mathbf{g}_{\psi} &= \mathbf{f}_{z'}, & \text{the new } z \text{ axis } (z'). \end{aligned} \quad (3.36)$$

Notice that the unit vectors \mathbf{g}_{λ} are *not* orthogonal. Correspondingly, we must distinguish between contravariant and covariant components of a vector⁴³ along these unit vectors. We see that the $\delta\vartheta^{\lambda}$ are contravariant components of $\delta\varphi$. The magnitude of the infinitesimal rotation $\delta\phi$ defines a metric tensor for the $\delta\vartheta^{\lambda}$ "space" according to

$$(\delta\varphi)^2 = \sum_{\lambda, \mu} g_{\lambda\mu} \delta\vartheta^{\lambda} \delta\vartheta^{\mu}, \quad (3.37)$$

where $g_{\lambda\mu}$ is the metric tensor. From (3.35) and (3.36) we find that

$$(\delta\varphi)^2 = (\delta\phi)^2 + (\delta\theta)^2 + (\delta\psi)^2 + 2\cos\theta \delta\phi \delta\psi,$$

since only \mathbf{g}_{ϕ} and \mathbf{g}_{ψ} are not orthogonal. Therefore

$$(g_{\lambda\mu}) = \begin{pmatrix} 1 & 0 & \cos\theta \\ 0 & 1 & 0 \\ \cos\theta & 0 & 1 \end{pmatrix}, \quad \text{Det}(g_{\lambda\mu}) = \sin^2\theta, \quad (3.38)$$

where $\text{Det}(g_{\lambda\mu})$ is the determinant of $g_{\lambda\mu}$ considered as a matrix. To complete the picture we also construct the inverse of $g_{\lambda\mu}$. This is

$$(g^{\lambda\mu}) = \frac{1}{\sin^2\theta} \begin{pmatrix} 1 & 0 & -\cos\theta \\ 0 & \sin^2\theta & 0 \\ -\cos\theta & 0 & 1 \end{pmatrix}, \quad (3.39)$$

provided that $\text{Det}(g_{\lambda\mu}) = \sin^2\theta \neq 0$, and allows us to pass from the covariant components $\delta\vartheta_{\mu}$ to the contravariant components $\delta\vartheta^{\lambda} = \sum_{\mu} g^{\lambda\mu} \delta\vartheta_{\mu}$ in the usual manner. We note in passing that the singularity in $g^{\lambda\mu}$ at $\theta = 0$ is related to the fact that the Euler angles are not everywhere unique.

⁴³ See for example B. Spain, *Tensor Calculus*, Oliver and Boyd, Edinburgh, 1953.

We may also project $\delta\boldsymbol{\varphi}$ along the space-fixed or body-fixed axes. We have

$$\sum \mathbf{e}_i \delta\varphi_i = \sum_{\lambda} \mathbf{g}_{\lambda} \delta\vartheta^{\lambda} = \sum_{\alpha} \mathbf{f}_{\alpha} \delta\varphi'_{\alpha}, \quad (3.40)$$

where $\delta\varphi_i = (\delta\varphi_x, \delta\varphi_y, \delta\varphi_z)$ and $\delta\varphi'_{\alpha} = (\delta\varphi'_{x'}, \delta\varphi'_{y'}, \delta\varphi'_{z'})$ are the components of $\delta\boldsymbol{\varphi}$ along the space-fixed and body-fixed axes respectively. The transformation matrices from the set $\delta\vartheta^{\lambda}$ to the sets $\delta\varphi_i$ or $\delta\varphi'_{\alpha}$ are given by⁴⁴

$$(P_{i\lambda}) = (\mathbf{e}_i \cdot \mathbf{g}_{\lambda}), \quad (Q_{\alpha\lambda}) = (\mathbf{f}_{\alpha} \cdot \mathbf{g}_{\lambda}), \quad (3.41)$$

so that we can write

$$\delta\varphi_i = \sum_{\lambda} P_{i\lambda} \delta\vartheta^{\lambda} \quad (3.42)$$

$$\delta\varphi'_{\alpha} = \sum_{\lambda} Q_{\alpha\lambda} \delta\vartheta^{\lambda}. \quad (3.43)$$

It is clear that P and Q are *not* rotation matrices in the sense of (3.19), since $P^T \neq P^{-1}$. However, the inverse transformations

$$\delta\vartheta^{\lambda} = \sum_i (P^{-1})_{\lambda i} \delta\varphi_i = \sum_{\alpha} (Q^{-1})_{\lambda\alpha} \delta\varphi'_{\alpha} \quad (3.44)$$

are still possible if the matrices P and Q are non-singular.

To construct the matrices P and Q we require the projections of the \mathbf{g}_{λ} along the space-fixed unit vectors \mathbf{e}_i , and the body-fixed unit vectors \mathbf{f}_{α} . These projections are easy to compute. We observe from Fig. 3.5 that (θ, ϕ) are the polar angles for the direction of $\mathbf{f}_{z'}$, with respect to the space-fixed axes, while $(\theta, \pi - \psi)$ are the polar angles for the direction of \mathbf{e}_z with respect to the body-fixed axes. Therefore,

$$\mathbf{g}_{\phi} = \mathbf{e}_z = (-\sin\theta \cos\psi)\mathbf{f}_{x'} + (\sin\theta \sin\psi)\mathbf{f}_{y'} + (\cos\theta)\mathbf{f}_{z'} \quad (3.45)$$

$$\mathbf{g}_{\psi} = \mathbf{f}_{z'} = (\sin\theta \cos\phi)\mathbf{e}_x + (\sin\theta \sin\phi)\mathbf{e}_y + (\cos\theta)\mathbf{e}_z, \quad (3.46)$$

and, combining these,

$$\begin{aligned} \mathbf{g}_{\theta} &= \frac{1}{\sin\theta} (\mathbf{e}_z \times \mathbf{f}_{z'}) = -\sin\phi \mathbf{e}_x + \cos\phi \mathbf{e}_y \\ &= \sin\psi \mathbf{f}_{x'} + \cos\psi \mathbf{f}_{y'}. \end{aligned} \quad (3.47)$$

We use the above relations to express the transformation matrices, (3.41), in terms of Euler angles and find

$$P = \begin{pmatrix} 0 & -\sin\phi & \cos\phi \sin\theta \\ 0 & \cos\phi & \sin\phi \sin\theta \\ 1 & 0 & \cos\theta \end{pmatrix}; \quad Q = \begin{pmatrix} -\sin\theta \cos\psi & \sin\psi & 0 \\ \sin\theta \sin\psi & \cos\psi & 0 \\ \cos\theta & 0 & 1 \end{pmatrix}. \quad (3.48)$$

Clearly, $\text{Det } P = \text{Det } Q = -\sin\theta$, so that the transformations (3.42) and (3.43) are non-singular for $\sin\theta \neq 0$.

⁴⁴ The matrices P and Q were introduced by H.B.G. Casimir in his doctoral dissertation, *Rotation of a Rigid body in Quantum Mechanics*, Wolter's Uitgevers-Maatschappij, Groningen, 1931.

The components of the angular velocity vector $\boldsymbol{\omega} = \lim_{\delta t \rightarrow 0} \delta \boldsymbol{\phi} / \delta t$ follow immediately from (3.42) and (3.43). We have

$$\omega_i = \lim_{\delta t \rightarrow 0} \frac{\delta \phi_i}{\delta t} = \sum_{\lambda} P_{i\lambda} \dot{\theta}^\lambda \quad (3.49)$$

for the space-fixed components of $\boldsymbol{\omega}$, and

$$\omega'_\alpha = \lim_{\delta t \rightarrow 0} \frac{\delta \phi'_\alpha}{\delta t} = \sum_{\lambda} Q_{\alpha\lambda} \dot{\theta}^\lambda \quad (3.50)$$

for the body-fixed components of $\boldsymbol{\omega}$. Written out as matrix equations, (3.49) and (3.50) read

$$\begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}_{\text{space}} = \begin{pmatrix} 0 & -\sin \phi & \cos \phi \sin \theta \\ 0 & \cos \phi & \sin \phi \sin \theta \\ 1 & 0 & \cos \theta \end{pmatrix} \begin{pmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{pmatrix} \quad (3.51)$$

and

$$\begin{pmatrix} \omega_{x'} \\ \omega_{y'} \\ \omega_{z'} \end{pmatrix}_{\text{body}} = \begin{pmatrix} -\sin \theta \cos \psi & -\sin \psi & 0 \\ \sin \theta \sin \psi & \cos \psi & 0 \\ \cos \theta & 0 & 1 \end{pmatrix} \begin{pmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{pmatrix}. \quad (3.52)$$

We are now in a position to write down the angular velocity and angular momentum of a rigid body in terms of the Euler angles and their time derivatives. For convenience we choose the body-fixed axes $O_{x'y'z'}$ as principal axes of the rigid body in question and revert to our previous notation of $\omega_1, \omega_2, \omega_3$, etc. for the components of a vector along these axes.

The components of angular velocity along the principal axes are obtained from (3.52). The components of angular momentum along the principal axes are given by the relation (3.12), viz. $L_i = I_i \omega_i$, $i = 1, 2, 3$. We thus have

$$\begin{aligned} \omega_1 &= -\dot{\phi} \sin \theta \cos \psi + \dot{\theta} \sin \psi = L_1 / I_1 \\ \omega_2 &= \dot{\phi} \sin \theta \sin \psi + \dot{\theta} \cos \psi = L_2 / I_2 \\ \omega_3 &= \dot{\phi} \cos \theta + \dot{\psi} = L_3 / I_3. \end{aligned} \quad (3.53)$$

The kinetic energy is given by (3.7) with the angular velocity components given by (3.53). We write out the answer for the particular case where $I_1 = I_2 \neq I_3$: Such rigid bodies are called symmetric tops and will be of interest to us later on. For them,

$$T = \frac{1}{2} I_1 (\dot{\phi}^2 \sin^2 \theta + \dot{\theta}^2) + \frac{1}{2} I_3 (\dot{\phi} \cos \theta + \dot{\psi})^2. \quad (3.54)$$

We also observe from (3.7) and (3.12) that the kinetic energy may be written as the scalar product

$$2T = \sum_i \omega_i L_i = (\boldsymbol{\omega} \cdot \mathbf{L}). \quad (3.55)$$

Writing out this scalar product in terms of the components of $\boldsymbol{\omega}$ and \mathbf{L} along the triad of unit vectors \mathbf{g}_λ produces

$$2T = \sum_{\lambda} \dot{\theta}^\lambda L_\lambda = \text{scalar}. \quad (3.56)$$

If we multiply both sides of this equation by an infinitesimal time interval δt we get

$$2T\delta t = \sum_{\lambda} \delta\theta^\lambda L_\lambda, \quad (3.57)$$

where $\delta\theta^\lambda$ are the increments in the Euler angles θ^λ .

The L_λ are in fact the momentum variables canonical to the Euler angles θ^λ in the sense described by (1.72) of Chapter 1. From (3.54) the value of $2T\delta t$ is

$$\begin{aligned} 2T\delta t = & [I_1\dot{\phi}\sin^2\theta + I_3(\dot{\phi}\cos\theta + \dot{\psi})\cos\theta]\delta\phi \\ & + (I_1\dot{\theta})\delta\theta + [I_3(\dot{\phi}\cos\theta + \dot{\psi})]\delta\psi. \end{aligned} \quad (3.58)$$

Comparison of the right-hand members of (3.57) and (3.58) yields the L_λ :

$$\begin{aligned} L_\phi &= I_1\dot{\phi}\sin^2\theta + I_3(\dot{\phi}\cos\theta + \dot{\psi})\cos\theta \\ L_\theta &= I_1\dot{\theta} \\ L_\psi &= I_3(\dot{\phi}\cos\theta + \dot{\psi}). \end{aligned} \quad (3.59)$$

Since the L_λ appear as the components of a covariant vector in (3.57) (as they must), we know that the expressions (3.59) are also the physical components of the angular momentum \mathbf{L} along the unit vectors \mathbf{g}_λ . By contrast the covariant components of \mathbf{L} along these directions are $L^\lambda = \sum_{\mu} g^{\lambda\mu} L_\mu$, or

$$L^\phi = I_1\dot{\phi}; \quad L^\theta = I_1\dot{\theta}; \quad L^\psi = I_3\dot{\psi} + (I_3 - I_1)\dot{\psi}\cos\theta \quad (3.60)$$

using an obvious notation. Equation (3.60) answers the question that must have occurred to the reader by now, viz. what meaning can one attach to a quantity *moment of inertia* \times *rate of change of an Euler angle*?

We are now in possession of all of the ingredients to construct the Lagrange function of a rigid body with $I_1 = I_2 \neq I_3$ moving about a fixed point. If the forces are derivable from a potential function $V(\theta^\lambda)$, we have

$$L = \frac{1}{2}I_1(\dot{\phi}^2\sin^2\theta + \dot{\theta}^2) + \frac{1}{2}I_3(\dot{\phi}\cos\theta + \dot{\psi})^2 - V(\phi, \theta, \psi) \quad (3.61)$$

as a standard form for the Lagrange function.

Let us first verify that L_λ in (3.59) are indeed the canonical momenta.

We have

$$\begin{aligned}\frac{\partial L}{\partial \dot{\phi}} &= I_1 \dot{\phi} \sin^2 \theta + I_3 (\dot{\phi} \cos \theta + \dot{\psi}) \cos \theta = L_{\dot{\phi}} \\ \frac{\partial L}{\partial \dot{\theta}} &= I_1 \dot{\theta} = L_{\dot{\theta}} \\ \frac{\partial L}{\partial \dot{\psi}} &= I_3 (\dot{\phi} \cos \theta + \dot{\psi}) = L_{\dot{\psi}},\end{aligned}\tag{3.62}$$

which proves the assertion. The Lagrange function L may be used to study the motion of a rigid body about a fixed point once the form of $V(\theta^\lambda)$ is known.

3.7 Applications

The applications of the Lagrange function in (3.61) to the motion of rigid bodies with an axis of symmetry are manifold. However, we will restrict the discussion to two specific examples. As a first application, look at the motion of an ordinary top spinning on a rough surface so that the peg of the top does not wander around. Alternatively, imagine the peg of the top to be attached to a fixed point by a "universal joint", so that the top can move freely about the point of attachment. In either event we are dealing with the action of a rigid body about a fixed point, and if the motion is due to conservative forces only, (3.61) for the Lagrange function will suffice to study the motion.

Anyone who has watched an ordinary top in motion must have been struck by how peacefully the top moves: spinning very rapidly around its symmetry axis, while turning as a whole rather slowly about a vertical axis with the symmetry axis held at an angle to the vertical. This motion is referred to as a steady, or regular precession and is the easiest type of motion to discuss analytically. Often such more complicated motions occur. The most common of these is produced if the top is set spinning rapidly about its symmetry axis and then set down gently on its peg with this axis inclined to the vertical. Initially the top starts to fall under the action of gravity but, remarkably, catches itself and climbs back to (almost) its initial inclination to the vertical while performing a precessional motion about the vertical at the same time. This is called pseudo-regular precession since it differs from the pure precession by the tops performing an additional nodding (nutational) motion during precession. The ability of such spinning bodies to stabilize themselves against the force of gravity has fascinated scientists and mystics alike throughout the history of mechanics. It will be an interesting task to understand the reasons behind these rather surprising motions in terms of the dynamics governing the motion of a rigid body.

We take an ordinary top spinning on a horizontal surface as a prototype example of a rigid body moving about a fixed point. Thus, we

assume either that the surface is rough enough to prevent the peg of the top from wandering around on it, or that the peg has been attached to a fixed point by some device (like a universal joint) that does not interfere with the orientation of the top in any way. We shall also assume that the top is symmetric, i.e. it possesses an axis of symmetry which passes through its center of gravity G and the point of support O . This means that OG is a principal axis of inertia, that *any* pair of orthogonal axes in the plane normal to OG will serve as principal axes, and that the moments of inertia about these two axes are equal. Therefore, in treating the motion of a symmetrical body we are at liberty to choose axes that are only partially "frozen" into the body, without destroying the validity of relations like (3.7) and (3.12). In particular, this means that the pair of principal axes normal to OG need not rotate around OG with the top.

The top is positioned in space by rotating it through Euler angles ϕ, θ and ψ relative to the space-fixed axes O_{xyz} as prescribed by (3.30). Fig. 3.6 shows the final position of the top after such a rotation has been performed. In the present context the angles θ and ϕ give the polar angles of the symmetry axis OG , while ψ measures the amount of rotation of the top around OG . The unit vectors $\mathbf{g}_\phi, \mathbf{g}_\theta$ and \mathbf{g}_ψ that are associated with these angles are also shown. When the top moves the unit vectors \mathbf{g}_θ and \mathbf{g}_ψ move with it, while \mathbf{g}_ϕ always maintains its direction along O_z which we choose as the vertical direction. Changes in ϕ describe the precession, θ the nutation (or nodding) of the top, while ψ measures the rotation about OG .

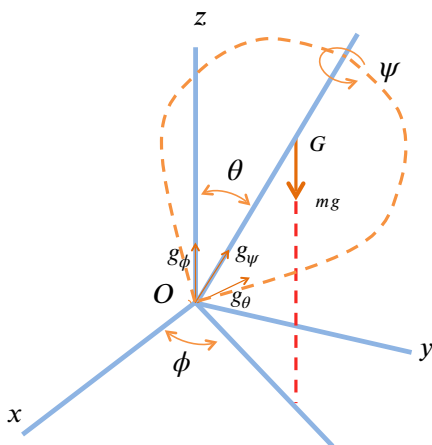


Figure 3.6: Rotation of the top with space-fixed axes.

Now, we must choose our principal axes at O : $\mathbf{g}_\psi = \overrightarrow{OG}$ is already a principal axis. Also we have remarked earlier that any axis in the plane normal to OG is a principal axis. Therefore, \mathbf{g}_θ also lies along a principal axis. The remaining unit vector

$$\mathbf{f} = (\mathbf{g}_\psi \cos \theta - \mathbf{g}_\phi) \operatorname{cosec} \theta$$

perpendicular to \mathbf{g}_θ and \mathbf{g}_ϕ and in the plane zOG is our other principal axis. Notice that the triad of partially frozen unit vectors $(\mathbf{f}, \mathbf{g}_\theta, \mathbf{g}_\psi)$ differs from the body-fixed triad $(\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3)$ in that it ignores the rotation ψ about \mathbf{g}_ψ . Therefore, the components of the angular velocity and angular momentum along *these* axes follow from (3.53) on setting $\psi = 0$ (but *not* $\dot{\psi} = 0$!) in those equations:

$$\begin{aligned}\omega_1 &= -\dot{\phi} \sin \theta = L_1/A \quad \text{along } \mathbf{f} \\ \omega_2 &= \dot{\theta} = L_2/A \quad \text{along } \mathbf{g}_\theta \\ \omega_3 &= \dot{\phi} \cos \theta + \dot{\psi} = L_3/C \quad \text{along } \mathbf{g}_\psi,\end{aligned}\tag{3.63}$$

if we denote the equal moments of inertia at O by $I_1 = I_2 = A$ and the unequal moment about OG by C .

Call $OG = h$; then, if the top moves in a constant gravitational field g that acts downward along $-z$, the potential energy is $V = mgh \cos \theta$, $m =$ mass of the top. Then, (3.61) reads

$$L = \frac{1}{2}A(\dot{\phi}^2 \sin^2 \theta + \dot{\theta}^2) + \frac{1}{2}C(\dot{\phi} \cos \theta + \dot{\psi})^2 - mgh \cos \theta.\tag{3.64}$$

We observe immediately that both angles ϕ and ψ are absent from L since the interaction energy with gravity is insensitive to changes in these angles. Thus, ϕ and ψ are cyclic and the corresponding canonical momenta are constants of motion. These momenta are given in the first and last lines of (3.62). Also, L does not depend on time explicitly. Therefore, the total energy is conserved during the motion. Notice, that by just examining the invariance properties of L we have discovered three constants of motion. They are

$$A\dot{\phi} \sin^2 \theta + C(\dot{\phi} \cos \theta + \dot{\psi}) \cos \theta = M\tag{3.65}$$

$$C(\dot{\phi} \cos \theta + \dot{\psi}) = K\tag{3.66}$$

$$\frac{1}{2}A(\dot{\phi}^2 \sin^2 \theta + \dot{\theta}^2) + \frac{1}{2}C(\dot{\phi} \cos \theta + \dot{\psi})^2 + mgh \cos \theta = E,\tag{3.67}$$

calling E the total energy and M and K the constant angular momentum projections along space axis O_z and symmetry axis OG respectively.

The above set of equations determines the top motion completely in terms of these constants, which are fixed by the initial conditions⁴⁵.

Let us examine their contents. Equation (3.66) insists that the angular momentum $L_3 = C\omega_3$ and therefore the angular velocity $\omega_3 = (\dot{\phi} \cos \theta + \dot{\psi})$ of the top along the symmetry axis is constant. Let us call this constant n , so that

$$n = \dot{\phi} \cos \theta + \dot{\psi} = K/C\tag{3.68}$$

can be used interchangeably with the constant angular momentum K if desired. The quantity n is called the *spin* of the top. Notice that it

⁴⁵ As an aside, notice that precisely the same set of constants E , M and K would be subject to quantization conditions if the top were considered to be a quantum mechanical system.

has contributions from both $\dot{\phi}$ and $\dot{\psi}$. Obviously, it will prove useful to eliminate the combination $\dot{\phi} \cos \theta + \dot{\psi}$ in equations (3.65) through (3.67); we find

$$Cn = K; \quad \frac{d\phi}{dt} = \frac{M - K \cos \theta}{A \sin^2 \theta}, \quad (3.69)$$

which combines with the energy (3.67) to give

$$\frac{1}{2}A\left(\frac{d\theta}{dt}\right)^2 = \left(E - \frac{K^2}{2C} - mgh \cos \theta\right) - \frac{1}{2A}\left(\frac{M - K \cos \theta}{\sin \theta}\right)^2. \quad (3.70)$$

The second of (3.69) and (3.70) express the precession $\dot{\phi}$ and nutation $\dot{\theta}$ in terms of the angular inclination of the top's symmetry axis; we have written these angular velocities as $d\phi/dt$ and $d\theta/dt$ respectively to emphasize the fact that the problem of finding $\phi(t)$ and $\theta(t)$ has been "reduced to quadratures" which is a formal way of saying that you know the answer if you can do the integrals. In fact (3.69) and (3.70) can be integrated in terms of elliptic functions⁴⁶, but the answer one gets is not particularly illuminating. Instead of carrying through the solution in general let us examine these equations for the particular example where the top starts off with an initial spin n about the symmetry axis that is inclined at an angle α to the vertical. The initial conditions for this motion are

$$\begin{aligned} \phi = 0, \quad \theta = \alpha, \quad \psi = 0 \\ \omega_1 = 0, \quad \omega_2 = 0, \quad \omega_3 = n \end{aligned} \quad (3.71)$$

at $t = 0$. The constants of motion therefore take on the values

$$K = Cn, \quad M = Cn \cos \alpha, \quad E = \frac{1}{2}Cn^2 + mgh \cos \alpha.$$

Entering this information into (3.69) and (3.70) we find that

$$\frac{d\phi}{dt} = n \frac{C}{A} \left(\frac{\cos \alpha - \cos \theta}{1 - \cos^2 \theta} \right) \quad (3.72)$$

and

$$\begin{aligned} \left(A \sin \theta \frac{d\theta}{dt}\right)^2 &= (\cos \alpha - \cos \theta) \\ &\times [2Amgh(1 - \cos^2 \theta) - C^2n^2(\cos \alpha - \cos \theta)] \end{aligned} \quad (3.73)$$

after a slight rearrangement of terms. Let us supplement these equations with the Lagrange equation for θ ; this is

$$A\ddot{\theta} = \frac{\partial L}{\partial \theta} = \sin \theta [A\dot{\phi}^2 \cos \theta - Cn\dot{\phi} + mgh], \quad (3.74)$$

where we calculate $\partial L/\partial \theta$ from (3.64). Alternatively, we could obtain this relation by differentiating (3.73) with respect to time. We require

⁴⁶ A. Sommerfeld, *Lectures on Theoretical Physics, Vol 1, Mechanics*, Academic Bress Inc., New York, 1952.

it to establish the initial motion of the top. Since $\dot{\phi} = 0$ at $t = 0$ when $\theta = \alpha$ we see that $\ddot{\theta}(t = 0) = (mgh \sin \alpha)/A$ is positive ($0 < \alpha < \pi$), i.e. initially the top begins to fall to the ground as one might intuitively expect it to do. However, this downward motion of the symmetry axis ceases when the nutational angular velocity $\dot{\theta}$ vanishes. Equation (3.73) provides us with information on this aspect of the motion: $\dot{\theta}$ vanishes when $\cos \theta = \omega$ reaches one of the roots of the cubic equation

$$(\omega_0 - \omega)[2Amgh(1 - \omega^2) - C^2n^2(\omega_0 - \omega)] = 0$$

$$\omega = \cos \theta, \quad \omega_0 = \cos \alpha, \quad |\omega_0| < 1.$$

The first factor $(\omega_0 - \omega) = 0$ just assures us that the axis will return to its initial inclination $\omega_0 = \cos \alpha$ during the subsequent motion. The other orientation of the symmetry axis for which $\dot{\theta} = 0$ is supplied by the real roots of the quadratic

$$f(\omega) = 2Amgh(1 - \omega^2) - C^2n^2(\omega_0 - \omega) = 0 \quad (3.75)$$

that satisfy $-1 < \omega < 1$ (θ must be real). We observe that $f(-1)$ is negative, while both $f(\omega_0)$ and $f(+1)$ are positive. Hence, $f(\omega)$ has one real root between -1 and ω_0 . Therefore, the symmetry axis of a top started off with the initial conditions (3.71) will swing between the angle α and the larger angle β that satisfies $f(\cos \beta) = 0$. If the initial spin is large we have

$$\cos \beta - \cos \alpha \simeq -2 \frac{Amgh}{(Cn)^2} \sin^2 \alpha. \quad (3.76)$$

Thus, by making the initial spin n (more correctly the angular momentum $K = Cn$) large enough we can effectively suppress the nutational motion of the top. Thus, θ is forced to vary over the small angular range $\alpha \leq \theta \leq \beta$ and we can therefore replace θ by its average value over this interval. From (3.76) we find that this angular interval is

$$\delta\alpha = (\beta - \alpha) \simeq 2 \frac{Amgh}{(Cn)^2} \sin \alpha$$

approximately, so that we can replace θ by the average value

$$\langle \theta \rangle = \alpha + \langle \delta\alpha \rangle \simeq \alpha + \frac{Amgh}{(Cn)^2} \sin \alpha \quad (3.77)$$

in our equations. Thus, the average precession rate is given by (3.72) with θ replaced by $\langle \theta \rangle$. We get

$$\langle \dot{\phi} \rangle \simeq \frac{Cn}{A} \frac{\langle \delta\alpha \rangle}{\sin \alpha} = \frac{mgh}{Cn} \quad (3.78)$$

if Cn is large. This equation shows that the precession about the vertical axis is in a positive sense under these circumstances and independent of the angle of inclination α .

Our picture of the top motion is now complete: the top precesses slowly around the vertical with an average precession given by (3.78) while spinning around its symmetry axis which nods up and down between the angles α and $\beta = \alpha + \delta\alpha$. The motion is perhaps best visualized by following the path that the extremity of the symmetry axis OG would trace out on a sphere centered at O . This path will be bounded by two circles on the sphere corresponding to the limiting orientations $\theta = \alpha$ and $\theta = \beta$ at OG . The exact nature of this path depends on the starting conditions. For the particular starting conditions we have been discussing it is clear from (3.72) that the precession $\dot{\phi}$ vanishes whenever the symmetry axis returns to its initial orientation $\theta = \alpha$ (where $\dot{\theta} = 0$ also). The symmetry axis therefore periodically comes to rest momentarily in this position. The path that its extremity traces out on a sphere centered at O is characterized by cusps at the points $\theta = \alpha, \dot{\phi}, \dot{\theta} = 0$ as shown in Fig. 3.7.

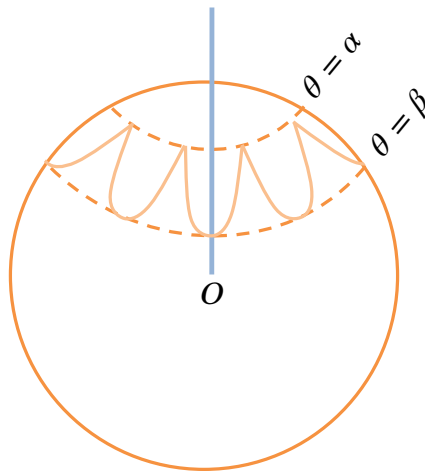


Figure 3.7: Cuspidal motion of top.

We can probe the problem for more detail by asking for example what the nature of the nutational motion is like. To investigate this point, we appeal to the equation of motion (3.74) for θ , set $\theta = \alpha + \varphi$, and use the unaveraged version of (3.78) with $\langle \delta\alpha \rangle$ replaced by φ for the precession, i.e.

$$\dot{\phi} = n \frac{C}{A} \frac{\varphi}{\sin \alpha}.$$

We substitute this value of $\dot{\phi}$ in (3.74) and find to first order in the small angle φ that

$$\ddot{\varphi} \simeq -\left(\frac{C}{A}n\right)^2\varphi + \frac{mgh}{A} \sin \alpha,$$

which has the solution

$$\varphi(t) = \frac{Amgh}{(Cn)^2} \sin \alpha \left[1 - \cos\left(\frac{C}{A}nt\right)\right], \quad (3.79)$$

if we incorporate the boundary conditions that φ must lie between O and the maximum value $\delta\alpha = \left[\frac{2Amgh}{C^2n^2}\right] \sin\alpha$. The nutation of the top is therefore a pure harmonic oscillation in the angle θ about the average orientation, (3.77), with the period

$$T_n = \frac{2\pi A}{n C}.$$

If we compare this with the period of average precession,

$$T_p = \frac{2\pi}{\langle\dot{\phi}\rangle} = 2\pi \frac{Cn}{mgh},$$

we find

$$\frac{T_n}{T_p} = \frac{Amgh}{(Cn)^2} \ll 1, \quad (3.80)$$

which is a precise physical statement of when our approximate treatment is valid: the precessional motion must be much slower than the nutational motion. Equation (3.80) also identifies $(Amgh)/(Cn)^2$ as the appropriate expansion parameter for this case. We note in passing that the two periods in (3.80) are incommensurate in general.

Equation (3.79) also allows us to compute the instantaneous precession to the same order of approximation. We have

$$\dot{\phi} \simeq \langle\dot{\phi}\rangle \left[1 - \cos\left(\frac{C}{A}nt\right)\right]$$

from (3.72), and integration with respect to time gives us ϕ :

$$\phi(t) = \langle\dot{\phi}\rangle t - \frac{Amgh}{(Cn)^2} \sin\left(\frac{C}{A}nt\right), \quad (3.81)$$

showing typical "drift" and oscillatory terms. Equations (3.79) and (3.81) specify the angles θ and ϕ as a function of time; we can get $\psi(t)$ from (3.68). Hence, our approximate solution is complete.

The motion we have just been discussing is called *pseudoregular* precession of a top because it differs from the true regular precession ($\theta = \alpha, \dot{\theta} = 0$) by terms of order T_n/T_p in the angular velocity of nutation, $\dot{\theta}$. Consequently, these small nutations are not visible to the casual observer and the top *appears* to perform a regular precession.

These observations naturally lead us to ask for the conditions under which a true regular precession is possible. Again, the equation of motion (3.74) provides the answer: for regular precession, both $\dot{\theta}$ and $\ddot{\theta}$ must vanish. This means $\partial L/\partial\theta = 0$, or that

$$\sin\alpha (Ap^2 \cos\alpha - Cnp + mgh) = 0 \quad (3.82)$$

for a regular precession with $\theta = \alpha$. We use p to denote the constant value of $\dot{\phi}$ that is allowed. If we exclude the special case $\sin\alpha = 0$, there are two allowed values of $\dot{\phi} = p$ in regular precession,

$$p = \left(Cn \pm \sqrt{(Cn)^2 - 4Amgh \cos\alpha}\right) / 2A \cos\alpha.$$

We notice that regular precession is possible only if $(Cn)^2 > 4Amgh \cos \alpha$; there is a minimum spin angular momentum required to stabilize the top.

Let us suppose again that Cn is large. Then, we find two physically distinct roots for p that can be classified as a fast and slow precession, i.e.

$$p \simeq \begin{cases} \frac{C}{A \cos \alpha} n, & \text{fast precession} \\ \frac{mgh}{Cn}, & \text{slow precession.} \end{cases} \quad (3.83)$$

Notice that both these precessions are in the same direction, and that the slow one coincides with the average precession of our first top, as given by (3.78). So the top we started off with initial conditions (3.71) nearly performs a regular precession. If we had provided for an initial precessional velocity $p = mgh/Cn$, it would have done so exactly. Clearly, other varieties of starting conditions that favor other types of motion that can be analysed in the same way as the examples we have discussed.

We have discussed the ordinary top in some detail because it forms a prototype example of how the motion of a rigid body may be investigated. We have also used Lagrange equations (and their consequences) throughout the discussion in order to stress once more the power and uniformity of such an approach. But having once obtained the answer it is always essential to try to understand the reasons behind our results. Perhaps, the most surprising aspect of the top motion is the way in which it sidesteps gravity instead of falling down in both its regular and pseudo-regular modes of precession. The reason for this lies *not* in the conservation of angular momentum (for there are torques present) but rather in the law of its change, as the following simple argument shows. We start by borrowing (1.14) from Chapter 1:

$$\frac{d\mathbf{L}}{dt} = \mathbf{N}; \quad d\mathbf{L} = \mathbf{N}dt, \quad (3.84)$$

where \mathbf{N} is the total torque that is responsible for changes in the angular momentum vector \mathbf{L} . From Fig. 3.6 we see that the torque \mathbf{N} for our ordinary top about the peg O is provided by the gravitational couple $\mathbf{N} = mgh \sin \theta \mathbf{g}_\theta$ that is always normal to the plane OzG . Also, in steady precession we have the condition that $\dot{\theta} = 0$: hence the only change in \mathbf{L} that can be brought about by \mathbf{N} is one of direction (the length of \mathbf{L} cannot change since it lies in the OzG plane and we know its projections along Oz and OG must be constant). The state of affairs is depicted in Fig. 3.8, which is the "skeleton" of Fig. 3.6. The angular momentum components are drawn in to show how \mathbf{L} is made up of the components Cn and $A\dot{\phi} \sin \alpha$ along and perpendicular to the symmetry axis of the top; α is the value of θ in steady precession as before and γ is the angle between the angular momentum \mathbf{L} and the symmetry axis. Since this entire

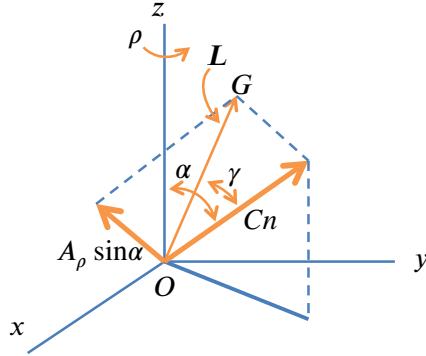


Figure 3.8: Directional change of angular momentum.

pattern precesses around the Oz axis with constant angular velocity p , we have immediately that $|d\mathbf{L}| = |\mathbf{L}| \sin(\alpha - \gamma) p dt$ and $d\mathbf{L}$ points along \mathbf{g}_θ . But

$$\begin{aligned} |\mathbf{L}| \sin(\alpha - \gamma) &= (|\mathbf{L}| \cos \gamma) \sin \alpha - (|\mathbf{L}| \sin \gamma) \cos \alpha \\ &= Cn \sin \alpha - Ap \sin \alpha \cos \alpha \end{aligned}$$

from the geometry of Fig. 3.8 so we have finally, using (3.84),

$$(Cn - Ap \cos \alpha) \sin \alpha p dt = mgh \sin \alpha dt,$$

which is the same condition as (3.82).

The magnitude and sign (does the angular momentum axis lie above or below the symmetry axis?) of the angle γ can also be read off from Fig. 3.8. Since $|\mathbf{L}| \sin \gamma = Ap \sin \alpha$, $|\mathbf{L}| \cos \gamma = Cn$, we have

$$\tan \gamma = \frac{Ap}{Cn} \sin \alpha, \quad (3.85)$$

which shows that Ap/Cn is the controlling ratio. Since we are dealing with steady precession, p and n are connected by (3.82) with the limiting values of p for fast and slow modes of precession given by (3.83). The latter equation and (3.85) combine to show that $\gamma \simeq \alpha$ in fast precession, while $\gamma \simeq 0$ in slow precession, i.e. the angular momentum vector aligns itself along the vertical axis in fast precession, and aligns itself along the spin, or symmetry axis, in slow precession.

If we are dealing with non-steady motions of the top, the above discussion is not complete. Changes in $\dot{\theta}$ away from zero introduce a component $A\dot{\theta}$ of \mathbf{L} along the unit vector \mathbf{g}_θ , so that both the direction and magnitude of \mathbf{L} change, subject always of course to the conditions that the projections M and K are constant. These restrictions mean that \mathbf{L} can at most "wave" back and forth through the plane OzG alternately leading and lagging the spin axis while keeping M and K constant. If the motion considered is pseudo-regular precession, (3.79) tells us that this

fluctuation in length and direction (out of the OzG plane) of \mathbf{L} is periodic with period T_n . Consequently $\tan \gamma(t)$ will vary with time like

$$\tan \gamma(t) \simeq \frac{A\langle\dot{\phi}\rangle \sin \alpha}{Cn} \left[1 - \cos\left(\frac{C}{A}nt\right)\right],$$

with an average value of $(A\langle\dot{\phi}\rangle \sin \alpha)/Cn$, i.e. just the value (3.85) with p replaced by the average precession.

3-8 General Motion of a Rigid Body

So far we have restricted the discussion to the case of a rigid body moving about a fixed point, since this had the advantage of focussing our attention on the pure rotational aspects of the motion and how these may be handled within a Lagrange framework. With this experience behind us, let us remove this restriction and consider the general case of unrestricted motion of a rigid body. In order to get a different perspective of the problem let us develop the general case starting afresh from first principles instead of using a Lagrange equation approach. At the end of our discussion we can then investigate to what extent Lagrange equations can be applied to the general problem of rigid body motion.

We have already set down the general equations of motion for any many-body system in Chapter 1. We recall for convenience the principles of linear and angular momentum:

$$\frac{d\mathbf{P}}{dt} = \mathbf{F}; \quad \frac{d\mathbf{L}}{dt} = \mathbf{N}, \quad (3.86)$$

which relate the rates of change of total linear and angular momentum to the total force (\mathbf{F}) and the total moment or torque (\mathbf{N}) that are responsible for these changes. At first sight, (3.86) appear to be very innocent. The trouble is of course that they only hold with respect to an inertial frame of reference (for our purposes, one fixed in space), while, for example, our expressions for the angular momentum were only simply related to angular velocity in a body-fixed frame which is *not* an inertial frame of reference. The calculation we must do is obvious: retain the simple relations between \mathbf{L} and $\boldsymbol{\omega}$ and modify (3.86) so that they hold in the body-fixed frame. (Note in passing that we have to do this for both equations, since they are coupled to each other by the condition that \mathbf{F} also provides the torque \mathbf{N} and consequently it would be extremely awkward to have the first equation expressed in an inertial frame and the second in a moving, non-inertial frame). But this transformation is simple to perform. We attach a set of axes to the center-of-mass G of the body: call $\boldsymbol{\Omega}$ the angular velocity of this set of axes. If the body has an arbitrary shape, these axes will also only be principal axes if they are rigidly "frozen" into the body; the angular velocity of the body will then also be $\boldsymbol{\omega} = \boldsymbol{\Omega}$. However, if the rigid body has an axis of symmetry

as our top had, we see that the principal axes need only be partially "frozen" into the body, in which case $\boldsymbol{\omega} \neq \boldsymbol{\Omega}$.

Consider now, how the effect of moving axes enters into the expression for the rate of change of a vector like \mathbf{L} . This has a time derivative $d\mathbf{L}/dt$ with respect to fixed axes. Suppose now that \mathbf{L} was rigidly attached to the moving axes that are rotating with angular velocity $\boldsymbol{\Omega}$; its rate of change would be zero in the moving system. In the fixed system, however, \mathbf{L} changes by $\delta\mathbf{L} = \delta t(\boldsymbol{\Omega} \times \mathbf{L})$ in time δt , see (3.2). In general \mathbf{L} will also change with respect to the moving axes. Calling this rate of change $\partial\mathbf{L}/\partial t$ we have

$$\frac{d\mathbf{L}}{dt} = \left(\frac{\partial\mathbf{L}}{\partial t}\right)_{\text{moving}} + (\boldsymbol{\Omega} \times \mathbf{L}), \quad (3.87)$$

where the differentiation on the right is now performed with respect to the moving axes. Equation (3.87) holds the key to our problem. We can now refer both sides of our basic equation (3.86) to moving axes. If these axes are principal axes (as they will be from now on) we possess simple expressions for the components of \mathbf{L} . In moving axes then we have

$$\begin{aligned} \left(\frac{\partial\mathbf{L}}{\partial t}\right)_{\text{moving}} + \boldsymbol{\Omega} \times \mathbf{L} &= \mathbf{N} \\ \left(\frac{\partial\mathbf{P}}{\partial t}\right)_{\text{moving}} + \boldsymbol{\Omega} \times \mathbf{P} &= \mathbf{F}. \end{aligned} \quad (3.88)$$

Let us break up each vector into its components along the principal axes and draw on the expressions (3.12) for the components of angular momentum. Then, we have

$$\begin{aligned} I_1\dot{\omega}_1 + \Omega_2 I_3 \omega_3 - \Omega_3 I_2 \omega_2 &= N_1 \\ I_2\dot{\omega}_2 + \Omega_3 I_1 \omega_1 - \Omega_1 I_3 \omega_3 &= N_2 \\ I_3\dot{\omega}_3 + \Omega_1 I_2 \omega_2 - \Omega_2 I_1 \omega_1 &= N_3 \end{aligned} \quad (3.89)$$

for the motion about the center-of-mass, and

$$\begin{aligned} \dot{P}_1 + \Omega_2 P_3 - \Omega_3 P_2 &= F_1 \\ \dot{P}_2 + \Omega_3 P_1 - \Omega_1 P_3 &= F_2 \\ \dot{P}_3 + \Omega_1 P_2 - \Omega_2 P_1 &= F_3 \end{aligned} \quad (3.90)$$

for the motion of the center-of-mass. We are accordingly treated to the full content of (3.88) which we see is not so simple after all. Equations (3.88), or in their component form above, are called Euler's equations. They were first constructed by him in 1758. Apparently, Euler was the first to recognize the importance of using moving axes in mechanics.

Before going on to consider applications of these equations to problems that merit them let us remark that we could discuss the ordinary top starting from these equations just as well. However, in the case of

the motion of a rigid body about a fixed point the reactions at that point are unknown in general and therefore also some of the force and torque components appearing on the right hand sides of (3.89) and (3.90) are unknown. Thus, it is much more practical to take moments about the fixed point. We introduce principal axes at this point (call it O as before) that rotate with angular velocity $\boldsymbol{\Omega}$, the unknown reactions have zero moments about O and we get

$$\left(\frac{\partial \mathbf{L}}{\partial t}\right)_{\text{moving}} + \boldsymbol{\Omega} \times \mathbf{L} = \mathbf{N}, \quad (3.91)$$

which is *not* the same as the first of (3.88) because \mathbf{L} and \mathbf{N} refer to the fixed point O and not to the center-of-mass G .

3-9 The Tipped-top

If a top like the one in Fig. 3.9(a) is set spinning rapidly on its broad end, it acts rather surprisingly. The thin peg describes a downward spiral and when it touches the floor the top rises rapidly on the peg so that it is spinning as in Fig. 3.9(b). Similarly if one takes an egg-shaped object as in Fig. 3.9(c) one discovers that it will only spin stably on its thin end as shown in Fig. 3.9(d). If set spinning as in Fig. 3.9(c) it also rapidly rises on to its thin end.

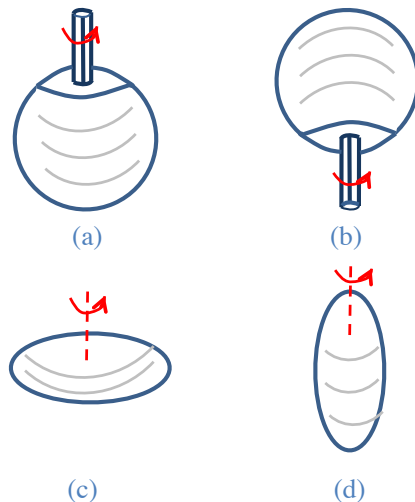


Figure 3.9: Rotating top ((a) and (b)) and rotating egg ((c) and (d)).

Quite a literature has grown up around problems of this sort⁴⁷ which concerns itself with the problem of the inverting top. The tipped-top provides us with an excellent system to study as an example of general rigid body motion. Moreover, there is the interesting question of whether we can understand the reasons for the peculiar behavior of the top, i.e. we have an experimental result that requires an explanation (a rather

⁴⁷ E.G. Gallop, Cambridge Phil. Soc. Transactions **19**, 356 (1904); M.M. Hugenholtz, Physica **18**, 515 (1952); D.G. Parkyn Math. Gazette XL, 260 (1956); Physica **24**, 313 (1958).

rare bird in Classical Mechanics!). We will in fact find that the problem of inverting tops yields to analysis in a straightforward manner.

Fig. 3.10 shows a tippe-top resting in an arbitrary orientation on a horizontal surface (a table).

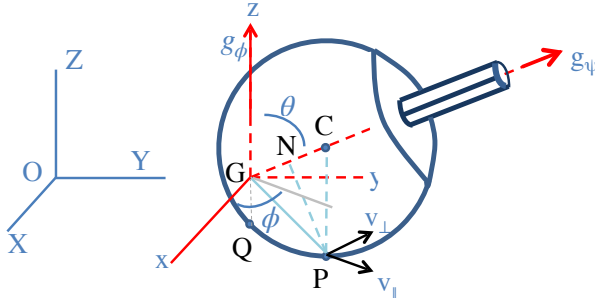


Figure 3.10: The tippe top.

- CP = a
- CG = h
- QG = a-h cos θ
- NP = a sin θ
- QP = h sin θ
- NG = h - a cos θ

We start the problem again by considering the degrees of freedom involved. If the top is required to be in contact with the table at all times, we introduce a constraint on the six degrees of freedom (three for translational motion plus three for rotational motion) that the rigid body has in arbitrary motion. Hence, we have five degrees of freedom and must therefore pick five independent coordinates. We do this in the following manner. Assume that the lower end of the tippe-top is spherical with radius a . The center of this sphere is at C on the symmetry axis of the top passing through its center-of-mass G and along the thin peg as shown. C will always be vertically above the point P where the top touches the table. We set up the usual unit vectors $\mathbf{g}_\phi, \mathbf{g}_\theta, \mathbf{g}_\psi$ attached to the center-of-mass G in order to define the Euler angles ϕ, θ, ψ , which orient the top with respect to axes G_{xyz} that move with G but keep their orientation in space. As in the case of the ordinary top \mathbf{g}_ϕ is vertical, \mathbf{g}_ψ points along the symmetry axis and $\mathbf{g}_\theta = (\mathbf{g}_\phi \times \mathbf{g}_\psi) / \sin \theta$ is normal to the vertical plane containing \mathbf{g}_ϕ and the symmetry axis. Since the top is symmetric about \mathbf{g}_ψ , the orthogonal triad $(\mathbf{f}, \mathbf{g}_\theta, \mathbf{g}_\psi)$ at G where

$$\mathbf{f} = \mathbf{g}_\psi \cot \theta - \frac{1}{\sin \theta} \mathbf{g}_\phi \tag{3.92}$$

lies in the vertical plane containing the symmetry axis, again form a set of principal axes. We call the two equal moments of inertia $I_1 = I_2 = A$ and $I_3 = C$ is the moment of inertia about the symmetry axis.

Notice that A and C now refer to the center-of-mass and not the point of support as was the case previously. The orientation of the tippe-top with respect to axes G_{xyz} is thus fully specified. Finally, we locate G with respect to the table top using axes O_{XYZ} . However, the vertical distance $Z = QG$ of G is not independent of θ because of our constraint on the top. Calling $h = CG$ the center-of-mass - center-of-curvature separation, we see from Fig. 3.10 that

$$Z = QG = (a - h \cos \theta) \quad (3.93)$$

which expresses our constraint equation on Z .

The system of forces acting on the top are, in addition to gravity at G and the normal reaction $\mathbf{R} = R\mathbf{g}_\phi$ at P , also forces of friction at P if the table is not perfectly smooth. The direction of the friction force will always oppose the motion of the contact point P on the spherical surface of the top. Let us break up the displacement of P into components parallel and perpendicular to the line QP . Then, to find these displacements we recall that the triad $(\mathbf{f}, \mathbf{g}_\theta, \mathbf{g}_\psi)$ moves with the top but does not share its rotation about \mathbf{g}_ψ . Hence, infinitesimal increments $\delta\theta^\lambda$ in the Euler angles (ϕ, θ, ψ) will generate the infinitesimal rotations

$$\delta\phi_1 = -\sin\theta\delta\phi, \quad \delta\phi_2 = \delta\theta, \quad \delta\phi_3 = \delta\psi + \delta\phi\cos\theta$$

along this triad. This result is geometrically obvious; alternatively we can use (3.43) with the Q -matrix given by (3.48) with ψ set equal to zero. The displacements of P along QP and normal to QP are thus, from Fig. 3.10

$$\begin{aligned} \delta G_{\parallel} - \delta\phi_2(QG) &= \delta G_{\parallel} - \delta\theta(a - h \cos \theta) \\ \delta G_{\perp} + \delta\phi_3(NP) - \delta\phi_1(NG) &= \delta G_{\perp} + [a\delta\phi_3 + \delta\phi(h - a \cos \theta)] \sin \theta, \end{aligned} \quad (3.94)$$

where δG_{\parallel} and δG_{\perp} symbolize the displacements of G itself along and normal to QP (such displacements are not governed by changes in the Euler angles). Dividing these expressions by the time interval δt during which they occurred, we get the velocity components of P ,

$$\begin{aligned} v_{\parallel} &= u - \dot{\theta}(a - h \cos \theta) \\ v_{\perp} &= v + [a\omega_3 + \dot{\phi}(h - a \cos \theta)] \sin \theta, \end{aligned} \quad (3.95)$$

where ω_3 is the angular velocity of the top along g_ψ . We have called the velocity components of G along and normal to QP , u and v . The friction force can now be displayed as the components

$$\begin{aligned} F_{\parallel} &= -\mu R \frac{v_{\parallel}}{|v_p|} \\ F_{\perp} &= -\mu R \frac{v_{\perp}}{|v_p|} \end{aligned} \quad (3.96)$$

with μ the coefficient of friction, R the normal reaction and $|v_p| = \sqrt{v_{\parallel}^2 + v_{\perp}^2}$ the speed of P .

The next step is to fill in (3.88) for our particular problem. We take moments about G . Then the force of gravity drops out, having no moment about G , and the total moment along \mathbf{g}_{ϕ} and \mathbf{g}_{ψ} arise solely from the frictional component F_{\perp} , while R and F_{\parallel} provide for moments along \mathbf{g}_{θ} . We get from Fig. 3.10

$$\begin{aligned} N_{\phi} &= F_{\perp}(QP) = F_{\perp}h \sin \theta & \text{(a)} \\ N_{\theta} &= -R(QP) - F_{\parallel}(QG) & \text{(b)} \\ N_{\psi} &= F_{\perp}(NP) = F_{\perp}a \sin \theta & \text{(c)} \end{aligned} \tag{3.97}$$

for the moments along \mathbf{g}_{ϕ} , \mathbf{g}_{θ} and \mathbf{g}_{ψ} respectively.

Let us calculate the rate of change $d\mathbf{L}/dt$ along the axes \mathbf{g}_{ϕ} , \mathbf{g}_{θ} and \mathbf{g}_{ψ} also. The angular velocity of the principal axes ($\mathbf{f}, \mathbf{g}_{\theta}, \mathbf{g}_{\psi}$) is

$$\boldsymbol{\Omega} = \mathbf{g}_{\phi}\dot{\phi} + \mathbf{g}_{\theta}\dot{\theta},$$

written in terms of its contravariant components $\dot{\phi}$ and $\dot{\theta}$. We write \mathbf{L} the same way, i.e.

$$\mathbf{L} = \sum_{\lambda} \mathbf{g}_{\lambda} L^{\lambda} = \mathbf{g}_{\phi} L^{\phi} + \mathbf{g}_{\theta} L^{\theta} + \mathbf{g}_{\psi} L^{\psi}$$

and find

$$\frac{d\mathbf{L}}{dt} = \sum_{\lambda} \mathbf{g}_{\lambda} \dot{L}^{\lambda} + \sum_{\lambda} \dot{\mathbf{g}}_{\lambda} L^{\lambda}. \tag{3.98}$$

The time derivatives of \mathbf{g}_{λ} are easy to obtain. We observe that \mathbf{g}_{ϕ} does not rotate at all while \mathbf{g}_{θ} and \mathbf{g}_{ψ} rotate with angular velocity $\boldsymbol{\Omega}$. Noting that

$$\begin{aligned} \mathbf{g}_{\phi} \times \mathbf{g}_{\theta} &= \mathbf{g}_{\phi} \cot \theta - \mathbf{g}_{\psi} \frac{1}{\sin \theta} \\ \mathbf{g}_{\theta} \times \mathbf{g}_{\psi} &= -\mathbf{g}_{\phi} \frac{1}{\sin \theta} + \mathbf{g}_{\psi} \cot \theta \\ \mathbf{g}_{\psi} \times \mathbf{g}_{\phi} &= -\mathbf{g}_{\theta} \sin \theta \end{aligned} \tag{3.99}$$

we have immediately that

$$\begin{aligned} \dot{\mathbf{g}}_{\phi} &= 0 \\ \dot{\mathbf{g}}_{\theta} &= \boldsymbol{\Omega} \times \mathbf{g}_{\theta} = \mathbf{g}_{\phi}(\dot{\phi} \cot \theta) - \mathbf{g}_{\psi}\left(\dot{\phi} \frac{1}{\sin \theta}\right) \\ \dot{\mathbf{g}}_{\psi} &= \boldsymbol{\Omega} \times \mathbf{g}_{\psi} = -\mathbf{g}_{\phi}\left(\dot{\theta} \frac{1}{\sin \theta}\right) + \mathbf{g}_{\theta}(\dot{\theta} \sin \theta) + \mathbf{g}_{\psi}(\dot{\theta} \cot \theta). \end{aligned} \tag{3.100}$$

Upon substitution for \mathbf{g}_λ in (3.98) and regrouping we find $d\mathbf{L}/dt$ to be given by

$$\begin{aligned}\frac{d\mathbf{L}}{dt} &= \mathbf{g}_\phi \left\{ \frac{1}{\sin \theta} \left[\frac{d}{dt} (L^\phi \sin \theta) - \dot{\theta} L^\psi \right] \right\} \\ &\quad + \mathbf{g}_\theta \left\{ \frac{dL^\theta}{dt} + \dot{\phi} L \sin \theta \right\} \\ &\quad + \mathbf{g}_\psi \left\{ \frac{1}{\sin \theta} \left[\frac{d}{dt} (L^\psi \sin \theta) - \dot{\phi} L^\theta \right] \right\},\end{aligned}\tag{3.101}$$

where the L^λ are given by (3.60) with $I_1 = A$ and $I_3 = C$, i.e.

$$L^\phi = A\dot{\phi}, \quad L^\theta = A\dot{\theta}, \quad L^\psi = C\omega_3 - A\dot{\phi} \cos \theta,\tag{3.102}$$

where ω_3 is again the angular velocity component of the top along \mathbf{g}_ψ . The expressions in curly brackets are of course contravariant components. The covariant components along the \mathbf{g}_λ follow with the help of (3.38). We equate these covariant components to the corresponding moments N_λ given by (3.97). Our basic equations for the motion about G then read

$$\begin{aligned}\frac{dL_\phi}{dt} &= F_\perp h \sin \theta \quad (\text{a}) \\ \frac{d}{dt} (A\dot{\theta}) + \dot{\phi} L^\psi \sin \theta &= -Rh \sin \theta - F_{||} (QG) \quad (\text{b}) \\ \frac{d}{dt} (A\omega_3) &= F_\perp a \sin \theta, \quad (\text{c})\end{aligned}\tag{3.103}$$

where

$$L_\phi = A\dot{\phi} + L^\psi \cos \theta = A\dot{\phi} \sin^2 \theta + C\omega_3 \cos \theta$$

and

$$L_\psi = L^\psi + A\dot{\phi} \cos \theta = C\omega_3$$

are the covariant components of the angular momentum along the vertical and along the symmetry axis as we already discovered in (3.59).

We supplement these with the equations of motion for the center-of-mass G . Since the velocity components u, v of G that appear in (3.95) are referred to the axis QP which is rotating with angular velocity $\dot{\phi} \mathbf{g}_\phi$ about the vertical, we must use (3.90) with $\boldsymbol{\Omega} = \dot{\phi} \mathbf{g}_\phi$. Then G moves according to

$$\begin{aligned}m(\dot{u} - v\dot{\phi}) &= F_{||} \quad (\text{a}) \\ m(\dot{v} + u\dot{\phi}) &= F_\perp \quad (\text{b}) \\ m \frac{d^2}{dt^2} (a - h \cos \theta) &= R - mg \quad (\text{c})\end{aligned}\tag{3.104}$$

in terms of acceleration components along and normal to QP and the vertical.

The equations (3.103) and (3.104) together with the information given by (3.95) and (3.96) define our problem completely. We observe immediately that the "fly in the ointment" is the force of friction. Without it, we would regain our two previous angular momentum constants $L_\phi = M$ and $L_\psi = K$. Moreover, the total energy would also be conserved (the top would not scrape on the table) and our problem for the tippe-top would reduce in all respects to that of the ordinary top. However, let us first see how far we can go without making specific assumptions about the nature of the motion. The first point to notice is, that while L_ϕ and $C\omega_3$ are no longer conserved quantities the linear combination $(aL_\phi - hC\omega_3)$ is. This follows from (3.103a) and (3.103c) upon multiplication by a and h respectively and subtracting. We have

$$\frac{d}{dt} [aA\dot{\phi} \sin^2 \theta + aC\omega_3 \cos \theta - hC\omega_3] = 0 \quad (3.105)$$

or

$$aA\dot{\phi} \sin^2 \theta + (a \cos \theta - h)C\omega_3 = \text{constant} \quad (3.106)$$

after inserting the value of L_ϕ given below (3.103). We have thus found one integral of motion even in the presence of friction. The latter relation, (3.106), is known as Jellett's integral. It seems first to have been discovered by him and is quoted in his book on friction⁴⁸. As far as is known, this is the only first integral of motion that exists for a top spinning on a rough surface.

⁴⁸ Jellet, *Theory of friction*, Chap. viii, 1872.

To make further progress we observe that (3.104) allow us to replace the unknown friction forces and normal reaction on the right of (3.103) by the accelerations they produce for G . In the case of R this is very helpful for we have *prescribed* how G must move vertically through the constraint equation (3.93). We eliminate R , F_\perp and F_\parallel from (3.103) in this way and find

$$\frac{dL_\phi}{dt} = m(\dot{v} + u\dot{\phi})h \sin \theta \quad (a)$$

$$\begin{aligned} \frac{d}{dt} [(A + mh^2 \sin^2 \theta)\dot{\theta}] - mh^2 \sin \theta \cos \theta \dot{\theta}^2 + \dot{\phi}L^\psi \sin \theta \\ = -mgh \sin \theta - m(\dot{u} - v\dot{\phi})(a - h \cos \theta) \end{aligned} \quad (b)$$

$$\frac{d}{dt} (C\omega_3) = m(\dot{v} + u\dot{\phi})a \sin \theta. \quad (c)$$

(3.107)

These equations still contain the unknown acceleration and velocity components of the center-of-mass G and so are not too much help as they stand. We have to provide information about the physical conditions at the point of contact as the following limiting cases will illustrate.

(a) *Pure sliding*. We assume the table is perfectly smooth so that the top slides freely. Then, the contact point P slides without friction on the table and the center-of-mass only has an acceleration in the vertical direction. The horizontal motion of G is one of constant velocity in a straight line. Steady precession in the pure sliding mode takes place with G at rest and P moving in a circle. Since $\theta = \alpha$, $\dot{\phi} = p$, $\omega_3 = n$ in addition to $u = 0$, $v = 0$, $\dot{u} = 0$, $\dot{v} = 0$ in steady precession, (3.107b) becomes

$$pL^\psi \sin \alpha = -mgh \sin \alpha, \quad (3.108)$$

which is just (3.82) again, but in a compact form. Except for one point: the sign of h is different. Setting $L^\psi = Cn - Ap \cos \alpha$, its steady precession value, we find that (3.108) has the two roots for p ($\sin \alpha \neq 0$),

$$Cn \simeq Ap \cos \alpha, \quad \text{or} \quad Cn \simeq -\frac{mgh}{p}, \quad (3.109)$$

connecting n and p in the fast and slow precession respectively, if Cn is large. The sign for the slow precession is different, indicating that if we started off the tippe-top like we did the ordinary top by spinning it and setting it down gently it would choose the *fast* precession root $Cn \simeq Ap \cos \alpha$, and the nutational motion consists of the top nodding up and down while G moves up and down vertically. The rest of the analysis proceeds as in the case of the ordinary top. As we have already mentioned, the three first integrals of motion, conservation of angular momentum along the vertical and symmetry axis, and conservation of energy, exist and allow us to duplicate the ordinary top discussion.

(b) *Pure rolling*. The top is said to roll if the point of contact P is always instantaneously at rest so that there is no relative motion between the surface of the top and the table at their common point of contact. The condition for pure rolling can be expressed as conditions on the angular velocity and orientation of the top and the velocity of G . From (3.95) the point P is instantaneously at rest if v_{\parallel} and v_{\perp} are zero, i.e.

$$\begin{aligned} u &= (a - h \cos \theta) \dot{\theta} \\ v &= -[a\omega_3 + (h - a \cos \theta) \dot{\phi}] \sin \theta, \end{aligned} \quad (3.110)$$

which tell us what the velocity components of G *must be* if the top rolls. This in turn means that the table must be rough enough to sustain G in a motion with u and v given by (3.110). Notice in passing that the pure rolling conditions are *non-holonomic*. We cannot integrate (3.110) in time until we know the motion of the top.

The usefulness of these constraints on the motion of G is that they allow us to eliminate friction entirely in terms of the motion it causes, i.e. rolling. The right sides of (3.107) then become *prescribed* functions of the Euler angles and their time derivatives. Let us look at (3.107c) in

particular. We obtain the combination

$$(\dot{v} + u\dot{\phi})a \sin \theta = -a \sin \theta \frac{d}{dt}(a\omega_3 \sin \theta) \\ - (h - a \cos \theta) \frac{d}{dt}(a\dot{\phi} \sin^2 \theta)$$

from (3.110) and insert it into the right side of (3.107c) after employing Jellett's relation, (3.106), to find $d/dt(a\dot{\phi} \sin^2 \theta)$. The result can be expressed as

$$\frac{1}{2} \frac{d}{dt} \left\{ \omega_3^2 \left[C + ma^2 \sin^2 \theta + m \frac{C}{A} (h - a \cos \theta)^2 \right] \right\} = 0,$$

or that

$$\omega_3^2 \left[C + ma^2 \sin^2 \theta + m \frac{C}{A} (h - a \cos \theta)^2 \right] = \text{constant} \quad (3.111)$$

is a constant of the motion. We have thus discovered another constant of motion for pure rolling, in addition to the conservation of energy (friction forces do no work in pure rolling motion) and Jellett's integral (which is always valid). Thus, we are once more armed with three first integrals of motion and we can discuss precession, nutation and the stability thereof as we did before for the ordinary top. We note one difference from case (a): In pure rolling the steady precession motion has both G and P going around in circles about each other while G still stays at a fixed height about the table (P instantaneously at rest does *not* mean P stays in the same spot!). The equation relating α , p , and n in steady precession follows from (3.107b) with \dot{u} set to zero, v given by (3.110):

$$pL^\psi \sin \alpha = -mgh \sin \alpha - m[an + (h - a \cos \alpha)p] \sin \alpha, \quad (3.112)$$

which differs in detail from the condition when pure slipping is present, but is still a quadratic equation for p and we expect slow and fast precession modes for large Cn as before.

(c) *Quasi steady motion.* If we assume that the friction forces and moments are small the motion of the top will be a quasi steady version of the precessional motion in pure sliding. For we may argue that, if the friction is small it acts like a disturbance of the steady motion which, as we know, produces a small oscillatory motion of the top about its steady precession orientation $\theta = \alpha$. (See problems). With one difference: The "small disturbance" one always invokes to perturb a system in order to ascertain its stability does just that and then goes away. This is not true for friction, which of course continues to act so that we expect a steady "drift" away from steady precession in addition to the oscillations we just mentioned. Now, we argue that the oscillations in θ around $\theta = \alpha$ are very rapid for large spin and hence will average out over times in excess of the nutation period T_n . Therefore, if we look only

at the time-averaged motion for times long compared with T_n , but short compared with the precession period T_p (so that the motion does not depart substantially from steady precession), we should be able to estimate the effects introduced by friction. However, we can see intuitively from (3.107a) what is going to happen: the frictional moment $F_{\perp} a$ tends to increase the projection of \mathbf{L} along the z -axis if F_{\perp} is positive, and decrease it if F_{\perp} is negative. A decreasing L_{ϕ} means that the angular momentum vector is falling and so too the symmetry axis, since from (3.107c) $C\omega_3 = L_{\psi}$ must also decrease. The top therefore begins to fall if $F_{\perp} < 0$. But the direction of F_{\perp} is determined by the velocity direction v_{\perp} of the contact point on the top (F_{\perp} is always opposite to v_{\perp}). Hence a falling top is characterized by $v_{\perp} > 0$, or

$$[an + (h - a \cos \alpha)p] \sin \alpha > 0, \quad (3.113)$$

(the velocity v of G is near zero) which for the tippe-top, where $Cn \simeq Ap \cos \alpha$, becomes⁴⁹

$$h + \frac{A - C}{C} a \cos \alpha > 0 \quad (3.114)$$

a purely geometrical condition for a given top. So, when the tippe-top spins on its thin peg we can again use (3.114) to ascertain whether the top falls or rises. Assuming that the tip of the peg is spherical, radius a , and setting $h \rightarrow -h$ (for the center-of-mass G now lies on the other side of C) we note $a \ll |h|$ so that (3.114) is violated: the top climbs up on its peg instead of falling.

For the more curious reader we present a more motivated derivation of the condition (3.113). The discussion starts with (3.103c); we assume that deviations away from steady precession are small so the F_{\perp} may be replaced by a constant value $F_{\perp}^{(0)}$ for quasi steady precession. The average change in the angular momentum Cn is thus (bars denote a time average)

$$C\bar{\delta}n \simeq F_{\perp}^{(0)} ta \sin \alpha \quad (T_n \ll t \ll T_p), \quad (i)$$

where T_n, T_p are the nutation and precession periods. The change (i) induces a corresponding change in p which, from Jellett's relation is

$$aA \sin^2 \alpha \bar{\delta}p \simeq (h - a \cos \alpha) C \bar{\delta}n. \quad (ii)$$

In the absence of friction, changes in θ respond to an outside disturbance according to

$$\ddot{\zeta} + \Omega_n^2 \zeta = 0, \quad (iii)$$

where $\Omega_n^2 = (2\pi)^2/T_n^2 = [(Cn - 2Ap \cos \alpha)^2 + (Ap \sin \alpha)^2]/(AA')$ and $A' = A + mh^2 \sin^2 \alpha$.

The result (iii) follows from (3.107b) of the text after setting $\theta = \alpha + \zeta$ and ignoring the forms of friction $F_{\parallel} \sim (\dot{u} - v\dot{\phi})$. The additional terms (i) and (ii) appear on the right side when friction (the $F_{\perp}^{(0)}$ dominantly) is present. Then, (iii) is replaced by

$$A'\ddot{\zeta} + A'\Omega_n^2 \zeta = -\kappa C \bar{\delta}n \quad (iv)$$

⁴⁹ This condition was first given by D.G. Parkyn, *Math. Gazette XL*, p. 260 (1956).

with $\kappa = [aAp \sin^2 \alpha + (h - a \cos \alpha)(Cn - 2Ap \cos \alpha)] \operatorname{cosec}^2 \alpha / (aA)$ if we continue to omit $F_{\parallel}^{(0)}$. The particular integral of (iv) gives the forced motion:

$$\bar{\zeta}(t) = -\frac{\kappa C}{A'\Omega_h^2} \delta \bar{n} = -\frac{1}{A'\Omega_h^2} \kappa a \sin \alpha F_{\perp}^{(0)} t. \quad (v)$$

We readily verify that κ is positive when $Cn \simeq Ap \cos \alpha$; hence $\zeta(t)$ increases (the top falls) when $F_{\perp}^{(0)} < 0$, or $v_{\perp}^{(0)} > 0$, which is the condition stated in (3.113). More generally, we point out that for an ordinary top where $h \rightarrow -h$ in (3.113) and $a \ll h$ we have the condition $(an - hp) > 0$ independent of the angle α , giving a critical precession beyond which the effect is reversed. D.G. Parkyn [*Physica* XXIV, p. 313 (1958)] has analysed the motion of a top with a rounded peg both theoretically and experimentally. He finds that an ordinary top in motion does effectively remain in the rolling mode, but that the motion is not in general one of pure rolling due to periodic sliding-rolling motions of the peg.

Recapitulation

If it seemed to the reader that the discussion of just two problems in rigid body mechanics took up far too much space, let him be mindful of the fact that (i) rigid body problems *are* difficult in general and (ii) the two problems we have discussed in detail serve as prototypes to illustrate the methods that one would employ in tackling any problem involving rigid body motion under given forces. Such methods are important and reach a high degree of sophistication in applied problems such as rocket guidance systems, gyrocompass design, servomechanisms etc. We do not discuss such applications here. On the more academic side, we have also avoided discussing the "egg rolling on a rotating paraboloid" types of problem, which tend to become rather devoid of physical content. However, such "academic" rigid body systems are indelibly woven into the historical fabric of mechanics and, like chess, they have their fascination. The reader who wishes to explore such topics further will do well to consult Routh's⁵⁰ *Advanced Rigid Dynamics*, or in a more modern (1965) vein, the treatise of Pars⁵¹.

⁵⁰ E.J. Routh, *Advanced Rigid Dynamics*, Sixth Edition, MacMillan and Co., Ltd., London, 1905.

⁵¹ L.A. Pars, *A Treatise on Analytical Dynamics*, John Wiley and Sons, Inc., New York, 1965.

Problems

3-1. Consider a rigid body in general motion about a fixed point O . Show that an instantaneous axis of rotation (i) exists and (ii) that this axis passes through O .

3-2. Find the principal moments of inertia and principal axes for the following rigid bodies at the points indicated. The mass of each body is M .

- (i) a square plate, side a , at one corner
- (ii) a rectangular plate, sides a and b , at one corner
- (iii) a semi-circular plate of radius a , at one of the "corners"
- (iv) a plate shaped like an equilateral triangle of side a , at any vertex

(v) a cube, side a , at the center-of-mass

3-3. Discuss the motion of a rigid body having an axis of symmetry and supported freely at its center-of-mass using the conservation laws only. Indicate how the direction of the angular velocity vector ω behaves with respect to the direction of the angular momentum vector \mathbf{M} .

3-4. A uniform solid circular cylinder of radius a rests on a horizontal plane, and an identical cylinder rests on it, touching it along the highest generator. If no slipping occurs, show that as long as the cylinders remain in contact,

$$\dot{\theta}^2 = \frac{12g(1 - \cos \theta)}{a(17 + 4 \cos \theta - 4 \cos^2 \theta)}$$

where θ is the angle which the plane containing the cylinder axes makes with the vertical (see Fig. 3.11).

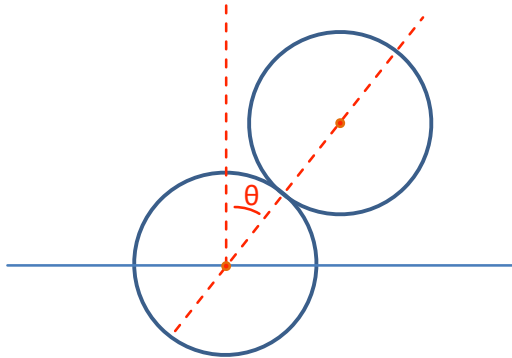


Figure 3.11: Sketch of two cylinders touching at a tangent point.

Show that the path described by the axis of the upper cylinder is

$$x = \frac{1}{3}a(\theta + 4 \sin \theta), \quad y = 2a(1 - \cos \theta)$$

relative to horizontal (x) and vertical (y) axes through the initial position of this axis.

3-5. A uniform right circular cone of semivertical angle α rolls *without slipping* (what is the condition for this?) on a plane inclined at an angle β with the horizontal, and is released from rest with the line of contact horizontal. Prove that the cone will remain in contact with the plane if

$$9 \tan \beta < \cot \alpha + 4 \tan \alpha.$$

3-6. An asymmetrical top rotates about its center of mass under no forces. The principal moments of inertia at the center-of-mass are $I_1, I_2, I_3, (I_1 > I_2 > I_3)$. If $\omega = (\omega_1, \omega_2, \omega_3)$ give the angular velocity

components along the principal axes, show that they have the form

$$\omega_1(t) = -A \operatorname{sech} \tau(t); \quad \omega_2(t) = B \tanh \tau(t); \quad \omega_3(t) = C \operatorname{sech} \tau(t),$$

if $M^2 = 2I_2T$ ($M =$ angular momentum, $T =$ kinetic energy) and $\omega_3 > 0, \omega_1 < 0$ at $t = 0$. Find A, B, C and $\tau(t)$. What happens as the time t increases indefinitely?

3-7. A gyro is set spinning with angular momentum \mathbf{m} along a direction making an angle α with the symmetry axis. Show that the symmetry axis precesses around the vector \mathbf{m} with angular velocity $|m|/I_1$. Show that this can also be written as $I_3\omega_3/I_1 \cos \alpha$, where ω_3 is the angular velocity along the symmetry axis.

3-8. A top is started spinning vertically so that θ and $\dot{\theta}$ are zero initially. Show that if $\omega_3^2 > 4mgh(I_1/I_3^2)$, the angle θ remains zero; if ω_3^2 is smaller than this the symmetry axis of the top will oscillate between O and an angle α . Find α .

3-9. Show that a steady precession of a top about the vertical is possible with its symmetry axis *horizontal* and find the relation between the precessional angular velocity and the "spin" of the top about its symmetry axis.

Chapter 4 Small Oscillations

4-1 Introduction

The systems of interacting particles we studied in the previous chapter in connection with the motion of rigid bodies were many-body systems of a rather special type because of the assumed nature of the interactions between particles (they were such as to render the system *rigid*). Here, we discuss another many-body problem of a rather special type: small oscillations of the constituent particles of the system about a stable equilibrium configuration. Such systems are realized, by for example, relaxing the condition of complete rigidity imposed in Chap. 3, but considering only small displacements away from equilibrium of the particles forming a rigid body. Notice that there is no guarantee that displacing particles in a many-body system away from their equilibrium positions in this manner always leads to oscillatory motion. For that to happen, the equilibrium configuration must be a *stable* one. This in turn is dictated by the nature of the potential energy function of the system in the neighborhood of the equilibrium configuration.

4-2 A simple example: Coupled pendulums.

To begin the study of small oscillations we look at a typical problem and try to guess a solution. Afterwards, we can proceed with the formal theory that leads to the solution in an organized and rigorous fashion. The rather innocent looking problem we discuss first is that of two simple pendulums each of length l , coupled to each other by a weightless spring, see Fig. 4.1(a).

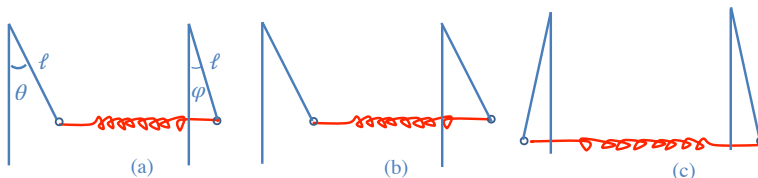


Figure 4.1: Coupled pendulums.

In terms of the angular displacements θ and ϕ the Lagrange function is readily found to be

$$L = \frac{1}{2}ml^2\dot{\theta}^2 - mgl(1 - \cos\theta) + \frac{1}{2}ml^2\dot{\phi}^2 - mgl(1 - \cos\phi) - \frac{1}{2}m\kappa l^2(\theta - \phi)^2,$$

if m is the mass of either bob, and κ is the spring constant. The last term in this expression provides a coupling between the free motion of each pendulum separately. Call x_1, x_2 the horizontal displacements of each bob from the equilibrium positions. Then, for *small* displacements ($l\theta, l\phi \ll 1$)

$$x_1 \simeq l\theta, \quad x_2 \simeq l\phi,$$

and the Lagrange function becomes

$$L \simeq \frac{1}{2}m\dot{x}_1^2 + \frac{1}{2}m\dot{x}_2^2 - \frac{1}{2}m\omega_0^2 x_1^2 - \frac{1}{2}m\omega_0^2 x_2^2 - \frac{1}{2}m\kappa(x_1 - x_2)^2,$$

where $\omega_0 = \sqrt{g/l}$ is the angular frequency of either pendulum oscillating freely. The equations of motion for x_1 and x_2 are

$$\frac{\partial L}{\partial x_k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_k}\right) = 0, \quad k = 1, 2,$$

or

$$\begin{aligned} \ddot{x}_1 + \omega_0^2 x_1 &= -\kappa(x_1 - x_2) \\ \ddot{x}_2 + \omega_0^2 x_2 &= -\kappa(x_2 - x_1). \end{aligned}$$

These equations are not independent because of the *foreign* coordinate x_1 or x_2 on the right hand sides of these equations respectively. If it were not for such *coupling terms* each pendulum would oscillate with its own frequency ω_0 . We try a solution of the form

$$x_1 = u_1 e^{i\Omega t}, \quad x_2 = u_2 e^{i\Omega t},$$

(we later can take real or imaginary parts, see remarks below (2.9)), where Ω is an unknown but common frequency and u_1, u_2 arbitrary amplitudes. This solution is compatible with the equations of motion if

$$\begin{aligned} (\omega_0^2 + \kappa - \Omega^2)u_1 - \kappa u_2 &= 0 \\ -\kappa u_1 + (\omega_0^2 + \kappa - \Omega^2)u_2 &= 0 \end{aligned}$$

or

$$\begin{bmatrix} \omega_0^2 + \kappa - \Omega^2 & -\kappa \\ -\kappa & \omega_0^2 + \kappa - \Omega^2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0.$$

Thus, we obtain a pair of *homogeneous* algebraic equations for the amplitudes u_1 and u_2 . These are non-zero only if the determinant of their coefficients vanishes, that is

$$\begin{vmatrix} \omega_0^2 + \kappa - \Omega^2 & -\kappa \\ -\kappa & \omega_0^2 + \kappa - \Omega^2 \end{vmatrix} = (\omega_0^2 + \kappa - \Omega^2)^2 - \kappa^2 = 0,$$

which provides for a determination of the unknown Ω . This is an algebraic equation of the second degree in Ω^2 . Hence, there are *two* possible frequencies

$$\Omega_1 = \omega_0, \quad \text{and} \quad \Omega_2 = \sqrt{\omega_0^2 + 2\kappa} \simeq \omega_0 + \frac{\kappa}{\omega_0},$$

the last approximation holding if the coupling is weak, $\kappa/\omega_0^2 \ll 1$, at which x_1 and x_2 can vibrate. The two values of Ω we have just found are called *eigenfrequencies* of the coupled system. The amplitudes of x_1 and x_2 in an *eigenvibration*, or *normal mode*, are found by returning to the homogeneous equations for u_1 and u_2 and substituting in the two values of Ω . Thus, for $\Omega = \Omega_1$ we find

$$\begin{bmatrix} \kappa & -\kappa \\ -\kappa & \kappa \end{bmatrix} \begin{bmatrix} u_1^{(1)} \\ u_2^{(1)} \end{bmatrix} = 0, \quad \text{or} \quad u_1^{(1)} = u_2^{(1)}$$

and for $\Omega = \Omega_2$

$$\begin{bmatrix} -\kappa & -\kappa \\ -\kappa & -\kappa \end{bmatrix} \begin{bmatrix} u_1^{(2)} \\ u_2^{(2)} \end{bmatrix} = 0, \quad \text{or} \quad u_1^{(2)} = -u_2^{(2)},$$

now adding superscripts 1 and 2 to the u 's to distinguish between the two normal modes. Thus, only the *ratios* of u_1 to u_2 are determined. The motion of the system in the normal mode $\Omega_1 = \omega_0$ thus has (α_1 is the phase of the arbitrary complex amplitude A_1)

$$x_1 = x_2 = \text{Re}(A_1 e^{i\Omega_1 t}) = |A_1| \cos(\Omega_1 t + \alpha_1)$$

or the pendulums oscillating in phase without ever stretching the spring, while the mode $\Omega_2 = \sqrt{\omega_0^2 + 2\kappa}$ has the pendulums oscillating out of phase,

$$x_1 = -x_2 = \text{Re}(A_2 e^{i\Omega_2 t}) = |A_2| \cos(\Omega_2 t + \alpha_2),$$

with $A_2 = |A_2| \exp i\alpha_2$ as before. The two normal modes are shown in Fig. 4.1(b) and (c).

The solutions we have just found are clearly very special. They are called the *eigensolutions*, or *eigenvibrations*, of the system. In terms of them, we can write the general solutions as the superpositions

$$x_1 = |A_1| \cos(\Omega_1 t + \alpha_1) + |A_2| \cos(\Omega_2 t + \alpha_2)$$

for x_1 , and

$$x_2 = |A_1| \cos(\Omega_1 t + \alpha_1) - |A_2| \cos(\Omega_2 t + \alpha_2)$$

for x_2 . For instance, if the left hand pendulum is pulled aside a (small) distance a and released, the initial conditions $x_1 = a$, $x_2 = 0$, $\dot{x}_1 = \dot{x}_2 = 0$ are met by the choice of constants

$$|A_1| = |A_2| = \frac{1}{2}a, \quad \alpha_1 = \alpha_2 = 0.$$

The subsequent motion of each pendulum is then described by

$$\begin{aligned}x_1 &= \frac{1}{2}a[\cos \Omega_1 t + \cos \Omega_2 t] = a \cos\left(\frac{\Omega_2 - \Omega_1}{2}t\right) \cos\left(\frac{\Omega_1 + \Omega_2}{2}t\right) \\x_2 &= \frac{1}{2}a[\cos \Omega_1 t - \cos \Omega_2 t] = a \sin\left(\frac{\Omega_2 - \Omega_1}{2}t\right) \sin\left(\frac{\Omega_1 + \Omega_2}{2}t\right).\end{aligned}$$

In the latter form these expressions show the familiar phenomenon of *beats*. This is most easily appreciated in the weak coupling limit when

$$\frac{1}{2}(\Omega_2 - \Omega_1) \simeq \frac{\kappa}{2\omega_0}, \quad \frac{1}{2}(\Omega_2 + \Omega_1) \simeq \omega_0.$$

Then

$$x_1 \simeq a \cos\left(\frac{\kappa t}{2\omega_0}\right) \cos \omega_0 t, \quad x_2 \simeq a \sin\left(\frac{\kappa t}{2\omega_0}\right) \sin \omega_0 t$$

showing that each pendulum oscillates with its natural frequency ω_0 but with an amplitude that varies slowly with time like $a \cos(\kappa t/2\omega_0)$ or $a \sin(\kappa t/2\omega_0)$. Thus, the amplitude of the right hand pendulum increases at the expense of the amplitude of the left hand pendulum and vice versa, each alternately coming to rest and reaching a maximum oscillation amplitude.

This example teaches us one very important fact. If we call

$$\operatorname{Re}(A_1 e^{i\Omega_1 t}) = \zeta_1 \quad \text{and} \quad \operatorname{Re}(A_2 e^{i\Omega_2 t}) = \zeta_2,$$

and note that

$$\zeta_1 = \frac{x_1 + x_2}{2}, \quad \zeta_2 = \frac{x_1 - x_2}{2},$$

then the equations of motion for x_1 and x_2 can be rewritten as

$$\frac{1}{2}(\ddot{x}_1 + \ddot{x}_2) + \frac{1}{2}\omega_0^2(x_1 + x_2) = 0, \quad \text{or} \quad \ddot{\zeta}_1 + \omega_0^2\zeta_1 = 0,$$

and

$$\frac{1}{2}(\ddot{x}_1 - \ddot{x}_2) + \frac{1}{2}(\omega_0^2 + 2\kappa)(x_1 - x_2) = 0, \quad \text{or} \quad \ddot{\zeta}_2 + (\omega_0^2 + 2\kappa)\zeta_2 = 0,$$

that is, the equations of motion *uncouple* in the coordinate ζ . Correspondingly the Lagrange function is reduced to a sum of squares,

$$L = \frac{1}{2}m(\dot{\zeta}_1^2 + \dot{\zeta}_2^2) - \frac{1}{2}m\omega_0^2\zeta_1^2 - \frac{1}{2}m(\omega_0^2 + 2\kappa)\zeta_2^2$$

when expressed in such coordinates, called the *normal coordinates* of the problem. It is our intention in the following to seek such coordinates in general for the solution of small oscillation problems. The eigenfrequencies will again appear as a byproduct of this procedure.

4-3 Statement of the Problem

Let us examine the motion of a many-body system consisting of N interacting particles. The Lagrange function is given by (2.105), that is

$$L = \sum_k \frac{1}{2} m_k \dot{x}_k^2 - V(x_1, x_2, \dots), \quad (4.1)$$

where the sum on k runs from 1 to $3N$ as there are $3N$ degrees of freedom. The potential V is due to the mutual interaction between particles only. We exclude for the time being the presence of an external field.

We are interested in the motion described by (4.1) in the neighbourhood of an *equilibrium configuration* (x_1^0, x_2^0, \dots) in which all particles are permanently at rest. This means that

$$\left(\frac{\partial V}{\partial x_k} \right)_{x_k=x_k^0} = 0 \quad (4.2)$$

at equilibrium, for all k . The set of $3N$ equations (4.2) define the equilibrium coordinates $x_k = x_k^0$ of the system. Since we are interested in small displacements of the system about the x_k^0 , we treat each displacement $x_k - x_k^0$ as small and expand L about its equilibrium value $L_0 = -V(x_1^0, x_2^0, \dots)$. Then,

$$L = L_0 + \sum_k \frac{1}{2} m_k \dot{x}_k^2 - \sum_{kl} \frac{1}{2} V_{kl} (x_k - x_k^0)(x_l - x_l^0) - \dots \quad (4.3)$$

The constants V_{kl} are given by

$$V_{kl} = \frac{\partial^2 V}{\partial x_k^0 \partial x_l^0} \quad (4.4)$$

and are symmetrical in the indices k and l , $V_{kl} = V_{lk}$.

The further development of the theory of small oscillations now proceeds on the assumption that higher order terms in (4.3) may be neglected. Measuring each x_k from its equilibrium value and dropping the constant L_0 our approximation to the full Lagrange function reads

$$L = \sum_k \frac{1}{2} m_k \dot{x}_k^2 - \sum_{kl} \frac{1}{2} V_{kl} x_k x_l. \quad (4.5)$$

Clearly this Lagrangian describes a system with restoring forces between particles proportional to particle displacements away from equilibrium. One then speaks of a system of coupled linear oscillators.

The difficulty with (4.5) lies with the coupling between the different x_k . We therefore try to decouple these coordinates by introducing new coordinates ζ_i such that L assumes the form

$$L = \sum_{i=1}^{3N} \left(\frac{1}{2} \dot{\zeta}_i^2 - \frac{1}{2} \Omega_i^2 \zeta_i^2 \right), \quad (4.6)$$

i.e. turns into a sum of $3N$ independent oscillators each having its own frequency Ω_i when expressed in the coordinates ζ . This form of L is often called its *normal form*.

4-4 Normal Modes of a System of Coupled Oscillators

The transformation leading to the normal form of L is usually accomplished in two steps: (i) introduction of *mass weighted* coordinates $x'_k = \sqrt{m_k}x_k$ instead of the x_k . Then, L assumes the form

$$L = \sum_k \frac{1}{2} \dot{x}'_k{}^2 - \sum_{kl} \frac{1}{2} W_{kl} x'_k x'_l, \quad (4.7)$$

where $W_{kl} = V_{kl}/\sqrt{m_k m_l}$. This is just a scale transformation. (ii) introduction of the ζ_i by means of a linear orthogonal transformation (a rotation)

$$x'_k = \sum_i U_{ki} \zeta_i \quad (4.8)$$

with constant coefficients U_{ki} . The velocities \dot{x}'_k then transform in the same way,

$$\dot{x}'_k = \sum_i U_{ki} \dot{\zeta}_i \quad (4.9)$$

since the coefficients U_{ki} are independent of time. The orthogonality of the transformation (4.8) is expressed, as in Chap. 3, by the conditions

$$\sum_i U_{ik} U_{il} = \sum_i U_{ki} U_{li} = \delta_{kl} \quad (4.10)$$

or

$$U^T U = U U^T = I \quad (4.11)$$

in matrix notation, where U is the array

$$U = \begin{pmatrix} U_{11} & U_{12} & \dots \\ U_{21} & U_{22} & \dots \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \cdot \end{pmatrix}.$$

As before U^T denotes the transpose⁵² of the matrix U , and I the unit matrix. In terms of the new coordinates, L reads

$$L = \sum_{ijk} \frac{1}{2} \dot{\zeta}_i U_{ki} U_{kj} \dot{\zeta}_j - \sum_{ijkl} \frac{1}{2} \zeta_i U_{ki} W_{kl} U_{lj} \zeta_j,$$

or

$$L = \frac{1}{2} \dot{\zeta}^T (U^T U) \dot{\zeta} - \frac{1}{2} \zeta^T (U^T W U) \zeta, \quad (4.12)$$

⁵² The matrix obtained by interchanging the rows and columns of U . The reader unfamiliar with elementary matrix operations is advised to consult A.C. Aitken, *Determinants and Matrices*, Seventh Edition, Oliver and Boyd, Edinburgh and London, 1951.

using the more compact notation of matrix multiplication. Here, ζ^T and ζ stand for the row and column vectors

$$\zeta^T = (\zeta_1, \zeta_2, \dots); \quad \zeta = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} = \{\zeta_1, \zeta_2, \dots\}.$$

In what follows we will often abbreviate the column vector by writing it as a row vector in curly brackets as shown above. Finally the matrix W stands for the symmetrical array

$$W = \begin{pmatrix} W_{11} & W_{12} & \dots \\ W_{12} & W_{22} & \dots \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \cdot \end{pmatrix}.$$

Since U has been taken as an orthogonal matrix the kinetic energy term in (4.12) already has the desired form. To obtain a diagonal form for the potential energy term shown in (4.6), we also require that

$$U^T W U = (\Omega_i^2 \delta_{ij}) = \begin{pmatrix} \Omega_1^2 & 0 & \dots \\ 0 & \Omega_2^2 & \dots \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \cdot \end{pmatrix}, \quad (4.13)$$

i.e. the matrix W is brought into a diagonal form by the orthogonal transformation U . It is known from matrix theory that such transformation matrices can always be found if, as here, the matrix to be diagonalized is real and symmetric.

Once L is in the form (4.6) one easily calculates that the equation of motion

$$\ddot{\zeta}_i + \Omega_i^2 \zeta_i = 0 \quad (4.14)$$

holds for each ζ_i . Hence

$$\zeta_i(t) = \text{Re}(C_i \exp i\Omega_i t) = |C_i| \cos(\Omega_i t + \delta_i), \quad (4.15)$$

where $C_i = |C_i| \exp \delta_i$ is a complex integration constant. Knowing the ζ_i as a function of time, and the matrix U , one can determine the x'_k (and hence the x_k) as a function of time from (4.8).

As in the example of the coupled pendulums, the frequencies Ω_i of each ζ_i are known as the *eigenfrequencies* of the system, the ζ_i as its *normal coordinates*. Thus, (4.8) simply says that the actual displacements x_k are given as a *linear superposition* of the normal modes of the system.

How to find the Ω_i and the ζ_i ? Equation (4.13) provides the answer to the first part of this query. We must diagonalize W . In the process one also obtains the matrix U and hence the ζ_i , as we now show.

To this end consider the operation of W on an arbitrary column vector \mathbf{u} . This is given by

$$W\mathbf{u} = \mathbf{v}, \tag{4.16}$$

with the interpretation: W operates on \mathbf{u} and produces a new column vector \mathbf{v} . We saw in Chap. 3 (3.27), that such an operation corresponds to a rotation of the vector \mathbf{u} into a new vector \mathbf{v} if the transformation is length preserving. This latter property is not necessarily true for an arbitrary matrix W , and \mathbf{v} can differ from \mathbf{u} in magnitude and direction. However, there exists a certain class of vectors, called *eigenvectors*⁵³ on which operation by W produces a change in magnitude only. Let \mathbf{u} be such a vector. Then

$$W\mathbf{u} = \lambda\mathbf{u}, \tag{4.17}$$

where the factor λ gives the change in length; λ is called the *eigenvalue* of W belonging to the eigenvector \mathbf{u} . Clearly, \mathbf{u} and λ are special vectors and numbers associated with the matrix W . Equation (4.17) represents an *eigenvalue problem* for these quantities. We determine them as follows. Write (4.17) in component form,

$$\sum_l W_{kl}u_l = \lambda u_k. \tag{4.18}$$

This form shows explicitly that the eigenvector components u_k satisfy $3N$ *homogeneous* equations with constant coefficients since W has dimensions $3N \times 3N$. Therefore, a non-trivial solution ($u_k \neq 0$ for at least one value of λ) exists only if the determinant of these coefficients vanishes⁵⁴

$$\text{Det}(W - \lambda I) = \begin{vmatrix} W_{11} - \lambda & W_{12} & \dots \\ W_{12} & W_{22} - \lambda & \dots \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \cdot \end{vmatrix} = 0. \tag{4.19}$$

The determinantal condition is obviously equivalent to a $3N^{\text{th}}$ order polynomial in λ whose $3N$ roots (not necessarily all different or all real⁵⁵) determine the allowed values of λ in (4.17). The relation (4.19) is called the *secular determinant*⁵⁶ for the eigenvalues of W . Let us assume that these roots are all distinct for the moment. (Prob. 4-2 presents an opportunity for worrying about the case where the roots are not all distinct - the so-called degenerate case). Call these values $\lambda_i, i = 1, 2, \dots, 3N$. Then, substituting these values of λ back into (4.18) provides us with the *ratios* $u_1 : u_2 : \dots$ of the components of the eigenvector \mathbf{u} .

⁵³ See for example H. Margenau and G.M. Murphy, *The Mathematics of Physics and Chemistry*, O. van Nostrand and Company, Inc., New York 1947, p. 304.

⁵⁴ H. Margnau and G.M. Murphy, *ibid.*

⁵⁵ The roots are all real if, as supposed here, W is real and symmetric, see Prob. 4-1.

⁵⁶ The name originated in the perturbation methods of celestial mechanics.

There are thus $3N$ distinct eigenvectors (one for each λ_1) which we label as $\mathbf{u}^{(i)}$ as in the example of Sec. 4-2, equation (4.17) should thus read

$$W\mathbf{u}^{(i)} = \lambda_i \mathbf{u}^{(i)}, \quad i = 1, 2, \dots, 3N, \quad (4.20)$$

that is, this equation is satisfied by a *whole set* of eigenvectors $\mathbf{u}^{(i)}$ belonging to distinct eigenvalues λ_i .

Now construct the product WU where the matrix U has the $3N$ eigenvectors $\mathbf{u}^{(i)}$ as columns, $U_{ki} = u_k^{(i)}$:

$$U = \begin{pmatrix} u_1^{(1)} & u_1^{(2)} & \dots \\ u_2^{(1)} & u_2^{(2)} & \dots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}. \quad (4.21)$$

Then, using (4.20) repeatedly, there results

$$\begin{aligned} WU &= \begin{pmatrix} W_{11} & W_{12} & \dots \\ W_{12} & W_{22} & \dots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} u_1^{(1)} & u_1^{(2)} & \dots \\ u_2^{(1)} & u_2^{(2)} & \dots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} = \begin{pmatrix} \lambda_1 u_1^{(1)} & \lambda_2 u_1^{(2)} & \dots \\ \lambda_1 u_2^{(1)} & \lambda_2 u_2^{(2)} & \dots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \\ &= \begin{pmatrix} u_1^{(1)} & u_1^{(2)} & \dots \\ u_2^{(1)} & u_2^{(2)} & \dots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \dots \\ 0 & \lambda_2 & \dots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} = U\Lambda, \end{aligned} \quad (4.22)$$

where Λ is the diagonal matrix

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots \\ 0 & \lambda_2 & \dots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}.$$

Hence

$$U^{-1}WU = \Lambda, \quad \text{a diagonal matrix} \quad (4.23)$$

provided that U is non-singular, i.e. $\det U \neq 0$, so that the inverse U^{-1} can be found. Thus, as a byproduct of the eigenvalue problem for a square matrix W we have shown how to construct a matrix U that diagonalizes W according to the transformation (4.23). The diagonal elements of the transformed matrix are then just the eigenvalues of W .

The *reduction to diagonal form* of W given in (4.23) is unique except for the order (your choice!) in which the eigenvalues λ_i appear along the diagonal. Operations like $U^{-1}WU$ on W (whether or not they produce a diagonal matrix) are called *similarity transformations*.

Equation (4.23) almost solves our problem for a single square matrix. Except that the operation in (4.13) is of the form $U^T W U$, not $U^{-1}WU$. However, if W is a symmetric matrix, (its elements are symmetric about the principal diagonal), then the transpose W^T equals W . In this case $U^{-1} = U^T$, as we now show. Multiply (4.20) on the left with $\mathbf{u}^{(j)T}$, the transpose of the eigenvector with eigenvalue λ_j ,

$$\mathbf{u}^{(j)T} W \mathbf{u}^{(i)} = \lambda_i \mathbf{u}^{(j)T} \mathbf{u}^{(i)}. \quad (4.24)$$

Interchanging i and j gives another relation

$$\mathbf{u}^{(i)T} W \mathbf{u}^{(j)} = \lambda_j \mathbf{u}^{(i)T} \mathbf{u}^{(j)}. \quad (4.25)$$

On the other hand the transpose of (4.24) is directly

$$\mathbf{u}^{(i)T} W^T \mathbf{u}^{(j)} = \lambda_i \mathbf{u}^{(i)T} \mathbf{u}^{(j)}. \quad (4.26)$$

If W is symmetric as supposed, $W = W^T$, and the left hand sides of (4.25) and (4.26) become identical. Therefore, it follows that

$$(\lambda_i - \lambda_j) \mathbf{u}^{(i)T} \mathbf{u}^{(j)} = 0. \quad (4.27)$$

If there is no degeneracy, (i.e. $\lambda_i = \lambda_j$ only if $i = j$) then this relation implies that

$$\mathbf{u}^{(i)T} \mathbf{u}^{(j)} = \sum_{k=1}^{3N} u_k^{(i)} u_k^{(j)} = 0, \quad \text{if } \lambda_i \neq \lambda_j. \quad (4.28)$$

This equation expresses the *orthogonality* of two eigenvectors belonging to two different eigenvalues, i.e. they are perpendicular to each other in a multidimensional sense.

The matrix product of U in (4.21) and its transpose U^T therefore has the value

$$U^T U = \begin{pmatrix} u_1^{(1)} & u_2^{(1)} & \dots \\ u_1^{(2)} & u_2^{(2)} & \dots \\ \cdot & \cdot & \dots \\ \cdot & \cdot & \dots \\ \cdot & \cdot & \dots \end{pmatrix} \begin{pmatrix} u_1^{(1)} & u_1^{(2)} & \dots \\ u_2^{(1)} & u_2^{(2)} & \dots \\ \cdot & \cdot & \dots \\ \cdot & \cdot & \dots \\ \cdot & \cdot & \dots \end{pmatrix} = \begin{pmatrix} \mathbf{u}^{(1)T} \mathbf{u}^{(1)} & 0 & \dots \\ 0 & \mathbf{u}^{(2)T} \mathbf{u}^{(2)} & \dots \\ \cdot & \cdot & \dots \\ \cdot & \cdot & \dots \\ \cdot & \cdot & \dots \end{pmatrix}$$

using the result (4.28) repeatedly. The value of $\mathbf{u}^{(i)T} \mathbf{u}^{(i)}$ for each i is left undetermined by (4.28). This is in accord with our previous information that only the ratios $u_1^{(i)} : u_2^{(i)} : \dots$ of the components of each eigenvector are determined by the eigenvalue equation. Therefore $\mathbf{u}^{(i)T} \mathbf{u}^{(i)} = \sum_k u_k^{(i)2}$

is an *arbitrary constant* (its length) for each eigenvector that is not predetermined. If we *normalize* our eigenvectors so that each one has the *same* length then $U^T U$ becomes proportional to the unit matrix. Obviously, we can then go the whole way and choose this common length to be unity. The eigenvectors are then said to be normalized to unity; they are orthogonal by (4.28) in any event. Such a set of eigenvectors is termed *orthonormal*, and is free of arbitrary constants. Hence, if

$$\mathbf{u}^{(i)T} \mathbf{u}^{(j)} = \delta_{ij}, \quad (4.29)$$

the transformation matrix U built out of such eigenvectors is orthogonal, $UU^T = U^T U = I$, or

$$U^{-1} = U^T, \quad (4.30)$$

so that (4.11) can be met. Thus, (4.23) reads

$$U^T W U = \Lambda \quad (4.31)$$

in this case. This transformation of W is called an *orthogonal transformation*. Hence: real, symmetric matrices may always be diagonalized by an orthogonal transformation.

The importance of symmetry in determining the transformation matrix U is illustrated by comparing the diagonalization of

$$W = \begin{pmatrix} 1 & 6 \\ 6 & 1 \end{pmatrix} \quad \text{and} \quad W' = \begin{pmatrix} 1 & 9 \\ 4 & 1 \end{pmatrix}.$$

Both have the same characteristic equation for their eigenvalues, viz.

$$\lambda^2 - 2\lambda - 35 = 0, \quad \text{or } \lambda = 7 \text{ and } -5$$

and hence the same diagonal form

$$\begin{pmatrix} 7 & 0 \\ 0 & -5 \end{pmatrix}.$$

The two unnormalized eigenvectors for W are

$$\mathbf{u}^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{for } \lambda = 7, \quad \mathbf{u}^{(2)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{for } \lambda = -5, \quad (i)$$

and for W'

$$\mathbf{u}^{(1)'} = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \quad \text{for } \lambda = 7, \quad \mathbf{u}^{(2)'} = \begin{pmatrix} 3 \\ -2 \end{pmatrix} \quad \text{for } \lambda = -5, \quad (ii)$$

respectively. Notice that $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$ are orthogonal,

$$\mathbf{u}^{(1)T} \mathbf{u}^{(2)} = (1 \ 1) \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0$$

but that

$$\mathbf{u}^{(1)T} \mathbf{u}^{(2)'} = (3 \ 2) \begin{pmatrix} 3 \\ -2 \end{pmatrix} = 5 \neq 0.$$

The eigenvectors of W' are *not* orthogonal to each other.

The eigenvectors of W and W' in (i) and (ii) can still have arbitrary length. We noted the advantage of normalizing them in the case of orthogonal eigenvectors. Doing so for the set (i) we obtain the transformation matrix

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad U^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = U^{-1},$$

so that

$$\begin{aligned} U^{-1} W U &= U^T W U = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 6 \\ 6 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 7 & -5 \\ 7 & 5 \end{pmatrix} = \begin{pmatrix} 7 & 0 \\ 0 & -5 \end{pmatrix}, \end{aligned}$$

as expected. Since the eigenvectors of W' set (ii) are not orthogonal there is really no point in normalizing them. We construct the transformation matrix U' directly from (ii) as

$$U' = \begin{pmatrix} 3 & 3 \\ 2 & -2 \end{pmatrix}; \quad U'^{(-1)} = \frac{1}{12} \begin{pmatrix} 2 & 3 \\ 2 & -3 \end{pmatrix},$$

so that

$$\begin{aligned} U'^{-1} W' U' &= \frac{1}{12} \begin{pmatrix} 2 & 3 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} 1 & 9 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} 3 & 3 \\ 2 & -2 \end{pmatrix} \\ &= \frac{1}{12} \begin{pmatrix} 2 & 3 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} 21 & -15 \\ 14 & 10 \end{pmatrix} = \begin{pmatrix} 7 & 0 \\ 0 & -5 \end{pmatrix}. \end{aligned}$$

We also take this opportunity to point out that interchanging the columns of U or U' merely reorders the appearance of the roots 7 and -5 to give

$$\begin{pmatrix} -5 & 0 \\ 0 & 7 \end{pmatrix} \quad \text{instead of} \quad \begin{pmatrix} 7 & 0 \\ 0 & -5 \end{pmatrix}$$

for the diagonal form of W or W' . (Can the rows of U or U' also be interchanged?)

We see that a knowledge of the eigenvalues and eigenvectors of the matrix W solves the problem of bringing the Lagrangian (4.12) into normal form (4.6), with eigenfrequencies Ω_i

$$\Omega_i = \sqrt{\lambda_i}, \quad i = 1, 2, \dots, 3N, \quad (4.32)$$

where λ_i are the roots of the secular determinant, (4.19).

The normal coordinates themselves are given by (4.15) as a function of time. Introducing these solutions into (4.8) one finds

$$x'_k = \sqrt{m_k} x_k = \sum_{i=1}^{3N} u_k^{(i)} C_i \cos(\sqrt{\lambda_i} t + \delta_i) \quad (4.33)$$

in terms of the eigenvectors $\mathbf{u}^{(i)}$ of W , for the mass scaled x'_k and actual x_k displacements of the system. Here, $u_k^{(i)}$ is the k^{th} component of the i^{th} eigenvector of the matrix U as constructed in (4.21). In words: each coordinate x_k is a linear superposition of normal coordinates with relative amplitudes determined by the eigenvector components $u_k^{(i)}$. Thus, each x_k does *not* oscillate with a definite frequency. Rather their time-dependence depends in a complicated way on how many normal coordinates have been excited and with what amplitudes and phases.

Often only a single normal mode is excited. Then *all* the x_k oscillate with the common frequency $\sqrt{\lambda_i}$ of *this* normal mode, viz.

$$\sqrt{m_k} x_k(t) = u_k^{(i)} C_i \cos(\sqrt{\lambda_i} t + \delta_i) \quad (4.34)$$

if only ζ_i is excited. The system is then said to be vibrating in its i^{th} *normal mode*. Notice that in any normal mode vibration all particles pass through their equilibrium positions simultaneously.

4-5 Applications I

Free oscillations

Consider the longitudinal oscillations of a system of two particles, each mass m , joined to a center particle of mass M by identical springs (a mechanical model for a linear XY_2 molecule, e.g. CO_2). The system is shown in Fig. 4.2.

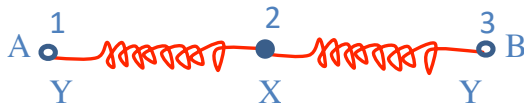


Figure 4.2: A linear XY_2 molecule.

Suppose that the spring constant for each spring is κ . The restoring forces for displacements of the particles along AB is taken to be proportional to the extensions in the springs. Hence, particles 1, 2 and 3 are subjected to forces

$$F_1 = -\kappa(x_2 - x_1), \quad F_2 = \kappa(x_1 - 2x_2 + x_3), \quad F_3 = -\kappa(x_3 - x_2),$$

where x_1, x_2 and x_3 are the displacements away from equilibrium of these particles. These forces are derivable from the potential function

$$V = \frac{1}{2} \frac{\kappa}{m} (x_1'^2 + 2x_2'^2 + x_3'^2 - 2x_1'x_2' - 2x_2'x_3'),$$

using the scaled coordinates $x_1' = \sqrt{m}x_1$, etc. of the problem. Thus, the W matrix reads

$$W = \begin{pmatrix} \frac{\kappa}{m} & -\frac{\kappa}{\sqrt{mM}} & 0 \\ -\frac{\kappa}{\sqrt{mM}} & \frac{2\kappa}{M} & -\frac{\kappa}{\sqrt{mM}} \\ 0 & -\frac{\kappa}{\sqrt{mM}} & \frac{\kappa}{m} \end{pmatrix}$$

for this potential. Its eigenvalue problem

$$W\mathbf{u} = \lambda\mathbf{u}$$

leads to the secular determinant

$$\begin{vmatrix} \frac{\kappa}{m} - \lambda & -\frac{\kappa}{\sqrt{mM}} & 0 \\ -\frac{\kappa}{\sqrt{mM}} & \frac{2\kappa}{M} - \lambda & -\frac{\kappa}{\sqrt{mM}} \\ 0 & -\frac{\kappa}{\sqrt{mM}} & \frac{\kappa}{m} - \lambda \end{vmatrix} = 0,$$

or

$$-\lambda\left(\frac{\kappa}{m} - \lambda\right)\left(\frac{\kappa}{m} + \frac{2\kappa}{M} - \lambda\right) = 0.$$

The roots are thus $\lambda_1 = 0$; $\lambda_2 = \kappa/m$, $\lambda_3 = \kappa/m + 2\kappa/M$. The associated normalized eigenvectors are obtained from the eigenvalue equation in the manner shown by the following calculations.

For $\lambda_1 = 0$,

$$\begin{pmatrix} \frac{\kappa}{m} & -\frac{\kappa}{\sqrt{mM}} & 0 \\ -\frac{\kappa}{\sqrt{mM}} & \frac{2\kappa}{M} & -\frac{\kappa}{\sqrt{mM}} \\ 0 & -\frac{\kappa}{\sqrt{mM}} & \frac{\kappa}{m} \end{pmatrix} \begin{pmatrix} \sqrt{m} \\ \sqrt{M} \\ \sqrt{m} \end{pmatrix} = 0, \quad \text{or } \mathbf{u}^{(1)} = \frac{1}{(2m + M)^{\frac{1}{2}}} \begin{pmatrix} \sqrt{m} \\ \sqrt{M} \\ \sqrt{m} \end{pmatrix}.$$

For $\lambda_2 = \kappa/m$,

$$\begin{pmatrix} 0 & -\frac{\kappa}{\sqrt{mM}} & 0 \\ -\frac{\kappa}{\sqrt{mM}} & \frac{2\kappa}{M} - \frac{\kappa}{m} & -\frac{\kappa}{\sqrt{mM}} \\ 0 & -\frac{\kappa}{\sqrt{mM}} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = 0, \quad \text{or } \mathbf{u}^{(2)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix},$$

and for $\lambda_3 = \kappa/m + 2\kappa/M$

$$\begin{pmatrix} -\frac{2\kappa}{M} & -\frac{\kappa}{\sqrt{mM}} & 0 \\ -\frac{\kappa}{\sqrt{mM}} & -\frac{\kappa}{m} & -\frac{\kappa}{\sqrt{mM}} \\ 0 & -\frac{\kappa}{\sqrt{mM}} & -\frac{2\kappa}{M} \end{pmatrix} \begin{pmatrix} \sqrt{M} \\ -2\sqrt{m} \\ \sqrt{M} \end{pmatrix} = 0, \quad \text{or } \mathbf{u}^{(3)} = \frac{1}{(2M + 4m)^{\frac{1}{2}}} \begin{pmatrix} \sqrt{M} \\ -2\sqrt{m} \\ \sqrt{M} \end{pmatrix}.$$

The eigenfrequencies of the system are thus

$$\Omega_1 = 0, \quad \Omega_2 = \sqrt{\kappa/m}, \quad \Omega_3 = \sqrt{\frac{\kappa}{m} + 2\frac{\kappa}{M}}.$$

The normal modes can be read off from the eigenvectors according to (4.34):

$$(1) \quad \Omega_1 = 0, \quad \sqrt{m}x_1 \sim \sqrt{m}\zeta_1, \quad \sqrt{M}x_2 \sim \sqrt{M}\zeta_1, \quad \sqrt{m}x_3 \sim \sqrt{m}\zeta_1,$$

or

$$x_1 = x_2 = x_3.$$

Therefore, all three particles suffer the same displacement. This mode corresponds to the displacement of the center-of-mass of the system as a whole without a change in the interparticle equilibrium distances. This cannot be a *vibrational* mode since the springs are not called into play. Indeed the equation of motion for ζ_1 confirms this:

$$\ddot{\zeta}_1 = 0, \quad \dot{\zeta}_1 = a, \quad \zeta_1 = a + bt$$

(a and b are constants). Therefore the center-of-mass moves with a constant velocity as it should.

$$(2) \quad \Omega_2 = \sqrt{\frac{\kappa}{m}}, \quad \sqrt{m}x_1 \sim \zeta_2, \quad \sqrt{M}x_2 \sim 0, \quad \sqrt{m}x_3 \sim -\zeta_2,$$

or

$$x_1 = -x_3, \quad x_2 = 0.$$

The outer particles move out of phase with equal displacements and the center particle is undisplaced.

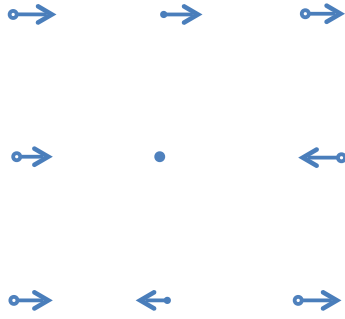
$$(3) \quad \Omega_3 = \sqrt{\frac{\kappa}{m} + 2\frac{\kappa}{M}}, \quad \sqrt{m}x_1 \sim \sqrt{M}\zeta_3; \quad \sqrt{M}x_2 \sim -2\sqrt{m}\zeta_3, \quad \sqrt{m}x_3 \sim \sqrt{M}\zeta_3,$$

or

$$x_1 = x_3, \quad x_2 = -2\frac{m}{M}x_1.$$

The outer particles move in phase with equal displacements while the center particle moves the fraction $2m/M$ of x_1 in the opposite direction, thus keeping the center-of-mass at rest. Notice that as $M/m \rightarrow \infty$ this mode describes two particles attached to an infinitely heavy center particle and vibrating in phase with frequency $\sqrt{\kappa/m}$ as one expects physically. The three normal modes are illustrated in Fig. 4.3.

Actually, we could have decided from the beginning that a linear XY_2 molecule has one zero frequency and two non-zero frequencies for longitudinal vibrations. We have just seen that the zero frequency motion pattern corresponds to the motion of the center-of-mass only. But (4.6) showed that there are as many frequencies Ω_i as there are degrees of freedom. Therefore one degree of freedom for XY_2 must correspond to a translation of the center-of-mass and this must be described by a normal coordinate of zero frequency since this normal coordinate *cannot* appear in L (otherwise it would spoil the momentum conservation). The remaining two degrees of freedom then possess normal coordinates

Figure 4.3: Normal modes of a linear XY_2 molecule.

with non-zero frequencies. In general then, an interacting system of N particles possesses $3N - 6$ non-zero eigenfrequencies and 6 zero eigenfrequencies corresponding to three coordinates describing the translation of the center-of-mass of the system in three dimensions and three coordinates describing the rotation of the system as a whole about its center-of-mass. None of these motions require the interparticle spacing to change but are possible motions, and hence must belong to zero eigenfrequencies of the system. The normal coordinates describing such motions must therefore reside completely in the kinetic energy of the system. The expression for L in (4.6) is thus more correctly written as $L_{int} + L_0$, where

$$L_{int} = \sum_{i=1}^{3N-6} \frac{1}{2} \dot{\zeta}_i^2 - \sum_{i=1}^{3N-6} \frac{1}{2} \Omega_i^2 \zeta_i^2, \quad (4.35)$$

describes the intrinsic vibration of the system, L_0 its translational and rotational energy.

4-6 Applications II

Forced oscillations

Let us briefly consider the situation where an external field $V_{ext}(x_1, x_2, \dots, t)$ is admitted to the problem. The Lagrange function (4.1) then becomes

$$L = \sum_k \frac{1}{2} m_k \dot{x}_k^2 - V(x_1, x_2, \dots) - V_{ext}(x_1, x_2, \dots, t).$$

Expanding this entire expression about the equilibrium configuration of the system again leads to

$$L \simeq T - V + \sum_k F_k^{(e)}(x_k - x_k^0),$$

where $T - V$ is the free oscillation Lagrangian of (4.5) and

$$F_k^{(e)} = -\left(\frac{dV_{ext}}{dx_k}\right)_{x_k=x_k^0}$$

the external force component on the motion described by coordinate x_k at equilibrium. Transforming to normal coordinates we find immediately that

$$L = \sum_i \left(\frac{1}{2} \dot{\zeta}_i^2 - \frac{1}{2} \omega_i^2 \zeta_i^2 \right) + \sum_i f_i^{(e)} \zeta_i,$$

where

$$f_i^{(e)} = \sum_k F_k^{(e)} \frac{U_{ki}}{\sqrt{m_k}}$$

is identified by the following equation of motion as the external force component on the i^{th} normal coordinate:

$$\ddot{\zeta}_i + \omega_i^2 \zeta_i = f_i^{(e)}.$$

Thus, the normal coordinates obey the equation of motion of a *driven oscillator* if external forces are present. To illustrate the effect of the $f_i^{(e)}$ suppose that they are periodic, with a common frequency ω . Then,

$$f_i^{(e)} = f_i^{(e)}(0) e^{-i\omega t},$$

where the $f_i^{(e)}(0)$ are constants. (Again the real or imaginary part of this expression may be used by those unwilling to accept the advantages of the complex notation!) With this force component acting, ζ_i has the special solution that oscillates with the frequency of the applied field:

$$\zeta_i^{(s)} = \frac{f_i^{(e)}(0) e^{-i\omega t}}{\omega_i^2 - \omega^2}.$$

Notice that the i^{th} normal mode can only be excited in this fashion if $F_k^{(e)}$ has a component "along" it, i.e. $f_i^{(e)} \neq 0$. The complete solution of the motion is obtained by adding to $\zeta_i^{(s)}$ any solution of the homogeneous equation (4.14) that results by setting all f_i equal to zero. If, as is usually the case, the homogeneous equation contains a damping term this part of the solution dies out in time and only $\zeta_i^{(s)}$ remains. For this reason $\zeta_i^{(s)}$ is called the *steady-state* solution (hence the superscript) in the presence of an external driving force.

Thus the steady-state solution tells us how the system responds to an external field. In fact this "response" is very useful in studying the eigenfrequencies of a system. Suppose for simplicity we have a linear array of N particles subjected to a common force $F_k^{(e)} \exp(-i\omega t) = F_0 \exp(-i\omega t)$ for all k (such as an electric field acting on a system of identically charged particles) and we determine its response to this field by measuring the *average* displacement of the system,

$$\langle x \rangle = \frac{1}{N} \sum_k x_k = \frac{1}{N} \sum_k \frac{U_{ki}}{\sqrt{m_k}} \zeta_i^{(s)}.$$

Filling in the value of $\zeta_i^{(s)}$ we find

$$\langle x \rangle = F_0 e^{-i\omega t} \sum_i \frac{a_i^2}{\omega_i^2 - \omega^2},$$

where $a_i = (1/\sqrt{N}) \sum_k U_{ki} / \sqrt{m_k}$. The ratio $\langle x \rangle / F_0$ of the maximum amplitude induced in $\langle x \rangle$ by a force strength F_0 is called a *response function* $\chi(\omega)$, or

$$\chi(\omega) = \sum_i \frac{a_i^2}{\omega_i^2 - \omega^2}.$$

We notice that $\chi(\omega)$ refers to properties of the *free* system only; it is independent of the exciting force. The important feature about $\chi(\omega)$ is that it has *poles* as a function of the applied frequency at the eigenfrequencies of the system under study (unless the corresponding a_i is zero of course). The measurement of $\chi(\omega)$ is thus one experimental way of probing the eigenfrequencies of an oscillatory system.

4-7 Applications III

Vibrations of non-linear molecules

Consider the "vibrating triangle" molecule X_3 as pictured in Fig. 4.4, consisting of three identical atoms situated at the vertices of an equilateral triangle. Label each vertex as 1,2 or 3 and set up parallel $x_p - y_p$ coordinate axes as shown where $p = 1, 2, 3$. The displacements of the atom in position p (note: the *position* not the atom is labelled) are then (x_p, y_p) or \mathbf{x}_p for short.

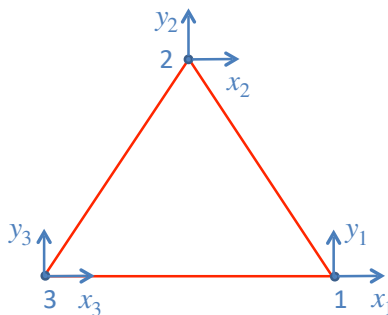


Figure 4.4: The molecule X_3 .

Assuming that the potential energy of the system depends quadratically on the extensions δl_{pq} of the sides of the triangle (linear restoring forces), one has

$$V = \frac{1}{2} \kappa [(\delta l_{12})^2 + (\delta l_{23})^2 + (\delta l_{31})^2]$$

with

$$\begin{aligned}\delta l_{12} &= \frac{1}{2}(x_1 - x_2) - \frac{\sqrt{3}}{2}(y_1 - y_2) \\ \delta l_{23} &= \frac{1}{2}(x_2 - x_3) + \frac{\sqrt{3}}{2}(y_2 - y_3) \\ \delta l_{31} &= x_1 - x_3.\end{aligned}\quad (4.36)$$

κ is the coupling constant. We can now substitute these expressions into the potential energy V , but the calculation requires courage. Instead, let us introduce new coordinates Q_i via the transformation

$$\{x_1, y_1, x_2, y_2, x_3, y_3\} = U\{Q_{A1}, Q_{A2}, Q_{E1}, Q_{E2}, Q'_{E1}, Q'_{E2}\}, \quad (4.37)$$

where

$$U = \frac{1}{2\sqrt{3}} \begin{pmatrix} \sqrt{3} & 1 & 2 & 0 & 2 & 0 \\ -1 & \sqrt{3} & 0 & -2 & 0 & 2 \\ 0 & -2 & -1 & -\sqrt{3} & 2 & 0 \\ 2 & 0 & -\sqrt{3} & 1 & 0 & 2 \\ -\sqrt{3} & 1 & -1 & \sqrt{3} & 2 & 0 \\ -1 & -\sqrt{3} & \sqrt{3} & 1 & 0 & 2 \end{pmatrix}, \quad (4.38)$$

which is readily demonstrated to be orthogonal, $U^{-1} = U^T$. Performing the transformation to the Q , we find that the Lagrange function for the X_3 molecule reads

$$L = \frac{1}{2}m\{\dot{Q}_{A1}^2 + \dot{Q}_{A2}^2 + \dot{Q}_{E1}^2 + \dot{Q}_{E2}^2 + \dot{Q}'_{E1}{}^2 + \dot{Q}'_{E2}{}^2\} - \frac{1}{2}\left(\frac{3\kappa}{2}\right)\{2Q_{A1}^2 + Q_{E1}^2 + Q_{E2}^2\} \quad (4.39)$$

if m is the mass of any atom. Thus, L is already in normal form. Only Q_{A1} , Q_{E1} and Q_{E2} appear in the potential energy. They must therefore be normal coordinates for vibrations. The frequencies are seen to be

$$\begin{aligned}\Omega_1 &= \sqrt{\frac{3\kappa}{m}} \quad \text{for } Q_{A1} \\ \Omega_2 &= \Omega_3 = \sqrt{\frac{3\kappa}{2m}} \quad \text{for } Q_{E1} \text{ and } Q_{E2}.\end{aligned}\quad (4.40)$$

Notice that the coordinates Q_{E1} and Q_{E2} vibrate with the same frequency, i.e. they are *degenerate*. The normal mode patterns are shown in Fig. 4.5(a), (b), and (c) as deduced from (4.3). The arrows indicate the amount and direction of displacement of each atom. The remaining coordinates Q_{A1} , Q'_{E1} , and Q'_{E2} have zero frequency and thus correspond to translations and rotations of the molecular framework as a whole. This expectation is borne out in Fig. 4.5(d), (e) and (f).

Thus, the problem is solved. We have given the solution in the above form, depending as it does on the pivotal transformation matrix U of

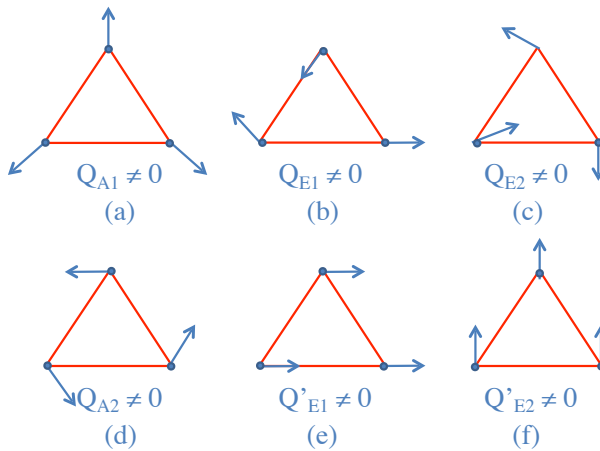


Figure 4.5: Normal modes of an X_3 molecule.

(4.38), in order to encourage the ambitious reader to tackle the next section with some sense of purpose. The point is: coordinates like the Q_i (even the curious subscripts used in the above problem has good reason) can often be determined from the *symmetry* that the molecule possesses in its undisturbed state. For instance, the X_3 molecule is unaffected by (a) rotating it through $\pm 120 (= \pm 2\pi/3)$ about an axis perpendicular to the molecular plane and passing through the center-of-mass, or (b) reflecting the structure in plane mirrors bisecting the three vertices of the triangle. The full set of such operations the symmetry of a molecule allows forms a *group*, the so-called *symmetry group* of the molecule. To benefit from this fact requires some of the machinery of a branch of mathematics called *group theory*. We consider some salient points of this subject in the following sections.

4-8 Point Groups and their Representations

It is *not* our intention in this short synopsis of the theory of point groups to give anything like a complete discussion of this important subject, but rather to provide a sufficient background for the more ambitious reader to appreciate the elegance of this powerful tool in the analysis of molecular vibrations. Most of the important theorems will simply be stated without proof. A rigorous discussion of all such theorems may, for example, be found in E.P. Wigner, *Group Theory* (Academic Press, 1959, translated from the German by J.J. Griffin).

A few definitions are in order: (i) A collection of objects A, B, C, \dots called *elements* form a *group* if the following four conditions are obeyed by every element of the group:

1) a law of combination called the *product*, is prescribed such that the product of any two elements produces a third element that also falls in

the group. We write the combination of A and B as $AB = C$ and call AB the product of A and B . Note carefully that ordinary multiplication is not implied (or excluded) by this definition of product.

- 2) the set of elements includes a unit element E such that $EA = AE = A$ holds for every element A of the group.
- 3) every element A of the group has an inverse element A^{-1} , also in the group, such that $XA = AX = E$.
- 4) the associative law holds for group multiplication: $A(BC) = (AB)C = ABC$.

Notice that the group multiplication need not (but can be) commutative, i.e.

$$AB \neq BA \tag{4.41}$$

in general. If an equality sign holds in (4.41) for every pair of elements of the group, it is called an *abelian* group.

The number of elements g in a group is called its *order*. It is possible to have groups with a finite or an infinite number of elements.

Exercise. Examine (i) the set of positive integers $1, 2, \dots, \infty$ and (ii) the set of elements $E = 1, B = -1, C = i, D = -i$ for group properties, choosing ordinary multiplication for the group multiplication in both instances.

The group multiplication among the elements of a group is best displayed by means of a group multiplication table like that shown in Table 4.1.

The table is constructed (and used) by reading off the intersection of the row and column identified by the two elements whose product is sought, e.g.

$$BC = D$$

from the intersection of row B and column C . Notice that each element appears once and only once in each row or column. This is a general group property (see Prob. 4-3).

Table 4.1: A group multiplication table for example (ii) of the text.

	E	B	C	D
E	E	B	C	D
B	B	E	D	C
C	C	D	B	E
D	D	C	E	B

The concept of conjugate elements and class

Two elements A and B are said to be *conjugate* if they are connected by a relation

$$A = C^{-1}BC,$$

where C is some element of the group. Collecting all conjugate elements together, we form a *class* of the group. We state without proof that the elements of a group can be segregated into various classes by pre- and postmultiplying each element with every element of the group such that no element appears in more than one class (see Prob. 4-4). Taking, for example, the group whose multiplication table is displayed in Table 4-1,

one finds each element is in a class by itself since the group is abelian (Prob. 4-5.) The *number* of classes a group possesses will be important in our applications later on.

(iii) Group representations. A very useful concept in group theory is that of a *representation* of the elements of a group by means of a group of square matrices that have the same *matrix multiplication table* as the multiplication table of the group they represent.

Thus, if $\Gamma(A)$ and $\Gamma(B)$ are matrices that are to represent the elements A and B of a group and $AB = C$, where C is another element in the group, then

$$\Gamma(A)\Gamma(B) = \Gamma(C) \quad (4.42)$$

must hold, where $\Gamma(C)$ is a matrix representing the element C . The law of combination in (4.42) is ordinary matrix multiplication.

The use of matrices to represent non-abelian groups is of course essential in order to mirror the non-commutative nature of the former. We make three further remarks about the matrices $\Gamma(A), \Gamma(B), \dots$. First, it is entirely possible for one matrix to represent more than one element of the group. Secondly the matrices $\Gamma(A), \Gamma(B), \dots$ are by no means unique. For if they form a representation then so do

$$U^{-1}\Gamma(A)U, \quad U^{-1}\Gamma(B)U, \quad \dots \quad (4.43)$$

where U is any square, invertible matrix. The matrices in (4.43) (call them $\Gamma(A), \Gamma(B), \dots$) form an *equivalent* representation. Equivalent representations are not counted as different representations of a group. The third point concerns the dimension of the matrices Γ . If $\Gamma'(A)$ and $\Gamma''(A)$ are representations of a group element A (equivalent or not) then so is

$$\Gamma(A) = \begin{pmatrix} \Gamma'(A) & 0 \\ 0 & \Gamma''(A) \end{pmatrix}, \quad (4.44)$$

as follows immediately from the rules of matrix multiplication. A representation like (4.44) that breaks down into matrices of smaller dimension is termed *reducible*. A similar property may hold for the constituent matrices in $\Gamma(A)$. The matrix $\Gamma(A)$ is reducible by construction. But a *similarity transformation* on $\Gamma(A)$ would generally destroy its "block" form by scrambling its rows and columns without, however, upsetting its reducibility. The property of being reducible is then by no means obvious and has to be tested for by group theoretic procedures that we indicate presently.

The dimension of the matrices providing a representation of a group is called the *dimension of the representation*. Clearly all the matrices forming a particular representation of a group must share this dimension. We denote the *collection* of such matrices by the single symbol Γ , and refer to this collection as the representation Γ .

Let us return to the property of reducibility. Given a representation Γ we say this representation is *reducible* if *all* these matrices can be brought into a block diagonal form of (4.44):

$$\Gamma(R) = \begin{pmatrix} D^{(1)}(R) & & & \\ & D^{(2)}(R) & & \\ & & D^{(3)}(R) & \\ & & & \dots \end{pmatrix}, \quad (4.45)$$

with the *same* similarity transformation for every element R in the group. If this cannot be done, the representation is *irreducible*. Irreducible representations will play a special role in our further considerations.

The submatrices $D^{(\mu)}(R)$ along the diagonal of $\Gamma(A)$ in (4.45) are all group representations of smaller dimension than $\Gamma(A)$. If they are in turn reducible a further similarity transformation is carried out reducing the matrix $\Gamma(A)$ still further. If no further such transformations are possible the representation Γ is said to be fully reduced. Suppose this is the case in (4.45). Then, the $D^{(\mu)}(A), D^{(\mu)}(B), \dots$ are termed the *irreducible* representations of the group.

From now on we reserve the symbol $D^{(\mu)}(R)$, with the superscript μ for the μ^{th} irreducible representation of the element R , where R stands for an arbitrary element of the group, and refer to the collection of such matrices as the representation $\Gamma^{(\mu)}$ of the group. The index μ thus distinguishes between different irreducible representations of the group. But what does "different" mean when applied to a representation? The set of matrices $\Gamma^{(\mu)}$ are clearly determined only up to a similarity transformation so that equivalent representations cannot be considered different. The remaining possibilities are (i) that the dimensions of the various representations $\Gamma^{(\mu)}$ differ, or (ii) if the dimensions are the same, that such representations are *inequivalent*. Thus, for two representations $\Gamma^{(1)}, \Gamma^{(2)}$ to differ they must either differ in dimension, or be inequivalent if they have the same dimension.

In the process of reducing Γ it is entirely possible that the same irreducible representation appears *more than once* in the reduced form; nor need all the possible irreducible representations of the group be present. To emphasize these points the reduction process is usually displayed by writing $\Gamma(R)$ as

$$\Gamma(R) = a_1 D^{(1)}(R) \oplus a_2 D^{(2)}(R) \oplus \dots = \sum_{\mu} a_{\mu} D^{(\mu)}(R) \quad (4.46)$$

for each group element R separately, or collectively as

$$\Gamma = \sum_{\mu} a_{\mu} \Gamma^{(\mu)},$$

where a_{μ} is an integer giving the number of times the representation $\Gamma^{(\mu)}$ appears in the reduction of Γ . Neither of these equalities should be

interpreted as sums in an arithmetic sense. They are simply a symbolic notation for how the representation of Γ is composed of other representations.

Next follow, without complete proofs, some theorems for the irreducible representations of *finite* groups:

Theorem I: a representation with square matrices having non-zero determinants is always equivalent to a representation with unitary matrices⁵⁷.

⁵⁷ A unitary matrix U has its inverse equal to its hermitian conjugate of the complex conjugate U^* of U .

Theorem II: (the Great Orthogonality Theorem): the set of inequivalent, unitary, irreducible matrices $D^{(\mu)}(R)$ of dimension n_μ that represent a group of order g , satisfy the relation

$$\sum_R D^{(\mu)*}(R)_{\alpha\beta} D^{(\nu)}(R)_{\gamma\delta} = \frac{g}{n_\mu} \delta_{\mu\nu} \delta_{\alpha\gamma} \delta_{\beta\delta}, \tag{4.47}$$

where the sum is over all the elements of the group. Here $D^{(\mu)}(R)_{\alpha\beta}$ is the $\alpha\beta^{\text{th}}$ element of the matrix $D^{(\mu)}(R)$, and the asterisk (*) denotes complex conjugation.

The matrix elements $D^{(\mu)}(R)_{\alpha\beta}$, for fixed $\mu, \alpha\beta$ can be viewed as the g components of a "vector"

$$\{D^{(\mu)}(A)_{\alpha\beta}, D^{(\mu)}(B)_{\alpha\beta}, \dots\}$$

in the space of the group elements, as R runs through all the elements of the group. Then, (4.47) says that these vectors form an orthogonal set. Since the triad $\mu, \alpha\beta$ distinguishes the different vectors, there are n_μ^2 of them per representation (i.e. for given μ) or

$$\sum_{\mu=1}^p n_\mu^2$$

in all, if the group possesses p distinct irreducible representations. But we cannot have more orthogonal vectors than the dimension of the space they span. Hence,

$$\sum_{\mu} n_\mu^2 \leq g.$$

Additional arguments have to be invoked (see, for example, E.P. Wigner, *ibid.*, p.115) to show that the equality sign holds, leading to Theorem III,

$$n_1^2 + n_2^2 + \dots + n_p^2 = g \tag{4.48}$$

as a fundamental relation between the *order* of a group and the *dimensions* of its various irreducible representations. This remarkable theorem often allows one to determine these dimensions (i.e. the numbers n_μ) uniquely if the number p of different irreducible representations is known. However, this latter number can be found as follows: The procedure makes use of the concept of the *character* $\chi(R)$ of the representation

of a group element R . This is simply defined as the trace of the matrix that represents R :

$$\chi(R) = \sum_{\alpha} \Gamma(R)_{\alpha\alpha},$$

the sum extending over all diagonal elements of the matrix $\Gamma(R)$. If the representation is an irreducible one, we indicate this by appending μ to $\chi(R)$ thus: $\chi^{(\mu)}(R) = \sum D^{(\mu)}(R)_{\alpha\alpha}$.

Since the trace of any square matrix is unaffected by a similarity transformation of that matrix, it is clear that group elements in the same *class* have the same character.

Now, we get a theorem about the character of a group by specializing Theorem II as follows: set $\alpha = \beta, \gamma = \delta$ and sum on α and γ . Then,

$$\sum_{\alpha, \gamma} \sum_R D^{(\mu)*}(R)_{\alpha\alpha} D^{(\nu)}(R)_{\gamma\gamma} = \frac{g}{n_{\nu}} \delta_{\mu\nu} \sum_{\alpha, \gamma} \delta_{\alpha\gamma},$$

or

$$\sum_R \chi^{(\mu)*}(R) \chi^{(\nu)}(R) = g \delta_{\mu\nu}, \quad (4.49)$$

since $\sum_{\alpha, \gamma} \delta_{\alpha\gamma} = \sum_{\alpha} 1 = n_{\mu}$. Thus, the characters $\chi^{(\mu)}(R)$ for various μ are also "orthogonal vectors" in group-element space. However, these vectors are not all different since we saw above that elements in the same class have the same character. If, however, we sum over all *classes* in (4.49) (C_k is any element of the k^{th} class containing N_k elements), then

$$\sum_{k=1} N_k \chi^{(\mu)*}(C_k) \chi^{(\nu)}(C_k) = g \delta_{\mu\nu}. \quad (4.50)$$

Thus, the characters $\chi^{(\mu)}$ also form an orthogonal set of vectors in *class-space*. There must be p such vectors (one for each irreducible representation) so that

$$p \leq \text{number of classes}$$

The equality sign is found to hold again, and so Theorem IV: the number of distinct irreducible representations

$$p = \text{number of classes in the group}. \quad (4.51)$$

No less remarkable than (4.48), this relation determines the number of irreducible representations of a group as the number of classes it contains. Armed with (4.48) and (4.51) it is usually possible to determine the dimensions n_{μ} of these irreducible representations uniquely⁵⁸.

We have just seen how the number of irreducible representations of a group and their dimensions may be found. A related question is *how many times* a given irreducible representation $\Gamma^{(\mu)}$ appears in the reduction of an arbitrary reducible representation Γ or the group, i.e. the determination of the integers a_{μ} in (4.46). The a_{μ} can be found once the

⁵⁸ This statement is true for all the groups of interest in molecular vibration problems.

characters of the representation Γ are known. We see from (4.45) that the character of a typical element R is given by

$$\chi(R) = \sum_{\mu=1}^p a_{\mu} \chi^{(\mu)}(R)$$

(p is as before the number of different irreducible representations) in terms of the characters $\chi^{(\mu)}(R)$ of the irreducible representations. Multiply this equation on the left with a particular $\chi^{(\mu)*}(R)$ and sum over all elements R of the group. Then,

$$g a_{\mu} = \sum_R \chi^{(\mu)*}(R) \chi(R)$$

on using the orthogonality of the $\chi^{(\mu)}$. Since the characters of all elements in a class are the same we can sum over classes instead and obtain

$$a_{\mu} = \frac{1}{g} \sum_{k=1}^p \chi^{(\mu)*}(C_k) \chi(C_k) N_k, \quad (4.52)$$

using the same notation as in (4.50).

	$N_1 C_1$	$N_2 C_2$	$N_3 C_3$
$\Gamma^{(1)}$	$\chi^{(1)}(C_1)$	$\chi^{(1)}(C_2)$	$\chi^{(1)}(C_3)$
$\Gamma^{(2)}$	$\chi^{(2)}(C_1)$	$\chi^{(2)}(C_2)$	$\chi^{(2)}(C_3)$
$\Gamma^{(3)}$	$\chi^{(3)}(C_1)$	$\chi^{(3)}(C_2)$	$\chi^{(3)}(C_3)$
...

Table 4.2: A character table.

(iv) The character table. From the foregoing it is clear that the characters $\chi^{(\mu)}(R)$ play an important role in characterizing a group representation (hence their name!). Furthermore, the $\chi^{(\mu)}(R)$ are insensitive to similarity transformations that produce equivalent representations, so that this non-uniqueness is not present in the characters. The characters of the various irreducible representations $\Gamma^{(\mu)}$ of a group are usually displayed in the form of a *character table* like that shown in Table ??.

The left hand column lists the different irreducible representations, the top row a typical element of each class, preceded by the number of elements N_k in that class. The characters $\chi^{(\mu)}(C_k)$ are then distinguished according to representation (μ) and class (k) within that representation.

4-9 Symmetry Operations

In this Section we examine more closely the concept *symmetry operation* that was mentioned at the end of Sec. 4-7. Consider an arbitrary polyatomic molecule in its equilibrium state. Let us be interested in the displacement⁵⁹ of this structure as a whole about a fixed point O in

⁵⁹ The word "displacement" is used here in a purely geometrical sense of moving the molecule as a rigid structure about its center-of-mass.

space (usually the center-of-mass), the atoms remaining in their equilibrium positions. The orientation in space of the molecular framework is fixed by fixing the positions of any three non-collinear atoms. The only geometrical displacements that are possible are therefore rotations of this structure about an axis passing through the fixed point and reflections of this structure in a plane containing the fixed point. Let us consider these two operations in more detail.

(i) Rotations. Denote the operation of rotating the molecular framework through an angle α about a direction \mathbf{e}_3 (usually the z direction) by $C(\alpha)$. The effect of $C(\alpha)$ on a point P attached to the framework is shown in Fig. 4.6.

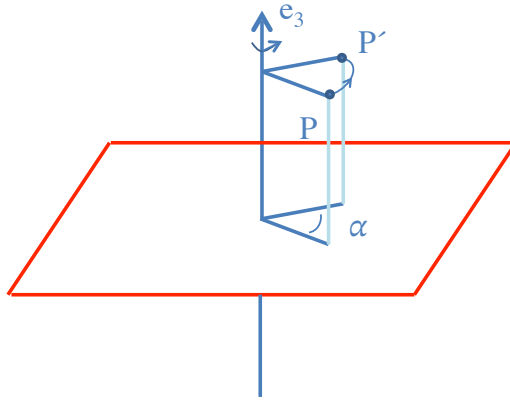


Figure 4.6: A rotation about \mathbf{e}_3 through an angle α .

The combination (product) of two rotations about the same axis is again a rotation by the sum of the angles. Furthermore, such operations commute:

$$C(\alpha)C(\beta) = C(\beta)C(\alpha) = C(\alpha + \beta). \quad (4.53)$$

A particular case of this relation arises if $\beta = -\alpha$. Then,

$$C(-\alpha)C(\alpha) = C(\alpha)C(-\alpha) = C(0) = E,$$

where E is the identity operation (the process of doing nothing). Thus, $C(-\alpha)$ is the *inverse* of the operation $C(\alpha)$,

$$C^{-1}(\alpha) = C(-\alpha), \quad (4.54)$$

a result that is geometrically obvious.

Often the rotation angle of interest is a rational fraction of 2π , $\alpha = 2\pi/n$, $n = 1, 2, \dots$. Then it is more convenient to write

$$C\left(\frac{2\pi}{n}\right) = C_n, \quad (4.55)$$

and specify n instead of α . We notice that performing the operation C_n n

times is equivalent to the identity operation:

$$\underbrace{C_n C_n \dots}_{n \text{ factors}} = C_n^n = C\left(n \frac{2\pi}{n}\right) = C(2\pi) = E,$$

that is, the operations (including the identity!)

$$C_n, C_n^2, \dots, C_n^{n-1}, C_n^n = E \quad (4.56)$$

are simply repeated in the set of operations

$$C_n^{n+1}, \dots, C_n^{2n-1}, C_n^{2n}.$$

In addition each operation in (4.56) has an inverse, and the triple product of any three operations obeys the associative law. Hence, the set of operations $C_n, C_n^2, \dots, C_n^{n-1}, C_n^n = E$ forms a *group* of order n in the sense of Sec. 4-6. This group of rotation elements about a common axis is usually denoted by C_n . It is an abelian group.

(ii) Reflections. Consider next the geometrical operation of a reflection of the molecular framework in a plane passing through O . Denote the plane and the operation of reflection in it by the symbol σ . The effect of σ on a point P is shown in Fig. 4.7. This figure shows two planes that are of particular interest: (a) a "vertical" plane σ_v passing through O that contains the rotation axis e_3 ; (b) the "horizontal" plane σ_h that also passes through O and is perpendicular to σ_v .

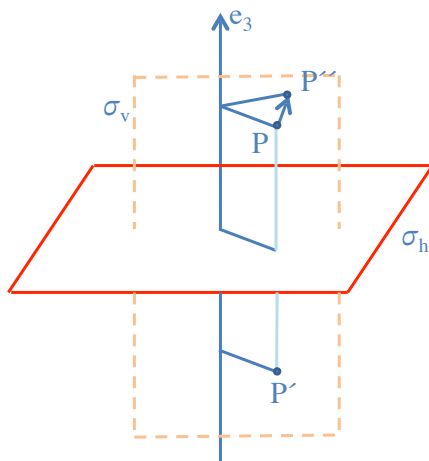


Figure 4.7: Reflections. P is taken into P' by σ_h , into P'' by σ_v .

The repetition of any reflection restores a structure to its original position. Therefore, a reflection is also its own inverse, so that

$$\sigma^2 = E, \quad \sigma^{-1} = \sigma. \quad (4.57)$$

So far the geometrical operations we have discussed in no way relate to the specific geometric shape of a particular molecule. However,

among all these operations of rotations, reflections, etc. one might find certain special ones that have the property of bringing the equilibrium framework of a specific molecule *into coincidence with itself*, that is, one is unable to tell whether these operations have been performed or not by just examining the molecular framework before and after the operation. Such operations are called *symmetry operations* and the corresponding symmetry in the molecule that admits them a *symmetry element* of the molecule. One thus has symmetry elements consisting of axes of rotation, planes of reflection, etc., depending on the equilibrium structure of the molecule being examined.

Take, for example, the X_3 molecule of Sec. 4-5. Considered as a plane structure in two dimensions its structure admits as symmetry operations (Fig. 4.8) (i) rotations by $\pm 2\pi/3$ (symbols C_3 and C_3^{-1}) about an axis \mathbf{e} perpendicular to the molecular plane and passing through the center-of-mass and (ii) reflections in three "vertical" planes (symbols $\sigma_1, \sigma_2, \sigma_3$) intersecting along \mathbf{e} and bisecting the bond angles at each X atom. The set of symmetry operations

$$E, C_3, C_3^{-1}, \sigma_1, \sigma_2, \sigma_3 \quad (4.58)$$

that includes the identity operation $E = C_3^3$ form a group, the so-called *symmetry group of the molecule* X_3 (Prob. 4-6). We will see that this symmetry group already predetermines to a large extent how a molecule can vibrate.

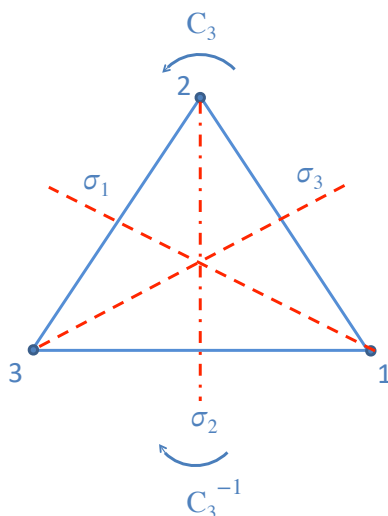


Figure 4.8: Symmetry operations for the X_3 molecule.

Returning to the symmetry group of X_3 in (4.58), (usually denoted by C_{3v} in the literature) we record in Table 4.3 its character table from D. Schonland⁶⁰, p.281. According to Table 4.3 the group C_{3v} has three classes consisting of the identity operation E , the two rotations C_3 and

⁶⁰ D. Schonland
Molecular Symmetry,
D. van Nostrand
Company Ltd.,
London and New
York, 1965.

C_3^{-1} , and the three reflections σ_1, σ_2 , and σ_3 respectively. Therefore this group has three irreducible representations (theorem IV). Calling their dimensions n_1, n_2 and n_3 we know that

$$n_1^2 + n_2^2 + n_3^2 = 6,$$

since the group has six elements (theorem III). This relation can be satisfied in only one way, viz.

$$n_1 = n_2 = 1, \quad n_3 = 2.$$

	E	$2C_3$	$3\sigma_v$
A_1	1	1	1
A_2	1	1	-1
E	2	-1	0
Γ	6	0	0
Γ_t	2	-1	0
Γ_r	1	1	-1
Γ_{vib}	3	0	1

Table 4.3: Character table for C_{3v} .

Therefore, two of the irreducible representations of C_{3v} , are one dimensional (called A_1 and A_2 in Table 4.3), the third one has dimension two (called E in Table 4.3 and *not* to be confused with the identity operation!).

The knowledge that X_3 has the symmetry group C_{3v} and that this group has three irreducible representations plays an important role in the analysis of the normal modes of X_3 . But to appreciate the procedure we need to know more about group representations and how they are found.

4-10 Group Representations.

We pointed out in Chap. 3 that the rotation of any vector \mathbf{Q} into a new vector \mathbf{q} by an operator A could be written

$$q_i = \sum_j A_{ij} Q_j, \tag{4.59}$$

where the matrix array

$$A_{ij} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$$

was called a representation of A in the basis vectors used as coordinate directions for the components of \mathbf{Q} and \mathbf{q} . A second rotation B carries \mathbf{q} into \mathbf{q}' , or

$$q'_i = \sum_j B_{ij} q_j = \sum_{jk} B_{ij} A_{jk} Q_k.$$

Thus the single rotation that carries \mathbf{Q} directly over into \mathbf{q}' is represented by the matrix product

$$C_{ik} = \sum_j B_{ij} A_{jk}$$

in terms of the individual representations of A and B . This relation shows that matrix arrays like (4.59) qualify as *matrix representations* for the group of operators A, B, C, \dots in the sense of Sec. 4-6 (iii).

The representations of rotations are now very easily found by choosing for \mathbf{Q} the position vector $\mathbf{r} = (x, y, z)$ of a point P in space and observing how P moves under a rotation. In fact we unwittingly found a representation of the rotation described by the three Euler angles (ϕ, θ, ψ) by this method in (3.34) of Chap. 3. With one important difference. Here we are moving the vector, not the coordinate system so that the angles in (3.31) *et seq.* must be replaced by their negatives for use in the present chapter. Let us find two important representations by considering the displacement of a point P in two dimensions.

(a) Representation of a rotation through an angle α about the z axis (the \mathbf{e}_3 direction). This operation is shown in Fig. 4.9.

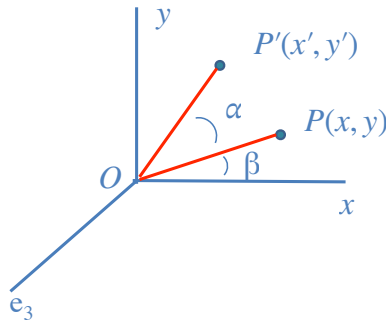


Figure 4.9: A rotation of the point P through an angle α in the $x - y$ plane.

The operation $C_z(\alpha)$ carries the point $P(x, y)$ into the point $P'(x', y')$ where

$$x' = OP \cos(\alpha + \beta), \quad y' = OP \sin(\alpha + \beta),$$

β being the angle OP makes with the x axis, or

$$x' = x \cos \alpha - y \sin \alpha, \quad y' = x \sin \alpha + y \cos \alpha.$$

Writing these equations in matrix form,

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

one finds a two-dimensional representation of $C_z(\alpha)$, viz.

$$C_z(\alpha) = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}. \quad (4.60)$$

(b) Representation of a reflection in a plane σ passing through \mathbf{e}_3 and making an angle α with the x axis. The operation is shown in Fig. 4.10.

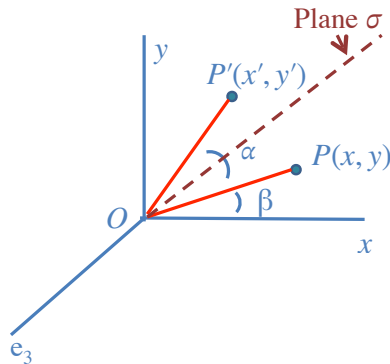


Figure 4.10: Reflection of the point P through a plane passing through \mathbf{e}_3 .

Defining the angle β as before one finds

$$x' = OP \cos(2\alpha - \beta), \quad y' = OP \sin(2\alpha - \beta),$$

or

$$x' = x \cos 2\alpha + y \sin 2\alpha, \quad y' = x \sin 2\alpha - y \cos 2\alpha,$$

so that

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

leading to the representation

$$\sigma(\alpha) = \begin{pmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{pmatrix} \quad (4.61)$$

for the reflection in the plane σ .

Instead of considering the transformation of the coordinates of a single point in two dimensions we may consider the transformation of all the displacement coordinates of an N -particle molecule. These coordinates generate a $3N \times 3N$ representation of the symmetry group of the molecule, the so-called *configuration space* representation. Because of its size, this representation is almost always reducible. Its reduction into the irreducible representations of the molecular symmetry group reveals important information about the nature of the normal modes of the molecule.

The preceding sentence is an important general statement that provides the link between abstract group theory and the small oscillations problem for molecules. While exploring its consequences, we carry along X_3 as an illustrative example. The symmetry group for X_3 is \mathcal{C}_{3v} (considered as a plane molecule in two dimensional space). Let us develop the

configuration space representation Γ for C_{3v} in the configuration space of X_3 . Looking back at Fig. 4.3 where we assigned the positions 1, 2 and 3 to the three X atoms and described the displacements of the atom in each of these positions by (x_1, y_1) , (x_2, y_2) and (x_3, y_3) , we now consider an arbitrary displacement of X_3 as in Fig. 4.11(a) (the arrows denote the extent and direction of displacement suffered by each X atom). "Freeze" the molecule in this displaced position and apply the operators of the group C_{3v} . Fig. 4.11(b) shows what the rotation C_3 about a z axis point out of the page does. After rotation, the atom in position 1 is endowed with the displacement that atom 3 had, 2 with the displacement 1 had, and 3 with the displacement 2 had, before rotation. Notice that the *labelling* of the three vertices is not altered by the symmetry operation.

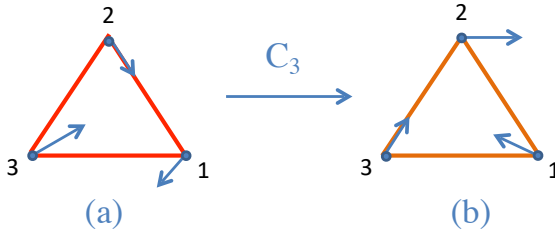


Figure 4.11: Action of C_3 on X_3 in a displaced configuration.

The effect of the rotation C_3 on the coordinates of each atom can be expressed analytically by means of the matrix (4.60) with $\alpha = 2\pi/3$. For instance if x'_1, y'_1 are the coordinates of the atom in position 1 after C_3 has acted, then

$$C_3 \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x'_1 \\ y'_1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x_3 \\ y_3 \end{pmatrix}.$$

Similar relations hold for the atoms in positions 2 and 3 before and after rotation so we can write

$$C_3 \begin{pmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \end{pmatrix} = \begin{pmatrix} x'_1 \\ y'_1 \\ x'_2 \\ y'_2 \\ x'_3 \\ y'_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & 0 & 0 & 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 & 0 & 0 & 0 \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \end{pmatrix}.$$

Looking back at (4.59), we see this means that the matrix

$$\Gamma(C_3) = \begin{pmatrix} 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & 0 & 0 & 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 & 0 & 0 & 0 \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 & 0 \end{pmatrix}$$

gives a 6×6 representation of the rotation C_3 . The two things that are striking about $\Gamma(C_3)$ are its "block" form, and the inordinate number of zeros that appear. In fact one can write

$$\Gamma(C_3) = \begin{pmatrix} \mathbf{0} & \mathbf{0} & C_3 \\ C_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & C_3 & \mathbf{0} \end{pmatrix}, \quad C_3 = \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{pmatrix} \quad (4.62)$$

in terms of 2×2 submatrices $\mathbf{0}$ and C_3 , where $\mathbf{0}$ is a 2×2 null matrix and C_3 the $\alpha = 2\pi/3$ version of (4.60). In a similar vein, we obtain the representations for the remaining operators of the group \mathcal{C}_{3v} :

$$\begin{aligned} \Gamma(C_3^{-1}) &= \begin{pmatrix} \mathbf{0} & C_3^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & C_3^{-1} \\ C_3^{-1} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad C_3^{-1} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{pmatrix} \\ \Gamma(\sigma_1) &= \begin{pmatrix} \sigma_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_1 \\ \mathbf{0} & \sigma_1 & \mathbf{0} \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{pmatrix} \\ \Gamma(\sigma_2) &= \begin{pmatrix} \mathbf{0} & \mathbf{0} & \sigma_2 \\ \mathbf{0} & \sigma_2 & \mathbf{0} \\ \sigma_2 & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \\ \Gamma(\sigma_3) &= \begin{pmatrix} \mathbf{0} & \sigma_3 & \mathbf{0} \\ \sigma_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_3 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{pmatrix} \\ \Gamma(E) &= \begin{pmatrix} I & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \quad (4.63)$$

The matrices for σ_1, σ_2 and σ_3 follow on setting $\alpha = -\pi/6, \pi/2$ and $\pi/6$ respectively in (4.61), and the matrix C_3^{-1} on setting $\alpha = -2\pi/3$ in (4.60). The representation $\Gamma(E)$ for the identity operation is quite obvious, but also formally follows on setting $\alpha = 0$ (or 2π) in (4.60).

The collection of matrices $\Gamma = \{\Gamma(E), \Gamma(C_3), \dots\}$ is clearly reducible since \mathcal{C}_{3v} only has one and two dimensional irreducible representations. Let us examine this reduction in some detail. First, we know that Γ must

be reducible into translational, rotational and vibrational representations Γ_t, Γ_r and Γ_{vib} generated by the translational, rotational and vibrational coordinates of the molecule. This is so because any one of these motions can be performed without exciting the others. We can thus write

$$\Gamma = \Gamma_t + \Gamma_r + \Gamma_{vib} \quad (4.64)$$

as a first step in the reduction of Γ . The representations on the right hand side of this equation are usually still further reducible. Of these, the reduction of Γ_{vib} is the important one since here we find information relating to the vibrational modes of the molecule.

To effect the reduction of Γ_{vib} we need to know the characters $\chi_{vib}(R)$ of this representation, see (4.46). But these are most easily found by subtraction, knowing the characters of Γ, Γ_t and Γ_r :

$$\chi_{vib}(R) = \chi(R) - \chi_t(R) - \chi_r(R) \quad (4.65)$$

in an obvious notation.

The characters $\chi(R)$ of the total representation Γ are read off directly from (4.62) and (4.63):

$$\chi(E) = 6, \quad \chi(C_3) = 0, \quad \chi(\sigma_1) = 0,$$

taking a typical element of each class. These numbers are appended to Table 4.3 next to the representation Γ .

The characters of Γ_t and Γ_r are also simple to calculate. Consider translations first. We take the center-of-mass coordinates of the molecule

$$\mathbf{X} = \frac{1}{\mu} \sum_p m_p \mathbf{x}_p \quad (4.66)$$

(\mathbf{x}_p is the displacement of the atom in the p^{th} position, m_p its mass and $\mu = \sum_p m_p$ the total molecular mass) and subject it to any symmetry operation R . Then, since R does not act on the symbol P ,

$$\begin{aligned} RX_i &= \frac{1}{\mu} \sum_p m_p (Rx_{pi}) = \frac{1}{\mu} \sum_p \sum_{j=1}^3 \Gamma_t(R)_{ij} x_{pj} \\ &= \sum_{j=1}^3 \Gamma_t(R)_{ij} X_j, \quad i = 1, 2, 3. \end{aligned}$$

Thus, \mathbf{X} transforms like the three dimensional vector that it is and the matrices $\Gamma_t(R)$, with matrix elements $\Gamma_t(R)_{ij}$, constitute a matrix representation of R . Its trace is

$$\chi_t(R) = \sum_{i=1}^3 \Gamma_t(R)_{ii}. \quad (4.67)$$

For X_3 there are six matrices $\Gamma_t(R)$ given by $C_3, C_3^{-1}, \sigma_1, \sigma_2, \sigma_3$ and \mathbf{I} in (4.62) through (4.63). The traces $\chi_t(R)$ of typical elements from each class are

$$\chi_t(E) = 2, \quad \chi_t(C_3) = -1, \quad \chi_r(\sigma_1) = 0$$

for X_3 . These numbers appear in Table 4.3 next to the representation Γ_t .

The representation Γ_r generated by the rotational coordinates is more tricky. Here, one must first decide how such a "coordinate" is defined. We know from our discussions of rigid body rotations that it is not generally possible to find coordinates such that their vanishing time derivatives implies that the body does not rotate. However, we can characterize a lack of rotation of the molecular framework by the vanishing of the angular momentum

$$\mathbf{L} = \sum_p m_p (\mathbf{r}_p \times \dot{\mathbf{x}}_p)$$

about its center-of-mass. Here \mathbf{r}_p measures the instantaneous position of the atom in position p from the center-of-mass and the \mathbf{x}_p have the same meaning as in (4.67). But $\mathbf{r}_p = \mathbf{r}_{op} + \mathbf{x}_p$ in terms of the equilibrium positions \mathbf{r}_{op} of each atom, so that one has

$$\mathbf{L} \simeq \sum_p m_p (\mathbf{r}_{op} \times \dot{\mathbf{x}}_p) = \frac{d}{dt} \sum_p m_p (\mathbf{r}_{op} \times \mathbf{x}_p)$$

to first order in the displacements \mathbf{x}_p . In this order, a vanishing \mathbf{L} means that the time derivative of

$$\mathbf{R} = \sum_p m_p (\mathbf{r}_{op} \times \mathbf{x}_p) \quad (4.68)$$

vanishes, thus placing \mathbf{R} on the same footing as \mathbf{X} , but for rotational instead of translational motion. We take \mathbf{R} as our "rotational" coordinate. It is an ordinary vector under rotation, but a *pseudovector* under reflections. In the present instance the single rotational coordinate R_z is found by direct calculation to be proportional to

$$(y_1 - y_3 - \sqrt{3}x_3),$$

which changes sign under σ_1 but is left unaffected by C_3 and E .

Hence

$$\chi_r(E) = 1, \quad \chi_r(C_3) = -1, \quad \chi_r(\sigma_1) = -1.$$

These numbers also appear in Table 4.3. The characters for the representation Γ_{vib} are now found by subtraction as

$$\chi_{vib} = 2, \quad \chi_{vib}(C_3) = 2, \quad \chi_{vib}(\sigma_1) = 1. \quad (4.69)$$

According to (4.46) the reduction of Γ_{vib} thus proceeds as follows:

$$\begin{aligned} \Gamma_{vib}(X_3) &= \frac{1}{6}(3+0+3)A_1 + \frac{1}{6}(3+0-3)A_2 + \frac{1}{6}(6+0+0)E \\ &= A_1 + E, \end{aligned} \quad (4.70)$$

if we use the relation (4.52) repeatedly.

Of what avail is this knowledge? We return to the dynamical description of a vibrating system to find out. The link to group theory is provided by the fact that the Lagrange function (see (4.7))

$$L = \sum_k \frac{1}{2} \dot{x}_k'^2 - \sum_k \frac{1}{2} W_{kl} x_k' x_l'$$

(notice the use of mass-scaled coordinates) is *invariant* under the symmetry operations of the molecular symmetry group. (This statement is of course equally true for the original Lagrangian given in the displacement coordinates but the theory is simpler to implement in the mass-scaled set as given above). To take our standard example of X_3 again, it is clear that the potential energy in the displacements pictured in Fig. 4.11 (a) and (b) is the same: all that has happened is that the physically equivalent X -atoms have changed places. Similarly, the kinetic energy term stays invariant under all symmetry operations of the group and therefore also L .

The same remarks are true if L is expressed in normal coordinates instead: L must be an invariant under all group operations. This requirement places certain restrictions on the way normal coordinates must transform under group operations and this in turn leads us directly to the information on the number and types of vibrational modes that a molecule has.

There are two cases:

(i) The system has no degenerate eigenfrequencies. This means that V has the form (the sum only includes vibrational normal coordinates, see (4.35))

$$V = \frac{1}{2} \Omega_1^2 \zeta_1^2 + \frac{1}{2} \Omega_2^2 \zeta_2^2 + \dots$$

when expressed in normal coordinates. The numbers $\Omega_1, \Omega_2, \dots$ are supposed to be all different. To maintain the same form for V under any symmetry operation R means that each normal coordinate must behave like

$$R\zeta_i = \pm \zeta_i \quad (4.71)$$

under a symmetry operation, i.e. they must be either symmetric (+1) or antisymmetric (-1). Thus, non-degenerate normal coordinates generate *one-dimensional* representations of the symmetry group. Such representations are obviously irreducible.

(ii) Some of the eigenfrequencies of the system may be degenerate. This means that some normal coordinates share a common frequency. The potential in this case looks like (again only vibrational coordinates are present)

$$V = \frac{1}{2} \Omega_1^2 \zeta_1^2 + \frac{1}{2} \Omega_2^2 \zeta_2^2 + \dots + \frac{1}{2} \omega_1^2 (\zeta_{11}^2 + \zeta_{12}^2 + \dots + \zeta_{1n_1}^2) \\ + \frac{1}{2} \omega_2^2 (\zeta_{21}^2 + \zeta_{22}^2 + \dots + \zeta_{2n_2}^2) + \dots + \frac{1}{2} \omega_\mu^2 (\zeta_{\mu 1}^2 + \zeta_{\mu 2}^2 + \dots + \zeta_{\mu n_\mu}^2) + \dots$$

where n_1 coordinates share the frequency ω_1 , n_2 coordinates the frequency ω_2 , etc. The non-degenerate coordinates still transform according to (4.71). But for a set of n_μ degenerate coordinates, $\zeta_{\mu 1}, \zeta_{\mu 2}, \dots, \zeta_{\mu n_\mu}$, all we can say is that

$$R\zeta_{\mu\alpha} = \sum_{\beta=1}^{n_\mu} D^{(\mu)}(R)_{\alpha\beta} \zeta_{\mu\beta}, \quad (4.72)$$

where $D^{(\mu)}(R)$ is an orthogonal matrix, in order to preserve the form of V . For (4.72) simply represents a *rotation* in the n_μ dimensional space of the normal coordinates $\{\zeta_{\mu\alpha}\}$, leaving the sum of squares

$$\zeta_{1\mu}^2 + \zeta_{2\mu}^2 + \dots + \zeta_{n_\mu\mu}^2$$

invariant. Clearly, the matrices with the $D^{(\mu)}(R)_{\alpha\beta}$ as matrix elements form an n_μ -dimensional representation of the molecular symmetry group. A little reflection shows that this must indeed be an *irreducible* representation for the potential energy has been reduced to a sum of squares.

Coordinates like the $\zeta_{\mu\alpha}$ which obey (4.72) are said to transform like (or belong to) the μ^{th} irreducible representation of the symmetry group. We have then that normal coordinates belonging to non-degenerate frequencies transform like the one dimensional representations of the group, those sharing a common eigenfrequency like multidimensional irreducible representations of the group, the dimension equalling the *multiplicity* f_μ of that degenerate eigenfrequency,

$$f_\mu = n_\mu.$$

We stress that these degeneracies are an inherent property of the molecule⁶¹ that cannot be destroyed without also destroying the symmetry of the molecule (by for example placing it in an external field with which it interacts).

Turning all this around now, we can say (i) that since the Lagrange function is brought into normal form by a linear orthogonal transformation (this is where the use of the mass-scaled coordinates is advantageous) $\mathbf{x}' = U\zeta$, the matrix U will likewise *serve to reduce the configuration space representation* Γ_{vib} (for the vibrational motion of the molecule) into *whatever irreducible representations it contains*, and (ii) that the dimensions of these irreducible representations tell us immediately how many non-degenerate and degenerate eigenfrequencies the molecule has, together with their multiplicities f_μ .

These remarks, and especially the reduction process, can be illustrated by looking back at (4.39) and (4.40) which show that one vibrational coordinate Q_{A_1} belongs to the frequency $\sqrt{3\kappa/m}$, and two vibrational coordinates Q_{E_1} and Q_{E_2} to the frequency $\sqrt{3\kappa/2m}$. This is in accord with the information contained in the reduction of Γ_{vib} for X_3 in (4.70).

⁶¹ As opposed to accidental degeneracies that can occur for special values of the coupling constants.

The transformation matrix U in (4.37) should reduce the configuration representation Γ of X_3 accordingly. This is indeed the case. Taking the representation matrix $\Gamma(C_3)$ for the rotation C_3 from (4.62) we find that

$$U^{-1}\Gamma(C_3)U = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & 0 & 0 & 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \quad (4.73)$$

fully reduces $\Gamma(C_3)$ into one dimensional and two dimensional irreducible representations of \mathcal{C}_{3v} . This reduced form agrees with the reduction

$$\Gamma = A_1 + A_2 + 2E \quad (4.74)$$

of Γ (see Prob. 4-8) and shows that the irreducible representation of C_3 are

$$D^{(A_1)}(C_3) = 1, \quad D^{(A_2)}(C_3) = 1, \quad \text{and} \quad D^{(E)}(C_3) = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}. \quad (4.75)$$

The irreducible representations for the remaining operators in \mathcal{C}_{3v} can be found in the same way. (Prob. 4-9).

4-11 Symmetry-Adapted Coordinates (SACs)

The appearance of two separate species of coordinates in the Lagrangian (4.29) for the X_3 molecule that are not coupled to each other suggests that coordinates belonging to different species of representation of the symmetry group never couple. This is indeed so. A term

$$(Q_{A_1}Q_{E1})$$

in the Lagrange function (4.39) would be inadmissible since according to (4.72) with $D^{(E)}(C_3)$ given by the last entry in (4.75) it would change into

$$Q_{A_1}\left(-\frac{1}{2}Q_{E1} - \frac{1}{2}\sqrt{3}Q_{E2}\right)$$

under action of the rotation C_3 , thereby changing the value of L . In fact, this simple remark leads to a very general discussion of the structure of the potential energy matrix U .

We introduce some terminology first. Coordinates like the Q_{A_1}, Q_{E1} , etc., that transform like the irreducible representations of the molecular symmetry group we choose to call *symmetry-adapted coordinates* (or SACs for short). They are not necessarily but can be *normal* coordinates of the system because of the non-uniqueness of the representation

matrices themselves. We restrict our attention for the moment to SACs not involving any linear or angular momentum of the molecule, since coordinates corresponding to such motions cannot occur in the potential function anyway. The symbol Q , suitably attired, is used to denote these coordinates. The SACs belonging to the μ^{th} irreducible representation respond to a symmetry operation R as in the equation

$$RQ_{\mu,\alpha} = \sum_{\beta=1}^{n_{\mu}} D^{(\mu)}(R)_{\alpha\beta} Q_{\mu,\beta}, \quad (4.76)$$

where n_{μ} is the dimension of the representation matrices $D^{(\mu)}(R)$. These coordinates themselves are said to define an *invariant subspace* of dimension n_{μ} . But we saw in the previous section that a given irreducible representation can occur more than once in Γ_v . This means there can be more than one *set* of coordinates $Q_{\mu,\alpha}$ that transform as in (4.76) and hence more than one invariant subspace of dimension n_{μ} . Two such subspaces are called *equivalent*. We need a notation to express this fact. Attach a subscript i to the representation index μ on $Q_{\mu,\alpha}$ thusly: $Q_{\mu_i,\alpha}$. If the representation μ is repeated a_{μ} times then

$$i = 1, 2, \dots, a_{\mu}$$

identifies the different sets of SAC's transforming according to the representation μ . Thus, different values of μ mean distinct subspaces, different values of i on the μ_i equivalent subspaces in this notation.

Now, to express the potential energy in terms of the set of SAC's

$$Q_{\mu_i,1}, Q_{\mu_i,2}, \dots, Q_{\mu_i,n_{\mu}}$$

where $i = 1, 2, \dots, a_{\mu}$ runs through all equivalent subspaces μ , and μ runs through all the distinct invariant subspaces admitted by the problem. We have

$$V = \sum_{\mu,\nu} \sum_{i,j} \sum_{\alpha,\beta} \frac{1}{2} W_{\mu_i,\alpha;\nu_j,\beta} Q_{\mu_i,\alpha} Q_{\nu_j,\beta}, \quad (4.77)$$

where the coefficients $W_{\mu_i,\alpha;\nu_j,\beta}$, like the W_{kl} of (4.5), form a symmetric matrix, but with rows and columns double-indexed by $\{\mu_i, \alpha\}$. A symmetry operation of V leads to

$$\begin{aligned} RV &= \frac{1}{2} W_{\mu_i,\alpha;\nu_j,\beta} D^{(\mu)}(R)_{\alpha'\alpha} D^{(\nu)}(R)_{\beta'\beta} Q_{\mu_i,\alpha'} Q_{\nu_j,\beta'} \\ &= \sum \frac{1}{2} W_{\mu_i,\alpha;\nu_j,\beta} Q_{\mu_i,\alpha} Q_{\nu_j,\beta}, \end{aligned}$$

since V is supposed to be invariant under all group operations. Therefore it follows that the relation

$$\sum_{\alpha'\beta'} D^{(\mu)}(R)_{\alpha'\alpha} D^{(\nu)}(R)_{\beta'\beta} W_{\mu_i,\alpha';\nu_j,\beta'} = W_{\mu_i,\alpha;\nu_j,\beta}$$

must be satisfied by the coefficients $U_{\mu_i\alpha;v_j\beta}$. This result holds for any symmetry operation R ; in particular it holds for the sum of all such operations. Summing both sides of this equation in R introduces the orthogonality properties of the $D^{(\mu)}(R)$ via the great orthogonality theorem, (4.43), (we only deal in *real* coordinates, therefore the matrices $D^{(\mu)}(R)$ are real and the complex conjugation $(*)$ in (4.43) is superfluous) and we find

$$g\delta_{\mu\nu}\delta_{\alpha\beta}\frac{1}{n_\mu}\sum_{\alpha=1}^{n_\mu}W_{\mu_i\alpha;v_j\alpha}=gW_{\mu_i\alpha;v_j\beta}.$$

This remarkable result says, first of all, that $W_{\mu_i\alpha;v_j\beta}$ is zero unless μ_i and v_j are equivalent representations ($\mu = \nu$), and $\alpha = \beta$. Furthermore, if these two conditions are satisfied then

$$\frac{1}{n_\mu}\sum_{\alpha=1}^{n_\mu}W_{\mu_i\alpha;\mu_j\alpha}=W_{\mu_i\alpha;\mu_j\alpha},$$

which can only be satisfied if both sides are *independent* of α . Thus,

$$W_{\mu_i\alpha;\mu_j\alpha}=W_{\mu_i\mu_j}\delta_{\mu\nu}\delta_{\alpha\beta}, \quad (4.78)$$

that is, the matrix elements $U_{\mu_i\alpha;v_j\beta}$ (a) only couple SACs lying in equivalent subspaces and (b) the coefficients of the cross products of all SACs in such a pair of equivalent subspaces are the same!

The final appearance of the potential in SACs is thus

$$V=\sum_{\mu}\sum_{i,j}\frac{1}{2}W_{\mu_i\mu_j}Q_{\mu_i\alpha}Q_{\mu_j\alpha}. \quad (4.79)$$

A similar reduction occurs for the kinetic energy. Furthermore, since the transformation from the configuration space coordinates to SACs is an orthogonal one if mass-scaled displacement coordinates are employed, the kinetic energy must appear as

$$T=\frac{1}{2}\sum_{\mu,l,\alpha}\dot{Q}_{\mu_l\alpha}^2+\dots \quad (4.80)$$

The ellipsis refers to SACs that describe the translation and rotation of the molecule that do appear in the kinetic energy but not in the potential energy (see (4.35)).

Putting (4.79) and (4.80) together we see that the determinantal equation (4.32) for the vibrational eigenfrequencies *factors* into subdeterminants of dimension equal to the number of equivalent irreducible representations that are present in Γ_{vib} . Furthermore there are as many degenerate eigenfrequencies as the dimension of any one of these equivalent representations, so that a given submatrix must be repeated in W as often as the dimension of the irreducible representation its SACs belong to.

The generality of these statements obscure their simplicity. An example will suffice as an illustration.

Consider the ammonia molecule NH_3 , the three H atoms forming an equilateral base triangle for a pyramid with the N atom at its apex. The symmetry group of NH_3 is clearly C_{3v} also, but the configuration representation Γ is now twelve dimensional (three degrees of freedom per atom). The characters of the total representation Γ as well as the characters of the translational and rotational representations are shown in Table 4.4 which repeats the character table of C_{3v} again.

	E	$2C_3$	$3\sigma_v$
A_1	1	1	1
A_2	1	1	-1
E	2	-1	0
Γ	12	0	2
Γ_t	3	0	1
Γ_r	3	0	-1
Γ_{vib}	6	0	2

Table 4.4: Character table for C_{3v} and characters for the NH_3 molecule.

The reduction of Γ_{vib} is easily found to be

$$\Gamma_{vib} = \frac{1}{6}(6+6)A_1 + \frac{1}{6}(6-6)A_2 + \frac{1}{6}(12)E = 2A_1 + 2E.$$

Therefore, NH_3 possesses two SAC's transforming like A_1 and two pairs of SAC's transforming like E . The corresponding structure of the W -matrix of (4.78) is shown in Fig. 4.12, the second and third 2×2 submatrices being identical.

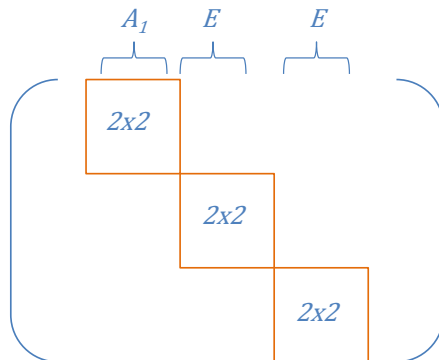


Figure 4.12: Structure of the W matrix for NH_3 .

Obviously, SACs are appropriate coordinates to employ in complicated vibrational problems. But how does one find them? A *systematic* application of group theoretic techniques provides a powerful method for doing so. We discuss the elements of this method next.

4-12 Projection Operators

One method of obtaining coordinates that belong to a prescribed irreducible representation of the molecular symmetry group is to try to construct them by inspection. But such an approach is obviously of limited value, depending as it does on the ingenuity of the user. We now show how a *systematic* procedure for finding SAC's can be developed, using some further results of group theory.

Consider again (4.76). Multiply both sides of this equation by $D^{(\nu)}(R)_{\gamma\delta}^*$ and then sum on all group operators R :

$$\sum_R D^{(\nu)}(R)_{\gamma\delta}^* R Q_{\mu,\alpha} = \frac{g}{n_\nu} \delta_{\mu\nu} \delta_{\gamma\alpha} Q_{\nu,\delta},$$

after invoking the great orthogonality theorem once more. Now set $\delta = \gamma$ on both sides of this expression and sum on $\gamma = 1, 2, \dots, n_\nu$. Then

$$\sum_R \sum_\gamma D^{(\nu)}(R)_{\gamma\gamma}^* R Q_{\mu,\alpha} = \sum_R \chi^{(\nu)}(R)^* R Q_{\mu,\alpha} = \frac{g}{n_\nu} \delta_{\mu\nu} Q_{\mu,\alpha},$$

or, writing μ for ν and ν for μ ,

$$\begin{aligned} P^{(\mu)} Q_{\nu,\alpha} &= 0, & \mu \neq \nu \\ &= Q_{\mu,\alpha}, & \mu = \nu, \end{aligned}$$

where the combination of operations

$$P^{(\mu)} = \frac{n_\mu}{g} \sum_R \chi^{(\mu)}(R)^* R \quad (4.81)$$

defines $P^{(\mu)}$. Thus, $P^{(\mu)}$ has the value one when operating on a coordinate $Q_{\mu,\alpha}$ that belongs to the irreducible representation μ and zero otherwise, i.e. $P^{(\mu)}$ is a *projection operator* onto the coordinates $Q_{\mu,\alpha}$.

We make use of this property as follows. The displacements x_k of each atom must be linear combinations of the symmetry-adapted coordinates $Q_{\mu_i,\alpha'}$. Moreover, if mass-scaled displacements $\sqrt{m_k} x_k = \tilde{x}'_k$ are used, it suffices to consider orthogonal transformations only:

$$\mathbf{x}' = U\mathbf{Q}, \quad \text{or} \quad \mathbf{Q} = U^T \mathbf{x}', \quad (4.82)$$

since $UU^T = I$. Operating on the k^{th} component of \mathbf{x}' with the projection operator $P^{(\mu)}$ gives

$$P^{(\mu)}(x'_k) = \sum_{i,\alpha} U_{k;\mu_i,\alpha} Q_{\mu_i,\alpha}, \quad (4.83)$$

containing (no summation on μ); the operator $P^{(\mu)}$ selects out of all the SAC's \mathbf{Q} in (4.82) those sets which belong to the μ^{th} irreducible representation only. On the other hand the left hand side of (4.83) can be calculated explicitly. It equals

$$P^{(\mu)}(x'_k) = \frac{n_\mu}{g} \sum_R \chi^{(\mu)*}(R) R x'_k = \frac{n_\mu}{g} \sum_R \sum_l \chi^{(\mu)*}(R) \Gamma(R)_{kl} x'_l, \quad (4.84)$$

where the $\Gamma(R)$ are the matrix representation of the group generated by the *complete* configuration space of the molecule. Consequently the "coordinates" $P^{(\mu)}(x'_k)$ that are *linear combinations* of the x_k may be used to construct a suitable set of Q 's. However, since each projection operator $P^{(\mu)}$ generates $3N$ linear combinations of coordinates that belong to μ , that are candidates for Q 's, and there are certainly less *independent* combinations than this per representation, there is considerable redundancy in the procedure. The problem (and its solution) is best illustrated by way of an example.

The projection operators for the irreducible representations A_1, A_2 and E_1 of \mathcal{C}_{3v} are

$$\begin{aligned} P^{(A_1)} &= \frac{1}{6}\{E + C_3 + C_3^{-1} + \sigma_1 + \sigma_2 + \sigma_3\} \\ P^{(A_2)} &= \frac{1}{6}\{E + C_3 + C_3^{-1} - \sigma_1 - \sigma_2 - \sigma_3\} \\ P^{(A_3)} &= \frac{1}{6}\{E - C_3 - C_3^{-1}\}, \end{aligned} \quad (4.85)$$

using (4.81) and the character table for \mathcal{C}_{3v} . Applying these operations in turn to the coordinates $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ of the X_3 molecule shown in Fig. 4.4 we obtain linear combinations of these coordinates that transform like A_1, A_2 and E . We choose the following combinations (going over into mass-scaled coordinates):

$$\begin{aligned} P^{(A_1)}(x'_1) &= \frac{\sqrt{3}}{12}\{\sqrt{3}(x'_1 - x'_3) - (y'_1 - 2y'_2 + y'_3)\} \sim Q_{A_1} \quad (i) \\ P^{(A_2)}(x'_1) &= \frac{1}{12}\{\sqrt{3}(y'_1 - y'_3) + (x'_1 - 2x'_2 + x'_3)\} \sim Q_{A_2} \quad (ii) \\ P^{(E)}(x'_1 + x'_2 + x'_3) &= x'_1 + x'_2 + x'_3 = Q'_{E1} \quad (iii) \\ P^{(E)}(y'_1 + y'_2 + y'_3) &= y'_1 + y'_2 + y'_3 = Q'_{E2} \quad (iv) \\ P^{(E)}(x'_1) &= \frac{1}{6}\{4x'_1 + x'_2 + x'_3 - \sqrt{3}(y'_2 - y'_3)\} \sim Q_{E1} \quad (v) \\ P^{(E)}(y'_1) &= \frac{1}{6}\{4y'_1 + y'_2 + y'_3 + \sqrt{3}(x'_2 - x'_3)\} \sim Q_{E2}. \quad (vi) \end{aligned} \quad (4.86)$$

From (4.86) (i) we can construct the first row of U in the inverse transformation

$$Q_{\mu_i, \alpha} = \sum_k U_{\mu_i, \alpha; k}^T x'_k \quad (4.87)$$

by writing

$$Q_{A_1} = \frac{1}{2\sqrt{3}}(\sqrt{3} \ -1 \ 0 \ 2 \ -\sqrt{3} \ -1)\{x'_1, y'_2, \dots\} \quad (4.88)$$

(the factor $1/2\sqrt{3}$ normalizes this vector to unity). The second row of U is obtained from (4.86) (ii) as

$$Q_{A_2} = \frac{1}{2\sqrt{3}}(1 \ \sqrt{3} \ -2 \ 0 \ 1 \ -\sqrt{3})\{x'_1, y'_2, \dots\}. \quad (4.89)$$

The row vectors appearing in Q_{A_1} and Q_{A_2} are automatically orthogonal to each other and to the row vectors Q_{E1} and Q_{E2} , etc. But the coordinates transforming like E have row vectors that are not necessarily orthogonal. Starting with

$$Q'_{E1} = \frac{1}{2\sqrt{3}}(2 \ 0 \ 2 \ 0 \ 2 \ 0)\{x'_1, y'_2, \dots\} \quad (4.90)$$

and

$$Q'_{E2} = \frac{1}{2\sqrt{3}}(0 \ 2 \ 0 \ 2 \ 0 \ 2)\{x'_1, y'_2, \dots\} \quad (4.91)$$

that *are* orthogonal to each other, we orthogonalize Q_{E1} and Q_{E2} (they are orthogonal to each other) to Q'_{E1} and Q'_{E2} . The result is

$$Q_{E1} = \frac{1}{2\sqrt{3}}(2 \ 0 \ -1 \ -\sqrt{3} \ -1 \ \sqrt{3})\{x'_1, y'_2, \dots\}$$

$$Q_{E2} = \frac{1}{2\sqrt{3}}(0 \ -2 \ -\sqrt{3} \ 1 \ \sqrt{3} \ 1)\{x'_1, y'_2, \dots\}$$

so that

$$\begin{pmatrix} Q_{A_1} \\ Q_{A_2} \\ Q_{E1} \\ Q_{E2} \\ Q'_{E1} \\ Q'_{E2} \end{pmatrix} = \frac{1}{2\sqrt{3}} \begin{pmatrix} \sqrt{3} & -1 & 0 & 2 & -\sqrt{3} & -1 \\ 1 & \sqrt{3} & -2 & 0 & 1 & -\sqrt{3} \\ 2 & 0 & -1 & -\sqrt{3} & -1 & \sqrt{3} \\ 0 & -2 & -\sqrt{3} & 1 & \sqrt{3} & 1 \\ 2 & 0 & 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} x'_1 \\ y'_1 \\ x'_2 \\ y'_2 \\ x'_3 \\ y'_3 \end{pmatrix} \quad (4.92)$$

and

$$\begin{pmatrix} x'_1 \\ y'_1 \\ x'_2 \\ y'_2 \\ x'_3 \\ y'_3 \end{pmatrix} = \frac{1}{2\sqrt{3}} \begin{pmatrix} \sqrt{3} & 1 & 2 & 0 & 2 & 0 \\ -1 & \sqrt{3} & 0 & -2 & 0 & 2 \\ 0 & -2 & -1 & -\sqrt{3} & 2 & 0 \\ 2 & 0 & -\sqrt{3} & 1 & 0 & 2 \\ -\sqrt{3} & 1 & -1 & \sqrt{3} & 2 & 0 \\ -1 & -\sqrt{3} & \sqrt{3} & 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} Q_{A_1} \\ Q_{A_2} \\ Q_{E1} \\ Q_{E2} \\ Q'_{E1} \\ Q'_{E2} \end{pmatrix} \quad (4.93)$$

give the relations between the configuration and symmetry-adapted coordinates, the latter relation serving to identify the origin of the matrix U in (4.38). We have succeeded in constructing the transformation matrix for the X_3 molecular vibration problem on the basis of symmetry arguments alone.

In terms of the SAC's Q_{A_1} , Q_{A_2} etc., the kinetic energy of X_3 assumes the form advertised in (4.80). In this case the potential energy also becomes diagonal, see (4.39), but this is an exception rather than a rule due to the high degree of symmetry X_3 exhibits. Usually some diagonalization among SAC's that belong to the same irreducible representation of the symmetry group of the molecule is necessary (Prob. 4-11).

4-13 Vibrations of Continuous Systems

We now inquire what happens to the vibrational modes of a system when the number of particles in it becomes infinite, that is, when it is no longer sensible (or desirable) to describe the motion of individual particles in the system. We restrict for the moment our considerations to linear systems in one dimension for simplicity. As a preamble to taking the limit $N \rightarrow \infty$, consider the transverse vibrations of a linear array of N identical, equally spaced particles that are joined together by identical springs. Particles 1 and N are attached to fixed points A and B by similar springs, Fig. 4.13. We use this system as a model for the vibrations of a continuous string tied down at A and B . We will clearly have to be careful about how the limit $N \rightarrow \infty$ obtains since this will determine the mass density distribution along the length of the string.

Looking at Fig. 4.13, we label the transverse displacement of particle k by y_k (we want to reserve x for measuring the distance *along* the string) and call their spacing a .

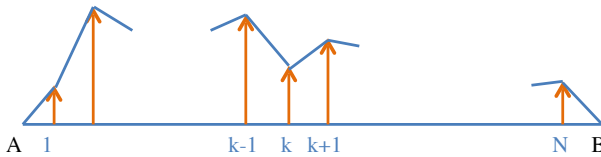


Figure 4.13: Linear chain of vibrating particles.

The points A and B are advantageously included in this scheme by calling their displacements x_0 and x_{N+1} and setting $y_0 = y_{N+1} = 0$. With this convention, the potential energy in an arbitrary displacement of the particle array is easily found. It is first of all clear that the potential energy associated with the k^{th} spring connecting particles $k - 1$ and k is

$$V_{k,k-1} = F\delta l_k \quad (4.94)$$

if F is the tension in the spring and δl_k its extension. But from Fig. 4.14

$$\delta l_k = a(\sec \theta - 1) \quad (4.95)$$

or

$$\delta l_k \simeq a\left(\frac{1}{2}\theta^2\right) \simeq \frac{a}{2}\left(\frac{y_k - y_{k-1}}{a}\right)^2 \quad (4.96)$$

if we make the small angle approximation for $\sec \theta$.

The potential energy of the entire system is thus

$$V = \sum_k V_k = \sum_{k=1}^{N+1} \frac{1}{2} Fa \left(\frac{y_k - y_{k-1}}{a}\right)^2, \quad (4.97)$$

so that the Lagrange function is given by

$$L = \sum_k \frac{1}{2} y_k'^2 - \sum_k \frac{1}{2} \frac{Fa}{m} \left(\frac{y_k' - y_{k-1}'}{a}\right)^2 \quad (4.98)$$

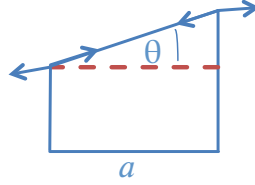


Figure 4.14: Geometry connecting two neighbouring particles.

after reverting to mass-scaled coordinates, $\sqrt{m_k}y_k = y'_k$. Hence, the eigenvalue problem (4.17) that is at the heart of the normal mode analysis reads (measure λ in units F/ma)

$$\begin{pmatrix} 2-\lambda & -1 & 0 & 0 & \cdots \\ -1 & 2-\lambda & -1 & 0 & \cdots \\ 0 & -1 & 2-\lambda & -1 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} u_1^\lambda \\ u_2^\lambda \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} = 0 \quad (4.99)$$

or

$$\begin{aligned} (2-\lambda)u_1^{(\lambda)} - u_2^{(\lambda)} &= 0 \\ -u_1^{(\lambda)} + (2-\lambda)u_2^{(\lambda)} - u_3^{(\lambda)} &= 0 \\ &\cdot \\ &\cdot \\ &\cdot \\ -u_{N-1}^{(\lambda)} + (2-\lambda)u_N^{(\lambda)} &= 0. \end{aligned} \quad (4.100)$$

This set of equations can be summarized by the single *recursion relation*

$$-u_{n-1}^{(\lambda)} + (2-\lambda)u_n^{(\lambda)} - u_{n+1}^{(\lambda)} = 0, \quad n = 1, \dots, N \quad (4.101)$$

if the *boundary conditions*

$$u_0^{(\lambda)} = 0, \quad u_{N+1}^{(\lambda)} = 0 \quad (4.102)$$

are obeyed for every λ . The relation (4.101) is obeyed by the set of functions

$$u_n = \sin n\varphi \quad (4.103)$$

where n is an integer, provided that

$$(2-\lambda) = 2 \cos \varphi, \quad \text{or} \quad \lambda = 2(1 - \cos \varphi). \quad (4.104)$$

The functions $\sin n\varphi$ automatically obey the first boundary condition in (4.102); the second one determines the allowed values of the angle φ and hence the eigenvalues through (4.104):

$$u_{N+1}^{(\lambda)} = \sin[(N+1)\varphi] = 0 \quad \text{if} \quad \varphi = \frac{m\pi}{N+1}, \quad (4.105)$$

where m is integral. Furthermore, $m \leq N$ since values of m equal to $N, N + 1, \dots$ just repeat the set of functions (4.103) for $m = 0, 1, \dots$. Finally, $m = 0$ is excluded since this gives the trivial solution $u_n = 0$ for all n . Hence the eigenvalues are

$$\lambda_m = 2(1 - \cos \frac{m\pi}{N + 1}), \quad m = 1, 2, \dots, N, \quad (4.106)$$

leading to the eigenfrequencies

$$\omega_m = \sqrt{\frac{2F}{\mu a^2}} \sqrt{1 - \cos(\frac{m\pi}{N + 1})}, \quad m = 1, 2, \dots, N \quad (4.107)$$

for a string carrying a mass $\mu = m/a$ per unit length. We notice from this expression that the ω_m bunch together as m increases until a maximum allowable frequency $\omega_{max} = \omega_N$ is reached at $m = N$. Frequencies higher than this cannot be propagated by the string.

The components of the m^{th} eigenvector are easily obtained from (4.103). They are

$$u_n^{(m)} = A \sin(\frac{m\pi n}{N + 1}), \quad n = 1, 2, \dots, N, \quad (4.108)$$

where A is a normalizing constant. Its value follows from the condition

$$\sum_{n=1}^N u_n^{(m)2} = A^2 \sum_{n=1}^N \sin^2(\frac{m\pi n}{N + 1}) = 1. \quad (4.109)$$

According to (4.108) the particles lie on a sine curve of wave length $2L/m$ in the m^{th} normal mode of vibration if the string has length L . The case for three particles, $N = 3$, is illustrated in Fig. 4.15.

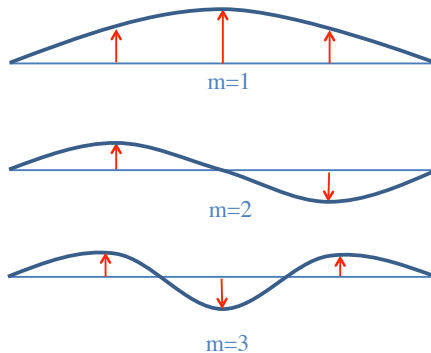


Figure 4.15: Transverse normal modes of three particles in a linear chain.

Consider now what happens as the number of particles increases without limit while the density, or mass per unit length, and the length L of the string are constant. This limiting procedure is characterized by the limits

$$a \rightarrow 0, \quad N \rightarrow \infty, \quad \text{but} \quad Na \rightarrow L, \quad \frac{m}{a} \rightarrow \mu, \quad (4.110)$$

where μ is the mass per unit length of a continuous string. The eigenvalues of the latter system are

$$\begin{aligned}\omega_m^2 &= \frac{F}{ma} \lambda_m = \lim_{a \rightarrow 0, N \rightarrow \infty} \frac{2F}{ma} \left[1 - \cos \frac{m\pi}{N+1}\right] \\ &= \lim_{a \rightarrow 0} \frac{2F}{\mu a^2} \left[1 - \cos \frac{m\pi a}{L}\right] \\ &= \lim_{a \rightarrow 0} \frac{2F}{\mu a^2} \left[1 - \left(1 - \frac{1}{2} \left(\frac{m\pi a}{L}\right)^2 + \dots\right)\right] \\ &= \frac{F}{\mu} \left(\frac{m\pi}{L}\right)^2,\end{aligned}\tag{4.111}$$

or

$$\omega_m = \sqrt{\frac{F}{\mu}} \left(\frac{m\pi}{L}\right), \quad m = 1, 2, \dots, \infty.\tag{4.112}$$

Notice that the frequency ω_m increases linearly with the mode number m *without limit* in contrast with (4.107) for the weighted string. The normal modes of this system are

$$y_k \sim \sin\left(\frac{mk\pi}{N+1}\right) \sim \sin\left(\frac{kam\pi}{L}\right).\tag{4.113}$$

But $ka = x$ is just the distance of the k^{th} particle from the point A .

Hence,

$$\lim_{a \rightarrow 0, N \rightarrow \infty} y_k \rightarrow y(x) = \sin\left(\frac{x\pi m}{L}\right)\tag{4.114}$$

is the normal mode displacement of the string at each point x along the length in the limit $N \rightarrow \infty$. This is a sine curve of wave length $2L/m$. These normal modes resemble the ones found for N finite in all respects except that there are an infinite number of them.

The result (4.114) is important from another entirely different point of view in that it shows that the displacement $y(x)$ of the string at x is the appropriate "coordinate" to use if we want to describe the dynamics of a continuous system directly without the intermediary procedure of passing through the discrete system first.

To see what changes in the Lagrange formalism are implied let us calculate (4.98) for the $N \rightarrow \infty$ case. Then,

$$\begin{aligned}L &\rightarrow \int \frac{1}{2} \mu \left(\frac{\partial y}{\partial t}\right)^2 dx - \int \frac{1}{2} \frac{F}{\mu} \left(\frac{\partial y}{\partial x}\right)^2 dx \\ &= \int \mathcal{L} dx,\end{aligned}\tag{4.115}$$

where

$$\mathcal{L} = \frac{1}{2} \mu \left(\frac{\partial y}{\partial t}\right)^2 - \frac{1}{2} F \left(\frac{\partial y}{\partial x}\right)^2\tag{4.116}$$

plays the obvious role of a linear density. It is called the *Lagrange density* of the system. Notice that \mathcal{L} depends on the derivatives $\partial y/\partial t$ and

$\partial y/\partial x$ of the "coordinate" y with respect to the parameters x and t . We therefore expect a Lagrange density of the form

$$\mathcal{L} = \mathcal{L}\left[y, \frac{\partial y}{\partial x}, \frac{\partial y}{\partial t}, t\right] \quad (4.117)$$

for a one dimensional system.

Such a system is called a *field*, described by the field amplitude $y(x, t)$. To find the equations of motion we return to (1.20) and write down the action S as

$$S = \int \int \mathcal{L} dx dt. \quad (4.118)$$

Variations in the value of S are now brought about by varying the field coordinates $y(x, t)$ at particular values of x and t . Calling δy the variation in y we get

$$\delta S = \int \int \left\{ \frac{\partial \mathcal{L}}{\partial y} \delta y + \frac{\partial \mathcal{L}}{\partial(\partial y/\partial x)} \delta\left(\frac{\partial y}{\partial x}\right) + \frac{\partial \mathcal{L}}{\partial(\partial y/\partial t)} \delta\left(\frac{\partial y}{\partial t}\right) \right\} dx dt, \quad (4.119)$$

since variations in y also induce variations in $\partial y/\partial x$ and $\partial y/\partial t$. In fact,

$$\delta\left(\frac{\partial y}{\partial x}\right) = \frac{\partial}{\partial x}(\delta y), \quad \text{and} \quad \delta\left(\frac{\partial y}{\partial t}\right) = \frac{\partial}{\partial t}(\delta y). \quad (4.120)$$

Therefore

$$\begin{aligned} \int \int \frac{\partial \mathcal{L}}{\partial(\partial y/\partial x)} \delta\left(\frac{\partial y}{\partial x}\right) &= - \int \int \frac{\partial}{\partial x} \frac{\partial \mathcal{L}}{\partial(\partial y/\partial x)} dx dt + \int_{t_1}^{t_2} dt \left[\frac{\partial \mathcal{L}}{\partial(\partial y/\partial x)} \delta y \right]_{x_1}^{x_2} \\ \int \int \frac{\partial \mathcal{L}}{\partial(\partial y/\partial t)} \delta\left(\frac{\partial y}{\partial t}\right) &= - \int \int \frac{\partial}{\partial t} \frac{\partial \mathcal{L}}{\partial(\partial y/\partial t)} dx dt + \int_{x_1}^{x_2} dx \left[\frac{\partial \mathcal{L}}{\partial(\partial y/\partial t)} \delta y \right]_{t_1}^{t_2}. \end{aligned} \quad (4.121)$$

The "surface terms" in both these expressions can be made to vanish by placing suitable boundary conditions on the variations δy as a function of x and t . In the former expression the time integral vanishes if either $y(x, t)$ vanishes as x_1 or x_2 approach $\pm\infty$ or if $y(x, t)$ obeys "periodic boundary conditions" at x_1 or x_2 ⁶²; the space integral in the latter expression vanishes if we impose the boundary condition

$$\delta y(x, t_1) = \delta y(x, t_2) = 0 \quad (4.122)$$

for all x , cf. (1.30). In view of these conditions the variation in the action now reads

$$\delta S = \int \int \left\{ \frac{\partial \mathcal{L}}{\partial y} - \frac{\partial}{\partial x} \frac{\partial \mathcal{L}}{\partial(\partial y/\partial x)} - \frac{\partial}{\partial t} \frac{\partial \mathcal{L}}{\partial(\partial y/\partial t)} \right\} \delta y dx dt, \quad (4.123)$$

which vanishes for an arbitrary variation δy if

$$\frac{\partial \mathcal{L}}{\partial y} - \frac{\partial}{\partial x} \frac{\partial \mathcal{L}}{\partial(\partial y/\partial x)} - \frac{\partial}{\partial t} \frac{\partial \mathcal{L}}{\partial(\partial y/\partial t)} = 0. \quad (4.124)$$

⁶² See, for example, L.I. Schiff, *Quantum Mechanics*, McGraw-Hill Book Company, Inc., New York, 1955, p43.

This is the Lagrange equation of motion for a one dimensional field $y(x, t)$. Notice that it is a *partial differential* equation, reflecting thereby an infinite number of degrees of freedom possessed by such a field.

If \mathcal{L} is given by (4.116), we get

$$\frac{\partial \mathcal{L}}{\partial y} = 0, \quad \frac{\partial \mathcal{L}}{\partial(\partial y/\partial x)} = -F\left(\frac{\partial y}{\partial x}\right), \quad \frac{\partial \mathcal{L}}{\partial(\partial y/\partial t)} = \mu\left(\frac{\partial y}{\partial t}\right), \quad (4.125)$$

so

$$-\frac{\partial}{\partial x}\left\{-F\frac{\partial y}{\partial x}\right\} - \frac{\partial}{\partial t}\left\{\mu\frac{\partial y}{\partial t}\right\} = 0, \quad (4.126)$$

or

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2}, \quad v = \sqrt{\frac{F}{\mu}} \quad (4.127)$$

is the equation of motion of a string under a constant tension F . Equation (4.126) is the one-dimensional "wave equation" for the propagation of waves along such a string with velocity $v = \sqrt{F/\mu}$.

4-14 The Oscillating String

As a first illustration of the application of (4.127) let us take a horizontal string under tension F that is tied down at both ends. The solution to this problem is actually contained in the expressions (4.112) and (4.114) but let us rederive these results directly. Equation (4.127) is a partial differential equation and as such admits solution by the separation of variables technique. Set

$$y(x, t) = u(x)f(t) \quad (4.128)$$

where u and f are functions of the variables shown in their arguments. Substituting back into (4.127) leads to

$$\frac{1}{u} \frac{d^2 u}{dx^2} = \frac{1}{v^2} \frac{1}{f} \frac{d^2 f}{dt^2}, \quad (4.129)$$

a relation that can only be satisfied if both sides equal a common constant (the separation constant) since the variables x and t have been segregated from each other. Calling the separation constant $-k^2$, one is lead to solve two ordinary differential equations

$$\begin{aligned} \frac{d^2 u}{dx^2} + k^2 u &= 0 \\ \frac{d^2 f}{dt^2} + (kv)^2 f &= 0. \end{aligned} \quad (4.130)$$

Assuming k^2 is positive, we find that $y = uf$ oscillates in both space and time according to

$$y(x, t) = e^{\pm ikx} e^{\pm i\omega t}, \quad \omega = kv. \quad (4.131)$$

The constant $\omega = kv$ is the *frequency* in time of the oscillation, k its *wave number*.

Any linear combination of the solutions (4.134) with specific choices of sign before k and ω are of course also solutions and we will make free use of this latitude. However, physical requirements, known as *boundary conditions*, on the system under discussion usually limit one to a particular combination. For example, in the case of the string tied down at both ends, we must require that

$$y(0, t) = y(L, t) = 0 \quad (4.132)$$

for all t , if L is the length of the string. This stipulation is said to give rise to a *standing wave* solution of the wave equation. The first boundary condition means that only the linear combination

$$\frac{1}{2i}(e^{ikx} - e^{-ikx}) = \sin kx \quad (4.133)$$

is permissible for the space part of $y(x, t)$. The second condition then fixes the value of the wave number k to be such that

$$\sin kL = 0, \quad \text{or} \quad k = k_m = \frac{m\pi}{L}, \quad m = 1, 2, \dots \quad (4.134)$$

Notice that $m = 0$ is excluded from this set as it gives a trivial solution to the problem (no displacement). Negative values of m are also excluded by the positive nature of the wave number k_m . Condition (4.134) thus gives us the eigenvalues of the vibrating string:

$$\omega_m = \sqrt{\frac{F}{\mu}}k_m = \sqrt{\frac{F}{\mu}}\left(\frac{m\pi}{L}\right), \quad m = 1, 2, \dots \quad (4.135)$$

in agreement with (4.112). The associated *eigenvectors* are

$$u_m(x) = \sin k_m x \quad (4.136)$$

apart from a normalizing constant. The function $u_m(x)$ is called an *eigenfunction* of the vibrating string. It is the obvious analog of $u_n^{(m)}$ in (4.108), with the index n replaced by the continuous index x , and the m moved down to agree with common usage. Multiplying $u_m(x)$ by the appropriate time factor, we get a normal mode of the string:

$$y_m(x, t) = u_m(x)e^{\pm i\omega_m t}. \quad (4.137)$$

We see then that there are an infinite number of such normal modes, so that the *most general* motion of the string tied down at both ends is given by the linear superposition

$$\begin{aligned} y(x, t) &= \sum_{m=1}^{\infty} C_m y_m(x, t) \\ &= \sum_{m=1}^{\infty} |C_m| \sin k_m x \cos(\omega_m t + \delta_m), \end{aligned} \quad (4.138)$$

containing an infinite number of adjustable complex constants $C_m = |C_m| \exp i\delta_m$. Clearly, the values of these constants are related to the initial conditions on the motion of the string. As usual, we are at liberty to specify its initial shape $y(x, 0)$ (the displacement of its constituent particles) and velocity $(\partial y / \partial t)_{t=0}$ (the velocity of its constituent particles) at $t = 0$. These functions are then expressed, according to (4.138) and its time derivative, by the *Fourier sine series*

$$y(x, 0) = \sum_{m=1} |C_m| \cos \delta_m \cdot \sin\left(\frac{m\pi x}{L}\right) \quad (4.139)$$

and

$$\left(\frac{\partial y}{\partial t}\right)_{t=0} = - \sum_m \omega_m |C_m| \sin \delta_m \cdot \sin\left(\frac{m\pi x}{L}\right) \quad (4.140)$$

in the interval $(0, L)$. The coefficients $|C_m| \cos \delta_m$ and $|C_m| \sin \delta_m$ are found by the usual inversion formulas

$$\begin{aligned} |C_m| \cos \delta_m &= \frac{2}{L} \int_0^L y(x, 0) \sin\left(\frac{m\pi x}{L}\right) dx \\ |C_m| \sin \delta_m &= - \frac{2}{\omega_m L} \int_0^L \left(\frac{\partial y}{\partial t}\right)_{t=0} \sin\left(\frac{m\pi x}{L}\right) dx \end{aligned} \quad (4.141)$$

that depend for their validity on the orthogonal nature

$$\frac{2}{L} \int_0^L \sin\left(\frac{m\pi x}{L}\right) \sin\left(\frac{n\pi x}{L}\right) dx = \delta_{mn} \quad (4.142)$$

of the set of eigenfunctions $\sin(m\pi x/L)$.

As a simple illustration, suppose the string is pulled aside a distance h at its center and released. The initial shape is thus

$$\begin{aligned} y(x, 0) &= x \frac{2h}{L}, \quad 0 < x < \frac{L}{2} \\ &= (L - x) \frac{2h}{L}, \quad \frac{L}{2} < x < L. \end{aligned}$$

We calculate $|C_m| \cos \delta_m$ from (4.141) as

$$|C_m| \cos \delta_m = \frac{2}{L} \cdot \frac{2h}{L} \cdot \left\{ \int_0^{L/2} x \sin\left(\frac{m\pi x}{L}\right) dx + \int_{L/2}^L (L - x) \sin\left(\frac{m\pi x}{L}\right) dx \right\}, \quad (4.143)$$

to find

$$\begin{aligned} |C_m| \cos \delta_m &= \frac{4h}{L^2} (1 + (-1)^{m+1}) \int_0^{L/2} x \sin \frac{m\pi x}{L} dx \\ &= \frac{8h}{\pi^2} \frac{(-1)^{(m-1)/2}}{m^2}, \quad m = 1, 3, 5, \dots \end{aligned} \quad (4.144)$$

and zero for even m . There is no initial velocity distribution along the string, so $|C_m| \sin \delta_m = 0$. The motion at subsequent times is therefore

given by the series

$$y(x, t) = \frac{8h}{\pi^2} \left\{ \sin \frac{\pi x}{L} \cos \omega_1 t - \frac{1}{3^2} \sin \frac{3\pi x}{L} \cos 3\omega_1 t + \frac{1}{5^2} \sin \frac{5\pi x}{L} \cos 5\omega_1 t - \dots \right\}, \quad (4.145)$$

containing no even modes.

4-15 Forced motion of a String

Up to now we have only considered the free oscillations of a string - we give the system some initial displacement and / or velocity and watch what it does. The next problem is to have some force (varying in space and time) act on the string. In what follows we shall always suppose that such forces are periodic in time, i.e. carry a time factor $\exp(i\omega t)$. The most important physical distinction between free and forced oscillations is then that the system oscillates with the frequency of the applied force rather than with any of its eigenfrequencies. A mathematical difference is that forced motion is described by an *inhomogeneous* differential equation. For suppose the string of Sec. 4-14 is subjected to forces distributed along its length according to $f^{(e)}(x) \exp(i\omega t)$ per unit length that act transversely. Their effect is included in (4.126) by simply adding this force to the left hand side of that equation. Hence,

$$\frac{\partial}{\partial x} \left\{ F \frac{\partial y}{\partial x} \right\} + f^{(e)}(x) e^{i\omega t} = \mu \frac{\partial^2 y}{\partial t^2} \quad (4.146)$$

describes the forced motion of a string of mass density μ under tension F . Assuming that F is constant throughout the length of the string, we try a solution

$$y(x, t) = u(x) e^{-i\omega t}, \quad (4.147)$$

and find that $u(x)$ satisfies

$$\left(\frac{d^2}{dx^2} + k^2 \right) u(x) = -\frac{f^{(e)}(x)}{F}, \quad k = \omega \sqrt{\frac{\mu}{F}}. \quad (4.148)$$

Notice that the dimension of $f^{(e)}/F$ is that of an inverse length, which agrees with the dimensions on the left.

It turns out that this equation is easily solved if we know the solutions of the *special* differential equation

$$\left(\frac{d^2}{dx^2} + k^2 \right) G_k(x, x') = \delta(x - x'), \quad (4.149)$$

where the forcing term is a delta function⁶³ $\delta(x - x')$ and is located at one point x' along the string. Its solution $G_k(x, x')$ is called a *Green function*.

⁶³ The symbol $\delta(x - x')$ is a mathematical artifice called the *Dirac delta function*. It is defined in such a fashion that

$$\int_a^b f(x) \delta(x - x') dx = f(x'), \quad (4.150)$$

where $a < x' < b$, and provided that $f(x')$ is finite.

From (4.149) we see that a Green function is nothing but the amplitude of displacement of the string under the action of the singular driving force $f(x) \sim \delta(x - x')$. To find $G_k(x, x')$, we note that as long as $x \neq x'$ the equation is homogeneous with the usual variety of solutions

$$\sin kx, \quad \cos kx, \quad (4.151)$$

or any combination of two of these. On the other hand, the presence of the delta function divides the x -axis into two regions, $0 < x < x'$ and $x' < x < L$. This is so because the singularity on the right hand side introduces a *discontinuity of slope* into the function $G_k(x, x')$ at $x = x'$. We see this by integrating both sides of (4.149) over the interval $x' - \varepsilon$ to $x' + \varepsilon$ around x' . Then, taking the limit,

$$\lim_{\varepsilon \rightarrow 0} \int_{x'-\varepsilon}^{x'+\varepsilon} k^2 G_k(x, x') dx = 0, \quad \lim_{\varepsilon \rightarrow 0} \int_{x'-\varepsilon}^{x'+\varepsilon} \delta(x - x') dx = 1, \quad (4.152)$$

but

$$\lim_{\varepsilon \rightarrow 0} \int_{x'-\varepsilon}^{x'+\varepsilon} \frac{d^2 G_k(x, x')}{dx^2} dx = \left(\frac{dG_k}{dx} \right)_{x=x'+\varepsilon} - \left(\frac{dG_k}{dx} \right)_{x=x'-\varepsilon}, \quad (4.153)$$

so that

$$\left(\frac{dG_k}{dx} \right)_{x=x'+\varepsilon} - \left(\frac{dG_k}{dx} \right)_{x=x'-\varepsilon} = 1 \quad (4.154)$$

is the "kink" or discontinuity, inflicted by the singular "function" $\delta(x - x')$ on G_k at $x = x'$. We now find the function G_k by patching together solutions of the homogeneous part of (4.149) in the two regions of space mentioned above. For $0 < x < x'$ we take

$$G_k(x, x') = A \sin k(x' - L) \sin kx \quad \text{for } x < x', \quad (4.155)$$

since G_k must vanish at end $x = 0$. Notice that the factor $A \sin k(x' - L)$ is a *constant* as far as x is concerned. Its presence simplifies the joining problem at $x = x'$, for this is automatic if we now interchange x and x' and take

$$G_k(x, x') = A \sin k(x - L) \sin kx' \quad \text{for } x > x'. \quad (4.156)$$

This also satisfied (4.149) for $x \neq x'$ and vanishes at the other end of the string $x = L$, as it should. The constant A is now found by requiring that $G_k(x, x')$ have a unit discontinuity in derivative at $x = x'$. The result is $A = 1/(k \sin kL)$, so that

$$\begin{aligned} G_k(x, x') &= \frac{1}{k \sin kL} \sin kx \sin k(x' - L), \quad 0 < x < x' \\ &= \frac{1}{k \sin kL} \sin kx' \sin k(x - L) \quad x' < x < L. \end{aligned} \quad (4.157)$$

The steady state displacement of the string under the influence of a general driving force can then be obtained by superposition:

$$y(x, t) = u(x)e^{-i\omega t} = - \int_0^L dx' \frac{f^{(e)}(x')}{F} G_k(x, x') e^{-i\omega t}, \quad (4.158)$$

a form reminiscent of $\zeta^{(s)}$ in Sec. 4-5. Note in particular that, considered as a function of the wave number k , the function $G_k(x, x')$ has poles at the allowed wave numbers k_m for the eigenvibrations of the *free* system ($k = 0$ is *not* a pole):

$$\sin kL = 0, \quad \text{or} \quad k = k_m = \frac{m\pi}{L}. \quad (4.159)$$

There is another form for $G_k(x, x')$ that is much more illuminating in determining the response $y(x, t)$ to the external field $f^{(e)}(x) \exp(i\omega t)$. This is the bilinear expansion of G_k into eigenfunctions of the homogeneous problem represented by the left hand side of (4.149). We write these eigenfunctions as

$$u_m(x) = \sqrt{\frac{2}{L}} \sin k_m x, \quad k_m = \frac{m\pi}{L}, \quad m = 1, 2, \dots \quad (4.160)$$

in the present case so that they obey the normalization condition (4.146) automatically,

$$\int_0^L u_m(x) u_n(x) dx = 0 \quad (4.161)$$

and suppose that, as a function of x , $G_k(x, x')$ has the expansion

$$G_k(x, x') = \sum_{m=1}^{\infty} a_m(x') u_m(x). \quad (4.162)$$

The expansion coefficients $a_m(x')$ depend on the *parameter* x' as indicated. Inserting this expansion into the differential equation (4.149) for G_k there results

$$\sum_{m=1}^{\infty} (-k_m^2 + k^2) a_m(x') u_m(x) = \delta(x - x'), \quad (4.163)$$

which can be solved for $a_n(x')$ by multiplying by $u_n(x)$ and integrating over x . The result

$$a_n(x') = \frac{u_n(x')}{k^2 - k_n^2}, \quad n = 1, 2, \dots \quad (4.164)$$

provides us with the symmetrical expression

$$G_k(x, x') = \sum_{m=1}^{\infty} \frac{u_m(x) u_m(x')}{k^2 - k_m^2} \quad (4.165)$$

for G_k . This expression is important because it shows immediately that only force distributions having a non-zero overlap with mode m can excite this mode:

$$\text{RHS of (4.158)} \sim \int_0^L u_m(x') f^{(e)}(x') dx' \neq 0, \quad (4.166)$$

if the m^{th} mode is to be excited.

The pole structure of G_k mentioned in connection with (4.159) is now self-evident of course and (4.165) is a very convenient formula for obtaining an approximation to G_k if k lies near one of the eigenvalues k_m . Then,

$$\frac{u_m(x)u_m(x')}{k^2 - k_m^2} \quad (4.167)$$

gives the dominant contribution to G_k . As an aside remark note that this expression has as residue at $k = k_m$, the product of the normalized eigenfunctions

$$\frac{1}{2k_m} u_m(x)u_m(x') \quad (4.168)$$

of the homogeneous problem. Therefore, if we know $G_k(x, x')$ from some other calculation as in (4.157), we can determine these eigenfunctions as the residues of G_k at the poles $k = k_m$. For instance this residue is

$$\lim_{k \rightarrow k_m} (k - k_m)G_k(x, x') = \frac{\sin k_m x \sin k_m x'}{2k_m}, \quad (4.169)$$

or

$$u_m(x)u_m(x') = \sqrt{\frac{2}{L}} \sin k_m x \cdot \sqrt{\frac{2}{L}} \sin k_m x', \quad (4.170)$$

confirming the result (4.160).

A more detailed discussion of the application of Green functions to calculate the response of continuous systems would take us too far afield. The interested reader is referred to P.M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Mc-Graw-Hill Book Company, Inc., New York, 1951, Vol. I, for further details of this method.

Problems

4-1. Show that the eigenvalues of a real, symmetric matrix are all real.

4-2. Reappraise (4.20) through (4.23) in the event that the matrix W has two equal roots, say $\lambda_1 = \lambda_2$. Test out your considerations on the matrix

$$\begin{pmatrix} 7 & -2 & 1 \\ -2 & 10 & -2 \\ 1 & -2 & 7 \end{pmatrix} \quad (4.171)$$

by explicitly constructing the matrix U that diagonalizes it.

4-3. Show that each element of a group appears once and once only in any row or column of the group multiplication table. (Hint: assume the converse and show that a violation of the group postulates arises).

4-4. Show that no element of a group can appear in more than one class of the group.

- 4-5. Show that for an abelian group the number of classes equals the order of the group.
- 4-6. Prove that operations $E, C_3, C_3^{-1}, \sigma_1, \sigma_2, \sigma_3$ listed in (4.54) form a group by showing that all the group postulates of Sec. 4-6 are satisfied.
- 4-7. Verify (4.73).
- 4-8. Verify that (4.74) gives the correct reduction of the configuration space representation of the X_3 molecule.
- 4-9. Find the irreducible matrix representation for all the operators of the group C_{3v} following along the lines leading up to (4.75).
- 4-10. Verify Table 4.4 in the text as the correct character table for the NH_3 molecule.
- 4-11. Set up and analyze the eigenvibration problem of a water molecule (H_2O) using the methods of group theory. Show that the secular determinant can be broken up in one 2×2 determinant and one 1×1 determinant. Construct the symmetry adapted coordinates for H_2O and solve explicitly for the vibrational frequencies under the assumption that the potential energy depends quadratically on both the OH bond extensions and the HOH bond angle increment.
- The H_2O molecule is a prototype of *non-linear* XY_2 molecules. The symmetry group you will come up with is called C_{2v} . It is an abelian group of order 4.
- 4-12. Consider, as a working hypothesis, that when the hammer of a piano hits a piano string it creates a "velocity profile"

$$v(x) = \frac{\bar{v}L}{\sqrt{\pi a}} e^{-(\frac{x-d}{a})^2} \quad (4.172)$$

of extent a about the point of impact d from one end of the string. Here L is the length of the string and \bar{v} its average initial velocity. Assume further that the string is negligibly displaced in the process. Calculate the motion of the string in the limit that $a \ll 1$ and show that even or odd modes can be suppressed by choosing the ratio d/L appropriately. Take this opportunity to discuss the energy distribution of the excited modes of the vibrating string. Show that low frequency modes contain most of the energy initially supplied to the string in the present example.

- 4-13. Consider a horizontal string of length L and mass density μ_0 tied down at both ends. In addition, the string carries N point masses

m equally spaced along the string. Present an analysis of the eigen-vibrations of this system. (Hint: write the mass density as $\mu(x) = \mu_0 + m \sum_{k=1}^N \delta(x - x_k)$ in the wave equation for a continuous string and construct suitable solutions).

Chapter 5 Fluid Motion

5-1 Introduction

In this Chapter we consider the dynamics of a fluid. By this concept we mean an interacting many-particle system containing a large number of particles per unit volume so that even a macroscopically small volume element in space contains a large number of particles. We may then examine the motion of such a system by considering the motion of a representative "fluid particle", i.e. a volume element in space containing many particles but macroscopically "small". We characterize the motion of such a fluid particle by the velocity of the fluid $\mathbf{u}(\mathbf{r}, t)$ at the position \mathbf{r} of the fluid particle and the density of the fluid at this point. Strictly speaking the "position" referred to by \mathbf{r} is fuzzy up to the size of the volume element delineating our fluid particle. Likewise, the velocity \mathbf{u} at the point \mathbf{r} is to be thought of as an average velocity throughout this volume element. One should carefully bear in mind these physical idealizations of the system in order to appreciate what is to follow.

A complete treatment of fluid motion within the confines of a single chapter is not practical. This extensive field deserves a text book for itself. Our purpose here is rather a much more modest one: to introduce the reader to the basic ideas and equations, with some applications of the laws of motion governing *perfect* fluids. A perfect fluid is one devoid of viscosity and hence also of most physical interest. Nevertheless, a study of such systems is essential as a first step to understanding the incredibly complicated field of real fluids.

Some kinematical concepts are in order first. We consider a finite volume ΔV in a fluid. It contains an amount of mass

$$\int_{\Delta V} \rho dV \quad (5.1)$$

if $\rho = \rho(\mathbf{r}, t)$ designates the fluid density at \mathbf{r} at time t and $dV = dx dy dz$ the volume element at position \mathbf{r} . If the fluid is in motion a net amount of fluid

$$\oint \rho u_n da \quad (5.2)$$

will pass out of ΔV in unit time. In this expression u_n is the fluid velocity normal to the surface element da pointing out of the volume ΔV and

the integral goes over the entire bounding surface of ΔV . This result follows from the observation that ρu_n is the mass of fluid transported per unit time across unit area perpendicular to u_n . A net outflow means mass is being lost from ΔV . This can only occur at the expense of a decrease in mass in ΔV since mass is conserved in a fluid. The decrease in mass per unit time is

$$-\frac{\partial}{\partial t} \int_{\Delta V} \rho dv = - \int_{\Delta V} \frac{\partial \rho}{\partial t} dv, \quad (5.3)$$

since ΔV is fixed. Equating this to the mass outflow, we obtain the conservation law

$$\int \frac{\partial \rho}{\partial t} dv + \oint \rho u_n da = 0. \quad (5.4)$$

The second member of this equation can be transformed into a volume integral using Gauss' theorem in reverse,

$$\oint \rho u_n da = \int_{\Delta V} \text{div}(\rho \mathbf{u}) dv \quad (5.5)$$

so that

$$\int_{\Delta V} \left\{ \frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{u}) \right\} dv = 0. \quad (5.6)$$

But the volume ΔV is arbitrary. Therefore, if we assume it is small enough so that ρ and \mathbf{u} can be replaced by their values "at the center" of ΔV (but still large enough to contain many particles), this equation reads

$$\left\{ \frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{u}) \right\} \int_{\Delta V} dv = 0, \quad (5.7)$$

or, equivalently,

$$\frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{u}) = 0. \quad (5.8)$$

This is the *equation of continuity* expressing the conservation of mass in fluid flow in differential form.

5-2 Euler's Equations for Fluid Motion

So much for kinematical preliminaries. We now go over to the task of examining the dynamics of fluid flow. The action principle of Chapter 1 will again be our guide. We take up this principle in the form of (1.31),

$$\int_{t_1}^{t_2} (\delta T + \delta W) dt = 0, \quad (5.9)$$

and tailor it to suit our problem. We straightaway state that we are only going to study *incompressible* fluids (in practice ordinary liquids like water or alcohol) that are characterized by exhibiting no noticeable

change in their density in space or time when subjected to external stresses. This means

$$\rho = \rho_0 \quad \text{a constant,} \quad \frac{\partial \rho}{\partial t} = 0, \quad \text{grad } \rho = 0, \quad (5.10)$$

and so

$$\text{div } \mathbf{u} = 0, \quad (5.11)$$

from the equation of continuity. Equation (5.9) is a *condition* the flow velocity must obey at each point in a fluid if the fluid is incompressible.

We can now fill in (5.9). The kinetic energy density of fluid contained in a volume element dv is obviously

$$\frac{1}{2} \rho_0 u^2, \quad (5.12)$$

so that

$$T = \int \frac{1}{2} \rho_0 u^2 dv, \quad \delta T = \int \rho_0 (\mathbf{u} \cdot \delta \mathbf{u}) dv, \quad (5.13)$$

if $\delta \mathbf{u}$ is the variation induced in \mathbf{u} by the virtual displacement $\delta \mathbf{r}$ of the fluid particle at \mathbf{r} . The virtual work is simply calculated as

$$\delta W = \int (\mathbf{F} \cdot \delta \mathbf{r}) dv \quad (5.14)$$

at any internal point in the fluid if \mathbf{F} is the applied force (per unit volume) at this point. Thus (5.9) reads

$$\int_{t_1}^{t_2} \int \{ \rho_0 \mathbf{u} \cdot \delta \mathbf{u} + \mathbf{F} \cdot \delta \mathbf{r} \} dv dt = 0 \quad (5.15)$$

for an incompressible fluid subjected to a force distribution \mathbf{F} per unit volume throughout its volume. Equation (5.15) does not account for the *internal* stresses in a fluid that render it incompressible. Microscopically these are due to the strong interparticle forces that act at liquid densities. Macroscopically we replace these forces of constraint (for that is what they are) by a condition of constraint of the fluid motion. This condition is just (5.11). This equation must be obeyed by all virtual motions of our fluid. Thus, $\text{div } \mathbf{u} = 0$ implies also

$$\text{div}(\delta \mathbf{u}) = 0, \quad \text{or} \quad \frac{d}{dt} [\text{div}(\delta \mathbf{r})] = 0, \quad (5.16)$$

since

$$\delta \mathbf{u} = \frac{d}{dt} (\delta \mathbf{r}), \quad (5.17)$$

according to a by now familiar argument. We would like (5.16) to be obeyed for all times. This is certainly the case if we restrict ourselves to virtual displacements that are subject to the constraint

$$\text{div}(\delta \mathbf{r}) = 0. \quad (5.18)$$

A Lagrange multiplier λ is necessary to include this constraint in (5.15), which then becomes

$$\int_{t_1}^{t_2} \int \{ \rho_0 \mathbf{u} \cdot \delta \mathbf{u} + \mathbf{F} \cdot \delta \mathbf{r} + \lambda \operatorname{div}(\delta \mathbf{r}) \} dv dt = 0. \quad (5.19)$$

Two transformations are now necessary. First we make use of the relation (5.17) to rewrite the first term as

$$\int_{t_1}^{t_2} \int \rho_0 \mathbf{u} \cdot \delta \mathbf{u} dv dt = - \int_{t_1}^{t_2} \int \rho_0 \frac{d\mathbf{u}}{dt} \cdot \delta \mathbf{r} dv dt, \quad (5.20)$$

after imposing the usual boundary conditions $\delta \mathbf{r}(t_1) = 0$, $\delta \mathbf{r}(t_2) = 0$, at the endpoints of the time integration. The second transformation involves the last integral whose integrand may be written as

$$\lambda \operatorname{div}(\delta \mathbf{r}) = \operatorname{div}(\lambda \delta \mathbf{r}) - \operatorname{grad} \lambda \cdot \delta \mathbf{r}. \quad (5.21)$$

Carrying out a spatial integration and using Gauss' theorem once more, one obtains (δr_n is the component of $\delta \mathbf{r}$ normal to the surface element da)

$$\int \lambda \operatorname{div}(\delta \mathbf{r}) dv = \oint \lambda \delta r_n da - \int \operatorname{grad} \lambda \cdot \delta \mathbf{r} dv. \quad (5.22)$$

Reserving remarks on the effect of the surface integral in this expression for later, we temporarily ignore it and obtain

$$\int_{t_1}^{t_2} \int \{ -\rho_0 \frac{d\mathbf{u}}{dt} + \mathbf{F} - \operatorname{grad} \lambda \} \cdot \delta \mathbf{r} dv dt = 0 \quad (5.23)$$

for the final form of the action principle. As before, we consider the time interval $t_2 - t_1 = \Delta t$ to be vanishingly small and conclude that

$$\left[\int (-\rho_0 \frac{d\mathbf{u}}{dt} + \mathbf{F} - \operatorname{grad} \lambda) \cdot \delta \mathbf{r} dv \right] \cdot \int_{\Delta t} dt = 0 \quad (5.24)$$

where the variables in the round brackets (\cdot) now refer to a particular instant of time t . But $\delta \mathbf{r}$ can now be chosen arbitrarily. We assume it vanishes everywhere but in a small volume ΔV surrounding the point \mathbf{r} . Then,

$$\left[(-\rho_0 \frac{d\mathbf{u}}{dt} + \mathbf{F} - \operatorname{grad} \lambda) \cdot \delta \mathbf{r} \right] \cdot \int_{\Delta V} dv \int_{\Delta t} dt = 0, \quad (5.25)$$

showing that the relation

$$-\rho_0 \frac{d\mathbf{u}}{dt} + \mathbf{F} - \operatorname{grad} \lambda = 0 \quad (5.26)$$

holds at each point \mathbf{r} in the fluid at time t . The components of this equation are easily identified physically. The term $-\rho_0 d\mathbf{u}/dt$ is the negative of the rate of change of momentum (per unit volume) of a fluid moving with velocity \mathbf{u} , and \mathbf{F} is the force on this unit volume. But what

is $-\text{grad}\lambda$? This clearly must be a *force of constraint* that arises in order to accommodate the condition (5.11) of incompressible flow. In this context λ can be identified with the *pressure* p at a point in the fluid since

$$-\text{grad } p \quad (5.27)$$

represents a pressure gradient or force per unit volume a fluid particle at \mathbf{r} will experience. This interpretation of pressure gives it the physically understandable meaning of an internal reaction in the fluid in order to obey the constraint of incompressibility. We note that p is an unknown that has to be determined along with \mathbf{u} from the equations of motion.

Equation (5.26), with the interpretation (5.27) of λ constitute the equations of motion we seek. They were first obtained by Leonhard Euler in 1755. To give Euler's form of these equations, some further transformations are necessary. We first carefully note the meaning of the differential quotient

$$\frac{d\mathbf{u}}{dt} \quad (5.28)$$

in (5.26). Since $\mathbf{u} = \mathbf{u}(\mathbf{r}, t)$, changes in \mathbf{u} will occur through its explicit dependence on time at a fixed point in space and due to the fact that the position \mathbf{r} of the fluid element we are examining changes with time. For example,

$$\frac{du_x}{dt} = \frac{\partial u_x}{\partial t} + u_x \frac{\partial u_x}{\partial x} + u_y \frac{\partial u_x}{\partial y} + u_z \frac{\partial u_x}{\partial z}, \quad (5.29)$$

where $\mathbf{u} = \{u_x, u_y, u_z\}$. Similar equations are found for u_y and u_z . One may write the summarizing vector expression as

$$\frac{d\mathbf{u}}{dt} = \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u}, \quad (5.30)$$

where $\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$, but this is only true if the cartesian components of \mathbf{u} are used. Now, the vector operation can be applied in the sense of the following formula

$$(\mathbf{u} \cdot \nabla)\mathbf{u} = \text{grad}\left(\frac{1}{2}u^2\right) - \mathbf{u} \times \text{curl } \mathbf{u}. \quad (5.31)$$

Therefore, we can write Euler's equation (5.26) in the symbolic form

$$\rho_0 \left(\frac{\partial \mathbf{u}}{\partial t} - \mathbf{u} \times \text{curl } \mathbf{u} \right) + \text{grad}\left(\frac{1}{2}\rho_0 u^2 + p\right) = \mathbf{F} \quad (5.32)$$

that is valid in any system of coordinates for \mathbf{u} . This has to be supplemented by the incompressible flow condition $\text{div } \mathbf{u} = 0$ in all our applications of course, giving us two equations for finding the unknowns u and p .

We yet have to account for the role placed by the ignored surface integral

$$\oint \lambda \delta r_n da = \oint p \delta r_n da \quad (5.33)$$

in (5.22). Its role can be elucidated by the following considerations. First, suppose the fluid surface is in contact with a rigid bounding wall. Then $\delta r_n = 0$ at every point on such a rigid surface and the integral must vanish. This cannot be the case, however, if the fluid has a *free surface*. For then, there is certainly no reason to maintain $\delta r_n = 0$. However, a second contribution to the virtual work comes into play at a free surface: One has to invoke a microscopic picture of the fluid to realize that an additional force $-F'_n$ per unit area pointing along the inward normal arises at a free surface due to the unbalanced internal forces on particles "at" the surface coming from particles below them. Consequently, δW has to be supplemented by

$$- \int F'_n \delta r_n da \quad (5.34)$$

at a free surface. Inclusion of this term in the action principle, together with (5.37), leads to a contribution

$$\int_{t_1}^{t_2} \int (p - F'_n) \delta r_n da dt, \quad (5.35)$$

that vanishes provided

$$p = F'_n. \quad (5.36)$$

Thus, the action principle also provides us with a *surface boundary condition* on the pressure at a free surface.

Before discussing specific applications of (5.32), some general remarks are in order. The first thing that strikes one about Euler's equation is its *non-linear* character. This fact alone accounts for the immense complexity of fluid flow problems, for it negates entirely the principle of superposing simple solutions of a problem to build up a general solution. One's natural inclination is to try to restore this principle by *linearizing* (5.32), that is, by dropping terms quadratic in the velocity \mathbf{u} . This underscores the problem of slowly-moving fluids and is the subject of subsequent sections. However, there is a particular case of (5.32) for which a first integral may readily be found. This is the case of irrotational flow:

$$\text{curl } \mathbf{u} = 0, \quad (5.37)$$

which is also steady ($\frac{\partial \mathbf{u}}{\partial t} = 0$). Then that equation reduces to

$$\text{grad} \left(\frac{1}{2} \rho_0 u^2 + p \right) = \mathbf{F} = -\text{grad } U, \quad (5.38)$$

if we assume in addition that \mathbf{F} is derivable from a potential field U per unit volume. Then space integration results in

$$\frac{1}{2} \rho_0 u^2 + p + U = \text{constant}, \quad (5.39)$$

a relation first obtained by Bernoulli in 1728. (Note the date! It was obtained *before* the work of Euler). Bernoulli's equation is perhaps the most fundamental relation in elementary fluid mechanics. In this equation $\rho_0 u^2/2$ is the kinetic energy per unit volume of fluid, U the external potential, and the pressure p a kind of potential energy per unit volume arising from internal forces. One surprising consequence of (5.39) is the drop in pressure it dictates with increasing velocity of fluid flow, a well-known (but often not appreciated) fact to most students of elementary physics.

5-3 Potential Flow

We showed in the previous section that Bernoulli's equation was the result of integrating Euler's equation under the condition (5.37) corresponding to irrotational flow. But we know from vector analysis that it is always possible to construct an irrotational vector by taking the gradient of any scalar function φ . Thus, $\text{curl } \mathbf{u} = 0$ identically if

$$\mathbf{u} = -\text{grad } \varphi \quad (5.40)$$

(the minus sign is by convention only). Since the flow we consider is also incompressible we can superpose this condition on (5.39) to obtain

$$\nabla^2 \varphi = 0 \quad (5.41)$$

that is Laplace's equation for φ . The function φ is called the *velocity potential* by analogy with potential theory of conservative force fields and flow governed by (5.40) and (5.41) is called *potential flow*. It is important to notice that (5.41) does not yet contain any dynamics. To obtain this we need to substitute (5.40) into Bernoulli's equation,

$$\frac{1}{2}\rho_0(\text{grad } \varphi)^2 + p + U = \text{constant}, \quad (5.42)$$

an equation which now determines the pressure field p of a given velocity potential.

A more general version of (5.42) is obtained if we keep the irrotationality condition (5.37) but allow the flow to be non-steady ($\partial U/\partial t \neq 0$). Then we find that

$$\text{grad}\left(-\rho_0 \frac{\partial \varphi}{\partial t} + \frac{1}{2}\rho_0(\text{grad } \varphi)^2 + p + U\right) = 0, \quad (5.43)$$

after using (5.40) for the first term in (5.32) and interchanging the order of differentiation $\partial/\partial t$ with grad. It therefore follows that

$$-\rho_0 \frac{\partial \varphi}{\partial t} + \frac{1}{2}\rho_0(\text{grad } \varphi)^2 + p + U = \text{constant} \quad (5.44)$$

once more, but where the "constant" can now be a function of time.

5-4 Water Waves

Let us apply (5.44) to the phenomenon of surface waves on an open sea. The coordinate system O_{xyz} we use is situated on the surface of the sea as illustrated in Fig. 5.1. The first step is to find a suitable velocity potential φ . We assume that all quantities associated with surface waves are *periodic* in time, with period $2\pi/\omega$.

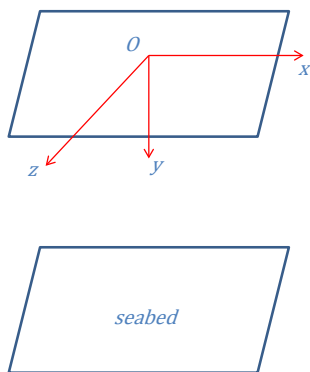


Figure 5.1: Coordinate system for discussing water waves on an open sea.

This means φ has the form

$$\varphi(x, y)e^{-i\omega t} \quad (5.45)$$

if we restrict the discussion to waves propagating in the x direction.

Thus, we must find the function $\varphi(x, y)$ that satisfies

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\varphi = 0. \quad (5.46)$$

Assuming separability, $\varphi = f(x)g(y)$, one finds that

$$\frac{1}{f} \frac{d^2 f}{dx^2} + \frac{1}{g} \frac{d^2 g}{dy^2} = 0, \quad (5.47)$$

or that

$$\frac{d^2 f}{dx^2} + k^2 f = 0, \quad \frac{d^2 g}{dy^2} - k^2 g = 0, \quad (5.48)$$

where k^2 is a separation constant. These two equations admit a variety of solutions. The ones we require are determined by *boundary conditions* at the sea bed and water surface. Suppose for simplicity that the sea is very deep. The φ should certainly vanish at the sea bed, $y \rightarrow \infty$, since the velocity u_y must vanish there. This dictates the choice

$$g(y) = e^{-ky} \quad (5.49)$$

for the function $g(y)$. The solutions for f are oscillatory in space as a function of x . We want here to study *travelling* waves on the surface of the sea, and so take

$$f(x) = e^{ikx}. \quad (5.50)$$

Our final solution for the velocity potential thus becomes

$$\varphi(x, y)e^{-i\omega t} = Ae^{-ky}e^{i(kx-\omega t)}, \quad (5.51)$$

where A is an arbitrary amplitude. We now have to ensure that the velocity of flow given by this potential is consistent with Bernoulli's equation at the surface of the water. By doing so, we introduce a dynamic element into the calculation. This compatibility problem is only tractable in the event that we consider small deviations of the surface of the water from horizontal. Call this displacement η and count the downward direction as positive, see Fig. 5.2.

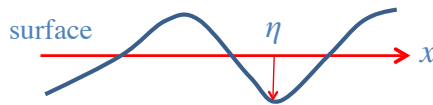


Figure 5.2: Deviations from the surface are designated by η .

Specifically we then assume that (i) this displacement is so small that all terms quadratic in η can be ignored, (ii) that the only restoring force is due to gravity so that $U = -\rho_0 g \eta$ and (iii) the atmospheric pressure at the surface can be ignored so that $p_s \simeq 0$. These assumptions, together with the assumed form (5.51) of the velocity potential yield

$$i\omega\rho_0 A e^{-k\eta} e^{i(kx-\omega t)} - \rho_0 g \eta = 0 \quad (5.52)$$

for (5.44), after setting the constant on the right hand side equal to zero. This is the only choice that does not interfere with the propagation of a travelling wave on the surface. But

$$k\eta \ll 1, \quad (5.53)$$

so $\exp(-k\eta)$ can be replaced by unity in (5.52) and we get

$$\eta(x, t) = \frac{i\omega}{g} A e^{i(kx-\omega t)}. \quad (5.54)$$

This relation provides an alternative way of calculating the velocity of the water at the surface, in particular the "sinking speed"

$$\frac{\partial \eta}{\partial t} = \frac{\omega^2}{g} A e^{i(kx-\omega t)} \quad (5.55)$$

in the y direction. But this must also be given by

$$u_y = -\text{grad}_y \varphi e^{-i\omega t} = k A e^{i(kx-\omega t)} \quad \text{at } y = 0. \quad (5.56)$$

Consequently, these two expressions are compatible only if ω depends on the wave number k like

$$\omega = \sqrt{gk}. \quad (5.57)$$

In turn this means that the *phase velocity* (the velocity with which planes of equal phase propagate)

$$u_p = \frac{\omega}{k} = \sqrt{\frac{g}{k}} = \sqrt{\frac{g}{2\pi}} \cdot \sqrt{\lambda} \quad (5.58)$$

is dependent on the wave length $\lambda = 2\pi/k$ of the oscillation. *Dispersion* is present; long wave lengths travel faster than short ones.

Since the restoring force causing this wave motion is gravity, the phenomenon we have just discussed is usually called a *gravity wave*. Waves encountered on the open sea are of this type.

Actually, the dispersion relation $\omega = \sqrt{gk}$ is only valid if the wave number k is not too high (the wave length not too short). At short wave lengths a new type of wave, called a *capillary wave*, can exist. This circumstance comes about because the hitherto neglected effect of the surface tension of the water enters the picture. The pressure at the surface due to surface tension is proportional to the *curvature* of the water surface (just like the restoring force on a string under tension):

$$p = T_0 \frac{\partial^2 \eta}{\partial x^2}. \quad (5.59)$$

Here, T_0 is a constant of proportionality called the surface tension (units: dynes/cm or Newtons/meter). For short wave lengths $\partial^2 \eta / \partial x^2$ is large and this term can no longer be neglected. Supplementing (5.52) on the left by this additional pressure term one has

$$i\omega\rho_0 A e^{i(kx-\omega t)} + T_0 \frac{\partial^2 \eta}{\partial x^2} - \rho_0 g \eta = 0 \quad (5.60)$$

for η . Assuming a travelling plane wave profile

$$\eta = \eta_0 e^{i(kx-\omega t)}, \quad (5.61)$$

for η once more, one finds that

$$i\omega A - k^2 \frac{T_0}{\rho_0} \eta_0 - g \eta_0 = 0. \quad (5.62)$$

The condition of matched sinking speeds becomes

$$kA = -i\omega\eta_0 \quad (5.63)$$

that provides a second relation between the unknown amplitudes A and η_0 . The two equations for A and η_0 are compatible only if

$$\begin{vmatrix} i\omega & -k^2 \frac{T_0}{\rho_0} - g \\ k & i\omega \end{vmatrix} = 0, \quad (5.64)$$

or

$$\omega = \sqrt{gk + k^3 \frac{T_0}{\rho_0}}. \quad (5.65)$$

If the wave length is short enough, the last term under the radical dominates, and we obtain the dispersion relation for pure capillary waves

$$\omega = \sqrt{T_0/\rho_0} k^{3/2}, \tag{5.66}$$

moving with a phase velocity

$$u'_p = \sqrt{\frac{T_0}{\rho_0}} \sqrt{k} = \sqrt{\frac{2\pi T_0}{\rho_0}} \frac{1}{\sqrt{\lambda}} \tag{5.67}$$

that increases with decreasing wave length. The phase velocity for a combined gravity-capillary wave motion can consequently be given by the curious form

$$u = \sqrt{u_p^2 + u_g^2}. \tag{5.68}$$

The wave length dependence of (5.68) as well as its gravity and

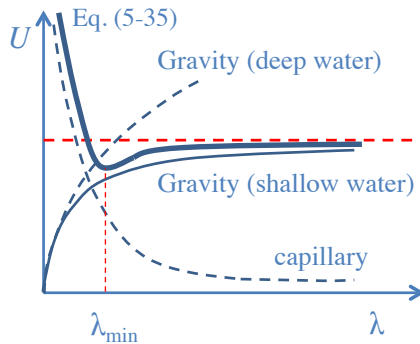


Figure 5.3: Phase velocity - wave-length relations for water waves.

capillary contributions are shown in Fig. 5.3. Notice that a minimum phase velocity occurs at the wave length

$$\lambda_{\min} = 2\pi \sqrt{\frac{T_0}{\rho_0 g}} \tag{5.69}$$

that serves as a rough dividing line between the wave lengths of capillary dominated ($\lambda \ll \lambda_{\min}$) and gravity dominated ($\lambda \gg \lambda_{\min}$) waves. For water ($T_0 = 73.05$ dynes/cm, $\rho_0 = 1$ gm/cc, $g = 980$ cm/sec²) this wave length is

$$\lambda_{\min} = 1,72 \text{ cm}, \tag{5.70}$$

giving a minimum phase velocity

$$u_{\min} = 23,13 \text{ cm/sec}. \tag{5.71}$$

Waves of a smaller velocity than this cannot propagate in water!

In the other limit where the wave number in (5.51) becomes very small (long wave lengths) we again get into trouble, but from a different

source; the sea bed. For if $\lambda \sim d$, the depth of the water, the boundary condition at the sea bed on the velocity potential has to be considered. We leave it to the reader to find the appropriate velocity potential and to show that there is no dispersion of water waves in the *shallow water* limit $\lambda \gg d$ (see Fig. 5.1 and Prob. 5-3).

5-5 Circular Waves on a Pond

Up to this point we have inquired into the manner of propagation of waves on the surface of a liquid without asking how the wave was generated in the first place. However, anyone who has ever thrown stones into a quiet pond or watched an insect swimming on its surface must have wondered about the beautiful wave patterns that develop on the surface of the water. As an interesting conclusion to our discussion of wave motion, we give a short mathematical analysis of the former problem.

Consider then, a small stone dropped into a pond, or a small insect settling on it. We assume that either event causes a small depression in the surface without imparting appreciable velocity to the water. Set up polar coordinates $x = r \cos \phi$, $z = r \sin \phi$ for a point P on the surface of the pond, see Fig. 5.4.

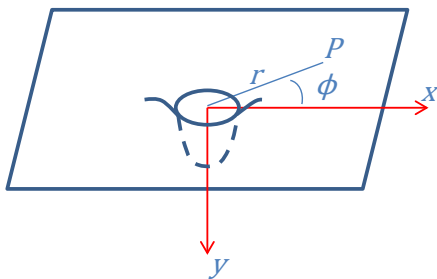


Figure 5.4: Initial disturbance of pond.

Now consider an initial disturbance

$$\eta(r, t = 0) = \frac{v_0}{\pi a^2} e^{-r^2/a}, \quad r = \sqrt{x^2 + a^2}, \quad (5.72)$$

v_0 being the volume of water displaced and a the "size" of the stone (or insect). The problem is then to determine how this disturbance propagates on the surface of the pond. Obviously such a disturbance is going to cause a wave pattern having a *cylindrical* symmetry about the vertical y axis, i.e. we are now dealing with a cylindrical rather than a plane wave propagation pattern.

We confine the discussion to a *deep* pond and consider a periodic disturbance first. If the wave pattern is to have a cylindrical symmetry instead of a plane one, we can expect the velocity potential to be given

by the expression (5.51) but with $\exp(ikx)$ replaced by a radial function $f(r)$:

$$\phi e^{-i\omega t} = A e^{-ky} f(r) e^{-i\omega t}. \quad (5.73)$$

Substituting this expression into Laplace's equation one readily finds that $f(r)$ must satisfy

$$\frac{d^2 f}{dr^2} + \frac{1}{r} \frac{df}{dr} + kf = 0, \quad (5.74)$$

that is Bessel's equation in the variable kr . Thus $f(r)$ is proportional to the Bessel function

$$f(r) = J_0(kr) \quad (5.75)$$

of order zero since ϕ must be regular at the origin.

The rest of the calculation for deep water plane waves transfers over to the present case with $J_0(kr)$ replacing $\exp(ikx)$ everywhere. Thus, the depression of the surface is now

$$\eta(r, t) = \frac{i\omega}{g} A J_0(kr) e^{-i\sqrt{gk}t}, \quad (5.76)$$

instead of (5.54), while the dispersion law (5.57) for deep water waves remains unchanged: $\omega = \sqrt{gk}$.

The problem is now to find a solution like $\eta(r, t)$ in (5.76) that satisfies Bernoulli's equation and *in addition* coincides with the prescribed shape (5.72) of η at $t = 0$. The solution to this problem can be solved by using the superposition principle. We only have to realize that the linear superposition

$$\int_0^\infty A(k) J_0(kr) e^{-i\sqrt{gk}t} dk \quad (5.77)$$

of the solutions (5.76) for arbitrary amplitudes $A(k)$ is also a solution of Laplace's equation. Replacing $A \exp(i(kx - \omega t))$ in (5.52) by this form of solution leads to the expression

$$\eta(r, t) = \int_0^\infty \frac{i\sqrt{gk}}{g} A(k) J_0(kr) e^{-i\sqrt{gk}t} k dk \quad (5.78)$$

for η . Equation (5.78) solves our problem. We only have to determine $A(k)$ such that $\eta(r, t = 0)$ assumes the form (5.72) at $t = 0$:

$$\frac{v_0}{\pi a^2} e^{-r^2/a^2} = \int_0^\infty i \frac{\sqrt{gk}}{g} A(k) J_0(kr) k dk. \quad (5.79)$$

The value of the coefficient $i\sqrt{gk}/g A(k)$ is given by the standard Fourier-Bessel inversion formula⁶⁴

$$i \frac{\sqrt{gk}}{g} A(k) = \frac{v_0}{\pi a^2} \int_0^\infty J_0(kr) e^{-r^2/a^2} r dr = \frac{v_0}{2\pi} e^{-(ka/2)^2}. \quad (5.80)$$

⁶⁴ P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill Book Company Inc., New York, 1953, Vol. I, p 766.

Hence,

$$\eta(r, t) = \frac{v_0}{2\pi} \int_0^\infty e^{-(ka/2)^2} J_0(kr) e^{-i\sqrt{gkt}} k dk. \tag{5.81}$$

This equation gives η for all r and t but does not yet tell us how the surface moves. We can obtain a qualitative idea of this by performing the integral of k approximately. Consider the case of a disturbance with a very small spatial extent ($a \rightarrow 0$) but keep the volume v_0 of water displaced constant. This is called a point disturbance. Also recede to a distance r from the point disturbance such that

$$kr \gg 1 \tag{5.82}$$

for all wave numbers k that count (this procedure will be justified *a posteriori*). Then $J_0(kr)$ in (5.81) can be replaced by its asymptotic form⁶⁵ for large kr ,

$$J_0(kr) \rightarrow \sqrt{\frac{2}{\pi}} \frac{\cos(kr - \pi/4)}{\sqrt{kr}}. \tag{5.83}$$

This turns the expression for η into

$$\begin{aligned} \eta(r, t) &\simeq \sqrt{\frac{2}{\pi}} \frac{v_0}{2\pi} \int_0^\infty \frac{\cos(kr - \pi/4)}{\sqrt{kr}} e^{-i\sqrt{gkt}} k dk \\ &= \sqrt{\frac{2}{\pi}} \frac{v_0}{2\pi r^2} \int_0^\infty \cos(u - \pi/4) e^{-2i\sqrt{\zeta}\sqrt{u}} \sqrt{u} du, \end{aligned} \tag{5.84}$$

where the last form follows on a change of variable from k to $kr = u$, and upon collecting the r and t dependence of the integrand into the single parameter

$$\zeta = \frac{gt^2}{4r}. \tag{5.85}$$

If this parameter is large, the integral in (5.84) can be evaluated approximately by the method of *saddle point* integration. To appreciate this procedure we first write the cosine as a sum of exponentials, so that

$$\int_0^\infty \cos(u - \frac{\pi}{4}) e^{-2i\sqrt{\zeta}\sqrt{u}} \sqrt{u} du = I_1 + I_2, \tag{5.86}$$

where

$$I_1 = \frac{1}{2} \int_0^\infty e^{i(u - 2\sqrt{\zeta}\sqrt{u} - \frac{\pi}{4})} \sqrt{u} du \tag{5.87}$$

and

$$I_2 = \frac{1}{2} \int_0^\infty e^{-i(u + 2\sqrt{\zeta}\sqrt{u} - \frac{\pi}{4})} \sqrt{u} du. \tag{5.88}$$

Take I_1 first. If ζ is large then the phase of the exponential

$$e^{if(u)}, \quad f(u) = u - 2\sqrt{\zeta}\sqrt{u} - \frac{\pi}{4} \tag{5.89}$$

changes rapidly with u giving rise to strong cancellations in the integrand. Thus the main contribution to the integral I_1 must come from a

⁶⁵ See for example W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Functions of Mathematical Physics*, Chelsea Publishing Company, New York, 1949.

region around a value for $u = u_0$ that makes the phase $f(u)$ stationary (if there is such a value). This value of u is determined by the condition $f'(u) = 0$, or

$$1 - \sqrt{\xi/u} = 0. \tag{5.90}$$

Thus $u_0 = \xi$ in the case of I_1 . Expanding $f(u)$ about this point we find

$$f(u) = -\left(\xi + \frac{\pi}{4}\right) + \frac{(u - \xi)^2}{4\xi} + \dots \tag{5.91}$$

Consequently

$$I_1 \simeq \frac{1}{2} e^{-i(\xi + \frac{\pi}{4})} \sqrt{\xi} \int_0^\infty e^{\frac{i(u-\xi)^2}{4\xi}} du. \tag{5.92}$$

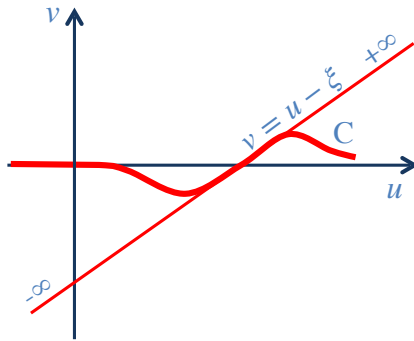


Figure 5.5: Contour C for the integral.

To evaluate this integral, we deform the path of integration along the real axis into the contour C shown in Fig. 5.5. This is legitimate since we have crossed no singularities of the integrand in the process. Thus, setting $z = u + iv$, we calculate

$$\int_C e^{\frac{i(z-\xi)^2}{4\xi}} dz. \tag{5.93}$$

But the point $z = \xi$ is a saddle point for the integrand in (5.93) (a minimum or a maximum depending on how you approach it), which therefore grows or recedes as we move away from ξ , depending on the path followed. Now, since

$$(z - \xi)^2 = (u - \xi + iv)^2 = [(u - \xi)^2 - v^2] + 2iv(u - \xi), \tag{5.94}$$

it is clear that along the path

$$v = u - \xi, \quad dv = du, \tag{5.95}$$

the oscillatory part of (5.93) is removed and the value of the integrand recedes. Consequently, we may extend the limits of integration to $\pm\infty$ along *this* path with impunity to obtain the approximate evaluation

$$\int_C e^{\frac{i(z-\xi)^2}{4\xi}} dz \simeq (1 + i) \int_{-\infty}^{+\infty} e^{-\frac{v^2}{2\xi}} dv = 2\sqrt{\xi}\pi e^{\frac{i\pi}{4}}. \tag{5.96}$$

Putting all this together, we find that

$$I_1 \simeq \sqrt{\pi} \zeta e^{-i\zeta}. \quad (5.97)$$

Turning now to I_2 , we find its phase is stationary at the root of

$$1 + \sqrt{\frac{\zeta}{u}} = 0, \quad (5.98)$$

which does not lie in the region of integration $u > 0$. Consequently,

$$I_2 \simeq 0 \quad (5.99)$$

to the same order of approximation as our expression for I_1 . Thus

$$\int_0^\infty \cos(u - \frac{\pi}{4}) e^{-2i\sqrt{\zeta}\sqrt{u}} \sqrt{u} du \simeq \sqrt{\pi} \zeta e^{-i\zeta}, \quad (5.100)$$

and the value of the depression η is

$$\eta(r, t) \simeq \sqrt{\frac{2}{\pi}} \cdot \frac{v_0}{2\pi r^2} \sqrt{\pi} \zeta e^{-i\zeta} = \frac{v_0}{\sqrt{2}\pi r^2} \zeta e^{-i\zeta}, \quad (5.101)$$

or, taking the real part,

$$\eta(r, t) \simeq \frac{v_0}{\sqrt{2}\pi r^2} \zeta \cos \zeta, \quad \zeta = \frac{gt^2}{4r} \gg 1. \quad (5.102)$$

Thus $\eta(r, t)$ represents a circular wave pattern moving away from the center of disturbance with a wave length that increases with increasing r at fixed time, but with an amplitude that decreases like r^{-3} . The steady build-up without limit of the amplitude like t^2 at a fixed point on the surface of the pond is of course unphysical; it arises from the use of a point disturbance with infinite amplitude [(5.72) becomes infinite at $r = 0$ as $a \rightarrow 0$] and the neglect of all dissipative processes.

Because the wave length of η decreases as we approach the source of the disturbance, one may well wonder what role the surface tension of the water plays; for we saw in Sec. 5-3 that the propagation of wave lengths below 1,72 cm (wave numbers above $3,7 \text{ cm}^{-1}$) in water are controlled more by surface tension effects than by gravity. We leave it to the reader to show that for capillary waves on the surface of a pond, the controlling parameter is

$$\zeta' = \frac{9 T_0 t^2}{4 \rho_0 r^3} \gg 1 \quad (5.103)$$

instead of ζ , where T_0 is the surface tension and ρ_0 the density of the water. Surface tension effects dominate gravity effects if $\zeta' \gg \zeta$ and *vice versa* if $\zeta' \ll \zeta$ (both factors of course remaining $\gg 1$). The dividing line

separating the regions of capillary and gravity dominated wave motion occurs at roughly $\zeta = \zeta'$, or a distance

$$r_{\text{crit}} = \sqrt{\frac{9T_0}{\rho_0 g}} \simeq 1 \text{ cm for water} \quad (5.104)$$

from the source of the disturbance. This value of r is independent of t . At distances further away from the source of disturbance than r_{crit} gravity effects rapidly dominate the motion, in agreement with what is seen when a quiet pond is disturbed by a small stone or an insect alighting on it.

Problems

5-1. Obtain Euler's equations for fluid motion by applying Newton's laws to the motion of a small volume of fluid.

5-2. Construct the *streamlines* for the velocity potential (5.51) given in the text. (A streamline is a line that has the fluid velocity tangent to it at every point in space). Hint: the streamlines are determined by the set of equations

$$\frac{dx}{u_x} = \frac{dy}{u_y} = \frac{dz}{u_z}, \quad (5.105)$$

where $\mathbf{u} = (u_x, u_y, u_z)$ is the fluid velocity at (x, y, z) .

5-3. Show that $\omega = \sqrt{gd}k$ for waves on a shallow ($kd \ll 1$) pond of depth d .

5-4. Determine the wave form on the surface of a pond initiated by the disturbance (5.72) if the pond is very shallow. (See previous problem).

5-5. Worry about the convergence of the integral appearing in (5.84).

5-6. Verify the statements made in the text in connection with (5.103). Investigate the role of surface tension by calculating $\eta(r, t)$ on the assumption that surface tension alone is responsible for the wave motion. Indicate any modifications of Fig. 5.5 that your results suggest to be important.

5-7. Develop equations describing a sound wave in a gas. Obtain an expression for the velocity of sound in the gas paying particular attention to the condition under which you derive your result. Consider further the propagation of a periodic sound wave past a sharp edge, like a vertical wall. Indicate what one might expect as to the distribution of sound energy on the other side of the wall. (Hint: look up some characteristics of sound waves in, for example P.M. Morse, *Vibration and Sound*, McGraw-Hill Book Company, Inc., New York, 1948, and develop methods to treat this particular problem).

Chapter 6 Relativistic Mechanics

6-1 Introduction

In Chapter 1 we pointed out the necessity of a frame of reference for describing dynamical processes, and noted that Newton's law of motion assumes the simple form $\mathbf{p} = \mathbf{F}$ in an inertial frame of reference. We saw further that an arbitrary number of such inertial frames exist if one does, and that the laws of motion are identical in form in all such frames. The invariance of form of the laws of motion under change of inertial reference frame constitutes *Galileo's principle of relativity*. However, for this principle to be strictly valid it is necessary that interactions between the constituents of a dynamical system be *instantaneous*, that is to say any change of the state of motion in one of the constituents must immediately be sensed by all the other constituents in interaction with it. Now, experiment shows that instantaneous interactions do not in fact exist in nature. Instead, there is a finite time lag between the "emitting" and "sensing" of the interaction between two electric charges, for example, and thus a finite velocity of propagation of interaction. We call this velocity of propagation c and identify its physical significance in a moment.

Given that all interactions propagate with a universal velocity c it is easy to see that Galileo's principle of relativity breaks down. For according to the Galilean transformation of velocities given by (1.4) of Chapter 1, the velocity c will clearly depend on the frame of reference in which the interaction is viewed. Thus, the corresponding laws of motion will become inertial frame-dependent, in violation of Galileo's relativity principle.

We must now consider the significance of the velocity of propagation c . It is known experimentally that electromagnetic waves (light) propagate in vacuo with a velocity that is *independent* of the frame of reference in which this velocity is measured (the famous Michelson-Morley experiment). Therefore, the velocity of light is certainly a candidate for c as the *universal* velocity. The thoughtful reader may wonder about other candidates for c , like the velocity of neutrinos or gravitons, for example. However, the important point here is that there can be only *one* univer-

sal velocity if all the tenets of the special theory of relativity are to be satisfied. In what follows we will identify c with the velocity of light in vacuo, $c \simeq 3 \times 10^{10}$ cm/sec. However, its real significance for relativity theory as a universal velocity must always be borne in mind.

6-2 Einstein's Principle of Relativity

After the violation of Galileo's principle of relativity for light waves became apparent from Michelson's experiments in the 1880's, there were numerous attempts to modify the (then) current "aether" theory of light to account for this apparent violation. None of these attempts were fully successful, although some were extremely ingenious. There simply seemed to be no way of reconciling Galileo's principle with Michelson's measurements.

The way out of this dilemma was resolved by Einstein in 1905. Einstein's special theory of relativity, or the relativity theory of inertial frames, rests on two postulates: (a) The universal velocity c is the same in all directions for all observers moving uniformly with respect to each other, and (b) all laws of nature are identical in all inertial frames of reference. Postulate (b) sounds like Galileo's relativity principle again. However, we will see that postulate (a) compels us to rethink our (somewhat intuitive) concept of the absolute nature of time. We introduce an inertial frame Σ and a second one Σ' moving with a uniform velocity \mathbf{V} relative to Σ (see Fig. 6.1).

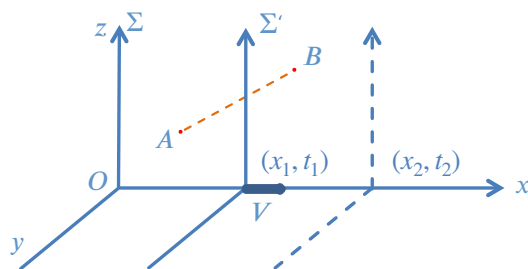


Figure 6.1: The inertial frames Σ and Σ' .

We attach observers rigidly to Σ and Σ' and supply them with identical devices for measuring distances and quantity of elapsed time in the respective frames. Both observers record the emission and absorption of a light ray travelling from A to B . The observer in Σ assigns coordinates (x_a, y_a, z_a) and a time t_a to the event of emission at A , and coordinates (x_b, y_b, z_b) and a time t_b to the event of absorption at B . Knowing the velocity of light to be c , he concludes that

$$c^2(t_b - t_a)^2 = (x_b - x_a)^2 + (y_b - y_a)^2 + (z_b - z_a)^2, \quad (6.1)$$

since both sides of this equation express the distance AB . The observer

in Σ' observes the same two events, assigns coordinates (x'_a, y'_a, z'_a, t'_a) and (x'_b, y'_b, z'_b, t'_b) to them, and concludes that

$$c^2(t'_b - t'_a)^2 = (x'_b - x'_a)^2 + (y'_b - y'_a)^2 + (z'_b - z'_a)^2, \quad (6.2)$$

since by postulate (a) the velocity c is the *same* in Σ and Σ' .

The appearance in these relations of time intervals $t_b - t_a$ and $t'_b - t'_a$ that depend on the reference frame to which the time measuring device (a "clock") measuring them, is attached, is something completely alien to the Newtonian idea of absolute time. Instead, we are forced to assert that there is no such thing as absolute time. The quantity of elapsed time depends on which observer makes the measurement.

We thus see that it becomes a natural idea to associate the *four* coordinates (x, y, z, t) with a "happening" like the emission or absorption of a light ray. The coordinates (x, y, z, t) are said to label an *event* in a given reference frame. An event can therefore be represented as a point, called a *world point* in a fictitious four-dimensional space, or *space-time*, with axes labelled by x, y, z and t . The motion of a particle is then represented by the curve (called a *world line*) traced out by a succession of world points in this four-dimensional space.

So far (6.1) and (6.2) refer to the motion of light rays. We now generalize the points A and B to refer to *any* two events and define the interval between these two events as

$$s_{ab}^2 = c^2(t_b - t_a)^2 - (x_b - x_a)^2 - (y_b - y_a)^2 - (z_b - z_a)^2. \quad (6.3)$$

If A and B are infinitesimally separated events at (x, y, z, t) and $(x + dx, y + dy, z + dz, t + dt)$, then $s_{ab} = ds$, where

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2. \quad (6.4)$$

We now assert that the infinitesimal interval ds between any two events is the same in all inertial frames. There are several ways of demonstrating this assertion. We follow the elegant argument of Landau and Lifshitz⁶⁶. Consider again the two frames Σ and Σ' . According to (6.1) and (6.2) $ds = 0$ in Σ and ds' in Σ' also vanishes, $ds' = 0$, for light ray propagation from A to B , the latter now being infinitesimally separated from A . Hence ds and ds' must be infinitesimals of the same order, or $ds' = \eta ds$, where η is a factor of proportionality. A little reflection shows that η can at most depend on the magnitude of the relative velocity \mathbf{V} of Σ' relative to Σ , i.e. $\eta = \eta(|\mathbf{V}|)$. Any spatial or temporal dependence in η would violate the homogeneity of space and time, while a directional dependence on \mathbf{V} would violate the isotropy of space. But if $\eta = \eta(|\mathbf{V}|)$, we may equally well view Σ from Σ' and conclude that $ds = \eta ds'$, or that $\eta^2 = 1$. However, the result must hold for any velocity, so that we cannot choose $\eta = +1$ for some velocities and $\eta = -1$ for others. Also

⁶⁶ L. Landau and E. Lifshitz, *The Classical Theory of Fields*, Addison-Wesley Inc., Cambridge, Mass., 1951), p.6.

$ds = \eta ds'$ must include the identity $ds = ds'$ when $\mathbf{V} = 0$. Hence, $\eta = 1$ and $ds = ds'$ has the same value in all inertial frames. Since $ds = ds'$ it follows directly that any finite interval has the same value, $s = s'$ in all inertial frames.

6-3 Proper time

We have seen that the time lapse between two events depends upon the frame of reference from which the events are viewed. So let us consider the special event of a clock *fixed rigidly* to Σ' and therefore moving with velocity \mathbf{V} . The moving clock is at rest in Σ' , so that $dx' = dy' = dz' = 0$ and it registers a time lapse dt' (say) in the same period that it is observed to move from (xyz) to $(x + dx, y + dy, z + dz)$ in Σ in a time dt as measured by a clock attached to Σ . The invariance of intervals lets us write

$$ds^2 = c^2 dt'^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2, \quad (6.5)$$

or

$$dt' = dt \sqrt{1 - \frac{v^2}{c^2}} = \frac{ds}{c}, \quad (6.6)$$

since $[(dx/dt)^2 + (dy/dt)^2 + (dz/dt)^2]^{1/2}$ is just the speed $v = |\mathbf{V}|$ of the clock attached to Σ' as seen from Σ . The interval dt' is called the *proper time*. It is the time recorded by a clock moving with a material particle. The proper time lapse for events recorded between t_1 and t_2 in Σ is given by integration as

$$t'_2 - t'_1 = \int_{t_1}^{t_2} dt \sqrt{1 - \frac{v^2}{c^2}}. \quad (6.7)$$

We comment that the speed v in (6.6) and (6.7) need *not* be uniform, i.e. v may vary with time. The point is simply that a clock performing an arbitrary motion can be considered as moving uniformly with whatever speed it has at each moment of time. We may therefore introduce a different inertial frame at each moment of time, and the argument leading to (6.6) goes through as before.

The relation (6.7) shows that the proper time interval $t'_2 - t'_1$ is always *less* than the time interval $t_2 - t_1$ as measured in Σ . The moving clock runs slower than the clock at rest relative to the observer. Results of this nature lead to several apparent "paradoxes" in the theory of relativity. However, since we intend to present only as much background of the special theory as is required for discussing relativistic mechanics in a coherent way, we resist the temptation of too much digression into those fascinating aspects of the special theory and refer the reader to the literature instead⁶⁷.

⁶⁷ A good reference for further reading is C. Moller, *The Theory of Relativity*, Oxford University Press, 1952.

6-4 The Lorentz Transformation

We now establish how the two sets of coordinates (x, y, z, t) and (x', y', z', t') of the same event in Σ and Σ' are related. To do so we introduce an imaginary time coordinate $x_4 = ict$ in place of t , and locate an event by $(x, y, z, t) \rightarrow (x_1, x_2, x_3, x_4)$ in Σ . The link between these coordinates and the corresponding coordinates (x'_1, x'_2, x'_3, x'_4) of the same event in Σ' then follows from the invariance property $s = s'$. Written out in terms of the coordinates of two events A and B , the invariance of s reads

$$-s^2 = (x_{1a} - x_{1b})^2 + (x_{2a} - x_{2b})^2 + (x_{3a} - x_{3b})^2 + (x_{4a} - x_{4b})^2 = \text{invariant} \quad (6.8)$$

if the imaginary-time coordinate x_4 is used. Equation (6.8) is just a statement of the invariance of "distance" between any two world points in the fictitious four-dimensional space, or *Euclidean space*, characterized by orthogonal axes O_{x_1, x_2, x_3, x_4} . (Alternatively, we could have introduced the notation $(x_0 = ct, x_1, x_2, x_3)$, which forms the basis of *Minkowski space*, but we will not use that here). Consequently, the transformation $(x_1, x_2, x_3, x_4) \rightarrow (x'_1, x'_2, x'_3, x'_4)$ must leave all distances in *this* four-dimensional space unchanged. If we disregard parallel displacement of axes, which merely relabels the coordinate origin, then the coordinates of the same event seen in Σ and Σ' must be connected by a rotation of coordinates in four dimensions,

$$\mathbf{x}' = A\mathbf{x}. \quad (6.9)$$

Here, A is formally the same as the rotation matrix in (3.13), $A^T A = A A^T = I$ and $\text{Det} A = +1$, while \mathbf{x}' and \mathbf{x} are column vectors of the coordinates of the same event in Σ and Σ' .

We now determine the elements of A under the assumption that the axes defining Σ and Σ' are parallel and that the origin of Σ' proceeds along the positive x -axis of Σ with velocity \mathbf{V} . Then the coordinates $x_2 = y$ and $x_3 = z$ remain unchanged, while $x_1 = x$ and $x_4 = ict$ must transform like

$$\begin{pmatrix} x' \\ x'_4 \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} x \\ x_4 \end{pmatrix} \quad (6.10)$$

in accordance with (?). The analogy here with a rotation in ordinary space is purely formal, however. The angle ϕ is merely a *parameter* in (6.10) (it will turn out to be pure imaginary) that has nothing to do with rotations in real space. However, the representation of the operator A by the 2×2 matrix on the right-hand side of (6.10) is very convenient, since the conditions on A given below (6.9) are now automatically satisfied. To determine A , we measure time from the moment that the reference frames Σ and Σ' are coincident. We follow the motion of the origin of Σ'

from Σ and deduce from the inverse transformation to (6.10) that

$$\begin{pmatrix} x \\ x_4 \end{pmatrix} \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} 0 \\ x'_4 \end{pmatrix}, \quad (6.11)$$

since now $x'_4 = 0$. This gives the ratio $x/x_4 = -\tan \phi$. However, $x/t = v$ is the speed of Σ' relative to Σ . Hence, $\tan \phi = iv/c$, so that

$$\cos \phi = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}; \quad \sin \phi = i \frac{v/c}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (6.12)$$

Inserting these expressions into (6.10), one obtains the *Lorentz transformation*

$$x' = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - vx/c^2}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (6.13)$$

after reverting to real time coordinates again. The inverse transformation follows upon reversing the sign of \mathbf{V} . (Σ has the velocity $-\mathbf{V}$ relative to Σ'):

$$x = \frac{x' + vt'}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad y = y', \quad z = z', \quad t = \frac{t' + vx'/c}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (6.14)$$

One notices that the transformation becomes singular for a reference frame moving with velocity c , and also that x and t both become imaginary for frames moving with $v > c$, corresponding to the impossibility of such motion for material particles⁶⁸.

The particular feature of the transformation (6.13) or its inverse (6.14) is that space and time coordinates get mixed up with each other. A knowledge of *both* x and t is necessary to compute x' and t' from (6.13).

We will see in the following examples of the use of the Lorentz transformation how essential this feature is if observers in Σ and Σ' are to arrive at consistent conclusions about the outcome of experiments performed in their respective frames.

(i) *Time-dilation*: Let us re-examine the phenomenon of time-dilation given by (6.6) or (6.7) from the point of view of the Lorentz transformation. A clock attached to the origin of Σ' records a time difference $t'_2 - t'_1$ in the time $t_2 - t_1$ (as seen by an observer in Σ) that the origin of Σ' moves from x_1 to x_2 as measured in Σ , see Fig. 6.1. Since the clock in Σ' is at the origin, the same two events (x_1, t_1) and (x_2, t_2) in Σ have coordinates $(0, t'_1)$ and $(0, t'_2)$ in Σ' . Hence, from (6.14) we conclude that

$$t_2 - t_1 = \frac{t'_2 - t'_1}{\sqrt{1 - \beta^2}} > t'_2 - t'_1, \quad \beta = \frac{V}{c}, \quad (6.15)$$

since the time-dilation factor $(1 - \beta^2)^{-1/2}$ is always greater than unity. Equations (6.15) and (6.7) agree of course for a constant time-dilation factor in the latter equation.

⁶⁸ This is the standard interpretation in classical physics. There are some dissenters, however. Following up ideas put forward by O.M.P. Bilaniuk, V.K. Deshpande and E.C.G. Sudershan, *Am. J. Phys.* **39**, 718 (1962), G. Feinberg, *Phys. Rev.* **159**, 1089 (1967) has developed a tentative theory of faster-than-light particles ("tachyons"). No experimental evidence exists to date for the existence of such objects.

Thus, moving clocks run slow in comparison with the clocks of the stationary observer. A similar analysis shows that the observer in Σ' (who views Σ as moving) claims that Σ' 's clock is running slow. Each observer thinks that the other's clock is slow!

(ii) *Simultaneity*: Time-dilation also plays havoc with the "common sense" idea of simultaneity, which now becomes a relative concept too: Events which occur simultaneously with respect to a stationary observer, are not simultaneous when viewed by a moving observer. For, suppose two events A and B a distance d apart in their rest frame Σ' occur simultaneously, i.e. have the same time coordinate (zero say) in *this* frame. Now let Σ' move with velocity \mathbf{V} along the x -axis of a second frame Σ as in Fig. 6.1. The events A and B have space-time coordinates (x_a, t_a) and (x_b, t_b) in Σ . Their time-separation in Σ is thus

$$\Delta t = t_b - t_a = \frac{vd}{c^2} \frac{1}{\sqrt{1 - \beta^2}}, \quad \beta = \frac{v}{c}, \quad (6.16)$$

by (6.14). But according to (6.15), the clock at A would register an interval

$$\Delta t' = \Delta t \sqrt{1 - \beta^2} = \frac{vd}{c^2} \quad (6.17)$$

in the interval Δt , i.e. would be found to be running *ahead* of the clock at B by $\frac{vd}{c^2}$. The moving observer would claim that the A and B events are not simultaneous.

Lest the reader feel that such a state of affairs is pure Lewis Carroll, we examine these conclusions by way of experiment. One of the best known examples of time-dilation shows up in the decay of μ -mesons that are produced in the upper atmosphere in cosmic ray phenomena. The μ -meson (muon) decays into an electron and two neutrinos with a lifetime of about 2 microseconds when the muon is *at rest*. Even moving with a velocity near that of light, this would imply a penetration depth into the earth's atmosphere now exceeding $d \simeq 600$ meters. Yet a copious flux of muons is observed at sea level ($d \simeq 2000$ meters). How does the muon get where it got? Time-dilation tells us that if the muon has a lifetime of $\tau_0 \simeq 2\mu$ sec in its rest frame, we observe it to decay at the slower rate of $\tau_0(1 - \beta^2)^{-1/2}$, and thus travel a distance $d = \beta c \tau_0(1 - \beta^2)^{-1/2}$ into our atmosphere. Experimentally, a factor $(1 - \beta^2)^{-1/2} \simeq 10$ is required to explain the observed muon flux at sea level. An independent measurement⁶⁹ of the muon energy (to find β) confirms this number.

(iii) *Lorentz contraction*: But how does this state of affairs appear to the moving muon? In its rest frame it still has a lifetime τ_0 which is too short to allow it time to penetrate down to sea level. The answer to this apparent paradox is provided by an effect similar to time dilation, *viz.* a moving measuring rod is shortened relative to its length at rest. This is called the Lorentz contraction. To prove this statement, we provide the observer in Σ' with a measuring rod which is laid along the x' axis.

⁶⁹ D.H. Frisch and J.H. Smith, Am. Journ. of Physics, **31**, 342 (1962).

The ends of the rod are observed by Σ' to have coordinates x'_1 and x'_2 , so its length is $l_0 = x'_2 - x'_1 > 0$ in the rest system. The *times* at which these measurements are made are immaterial. The length l_0 is called the *proper length* of the rod. What length does the observer in Σ measure? He assigns space-time coordinates (x_1, t_1) and (x_2, t_2) to the ends of the rod. Thus the length he gets will depend on the *times* that he makes his observation. Clearly what we mean by length of a rod is its length "now", i.e. he must measure the coordinates of both ends *simultaneously*. (Operationally this means placing observers along the x -axis in Σ with clocks that have previously been synchronized by light signals, say). These measurements will not be simultaneous in the rest system of the rod, but that is of no consequence. The recorded length, obtained in this way in Σ is noted as $l = x_2 - x_1$. But from (6.13)

$$l_0 = x'_2 - x'_1 = \frac{x_2 - x_1}{\sqrt{1 - \beta^2}} = \frac{l}{\sqrt{1 - \beta^2}}, \quad (6.18)$$

or

$$l = l_0 \sqrt{1 - \beta^2} < l_0. \quad (6.19)$$

The length l_0 of the moving rod is contracted by the factor $\sqrt{1 - \beta^2}$. Just as moving clocks ran slow, we now discover that moving rods get shorter. Observe the complementary nature of these two statements. From the point of view of the Lorentz transformation one cannot have one without also having the other, and it is this feature which removes the muon's dilemma. For, from the point of view of the muon, it is the atmosphere that is moving. Consequently the muon only has to go *this* distance in its own lifetime τ_0 , i.e. $d\sqrt{1 - \beta^2} = \beta c\tau_0$, which is exactly what our earthbound observer calculates. We record in passing that for a long time it was tacitly assumed that a Lorentz contraction would be visible as a flattening of a three dimensional object in the direction of its motion. This assumption is *false*, as has been shown by Terrell J.L. Terrell, *Phys. Rev.* **116**, 1041 (1959). See also V.F. Weisskopf, *Physics Today* **13**, 24 (1960). The point is that what one means by "see" is the ocular image created in one's brain by light rays entering the eye at a given instant in time. These rays necessarily leave an extended body at different times, and can be shown to give the viewer the impression of a rotation, rather than a compression of the object.

(iii) Doppler Shift

The discussion so far has carefully avoided the question of what happens to clocks and meter sticks when viewed from non-inertial frames of reference. Such questions properly belong to the province of Einstein's General Theory of Relativity. However, it is possible to illustrate some effects of non-inertial frames within the framework of the special theory by looking at the Doppler shift. The Doppler shift, which has nothing to do with non-inertial frames *per se*, refers to the

phenomena, shared by all wave-like disturbances, whereby the frequency of the emitted vibration appears to be higher (lower) to an observer moving towards (away from) the source. The analysis for a light emitter runs as follows. If the natural frequency of the source is ν_0 (i.e. the frequency recorded by an observer at rest relative to the source), then it has a frequency $\nu = \nu_0 \sqrt{1 - \beta^2}$ by time-dilation in the reference frame of an observer moving relative to the light source with speed $v = \beta c$. The time between successive light pulses in the observer's frame is thus ν^{-1} . However, in this time the pulse travels a distance c/ν , the source a distance $\beta c/\nu$, so that the observer records pulses spaced at a distance $(1 - \beta)c/\nu$ if he is approaching the source. The number of these pulses (which travel with velocity c) per second is thus recorded by the observer as

$$\nu_D = \frac{\nu}{1 - \beta} = \nu_0 \sqrt{\frac{1 + \beta}{1 - \beta}} \quad (6.20)$$

an equation which expresses the relativistic Doppler shift: The natural frequency of an approaching source is shifted to a higher value ("blued"), that of a receding source to a lower value ("reddened").

(v) *Falling Photons turn Blue*

By means of the Doppler shift we can study the effect of a uniform gravitational field on the frequency of a light source placed in that field. The argument is based on the experimental equivalence of gravitational and inertial mass, plus the validity of Newton's laws of motion. Given these facts, one demonstrates that the motion of a mass in falling under gravity with acceleration g is indistinguishable from the motion of the same mass moving in gravity-free space, but viewed from a *non-inertial* reference frame accelerating upwards with acceleration g . This statement is an example of Einstein's *equivalence principle*. Here it follows as a consequence of known experimental facts in the special case of the constant gravitational acceleration of a mass point. Einstein elevated it to a *general* principle, applicable to arbitrary gravitational accelerations, and embracing the motion of light rays as well as particles⁷⁰. We can use this principle to study the behavior of a light source in a uniform gravitational field. Consider a stationary source S of natural frequency ν_0 placed at a vertical height h above an observer on the surface of the earth. If h is small, the gravitational field in which source (and observer) find themselves has uniform value g ($\sim 980 \text{ cm/s}^2$) pointing from source to observer. The equivalence principle maintains that any effect of this gravitational field on the source S must be the same as what would be observed if source *plus* observer were moving *upwards* with acceleration in gravity-free space. Starting from rest in the latter situation, the observer reaches a velocity $v = gh/c$ in the time h/c that a light-pulse from S takes to reach him if we assume he does not change position appreciably during this time. But since the observer is moving

⁷⁰ A concise discussion may be found in L. Landau and E. Lifschitz, *loc. cit.*, Chapter 10.

towards the source, he records a Doppler shift in the source frequency of

$$\frac{\nu_D - \nu_0}{\nu_0} \simeq \beta = \frac{gh}{c^2} \quad (6.21)$$

by (6.20) with $\beta = gh/c^2 \ll 1$. This equation holds equally for a *stationary* source in a uniform gravitational field, by the equivalence principle. It says that a source at a higher gravitational potential $\phi_S = gh$ appears "blue-shifted" to an observer at a lower potential $\phi_0 = 0$, by an amount

$$\Delta\nu = \frac{\nu_0}{c^2}(\phi_S - \phi_0), \quad (6.22)$$

where $\phi_S - \phi_0 = gh$. The validity of (6.22), which coincides with the more general result for weak but not necessarily uniform fields, $\phi_S - \phi_0 \neq gh$, that one obtains in the General Theory⁷¹, has been demonstrated in an earthbound laboratory. In a remarkable experiment, Pund and Rebka⁷² have shown that photons emitted by ^{57}Fe were "blued" by just the predicted amount after "falling" a measured distance under gravity. Previously, the prediction (6.22) could only be tested in terms of the "red-shift" expected for light emitted by massive, distant stars (see Problems).

Two aspects of the result (6.22) should be emphasized. Firstly if observer and source interchange positions, so that the observer is at a higher potential relative to the source, the light reaching him will be "red-shifted". Stated more specifically in terms of atomic periods of the atoms constituting the light source, this means that all observers *agree* that the "clock" at a lower gravitational potential runs slower than an identical clock at a higher potential. This contrasts for example with the opinions of observers in uniform relative motion, who always maintain that the other's clock is running slow. The second point to make in regard to (6.22) is that the General Theory teaches us that the speeding up of an atomic clock in a gravitational field, rather than the loss of "gravitational potential energy" by the emitted photons is the correct physical interpretation of this relation.

An appreciation of the different behavior of clocks in inertial vs non-inertial reference frames is also the clue to understanding the oft-discussed "twin paradox". There now seems to be consensus that the travelling twin *does* age slower, and furthermore that *both* twins agree that this is so (if the effect of accelerations on the observations of the travelling twin are not considered the twins don't agree: That is the paradox). A careful discussion of the twin paradox is only really possible using the full machinery of the General Theory of Relativity. We refer the reader to Moller's book quoted at the beginning of this section for such discussions.

⁷¹ L. Landau and E. Lifschitz, *ibid.*, Chapter 10.

⁷² R.V. Pound and G.A. Rebka, *Phys. Rev. Lett.*, 4, 337 (1960).

6-5 Four-vectors and Lorentz-invariance

We saw in Sec. 6-4 that the coordinates of an event in space-time are related by $x' = Ax$, or

$$x'_\mu = \sum_{\nu=1}^4 A_{\mu\nu} x_\nu, \mu = 1, 2, 3, 4 \quad (6.23)$$

when viewed from two different inertial frames Σ or Σ' . Here, $A_{\mu\nu}$ form the elements of the orthogonal transformation matrix A , for which $\text{Det}A = +1$ must hold. For the special case considered previously, where Σ' moves along the positive x -axis of Σ with speed V and coincides with Σ at $t = 0$, (6.23) has the explicit form

$$\begin{pmatrix} x' \\ y' \\ z' \\ ict' \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ ict \end{pmatrix}, \quad (6.24)$$

where $\beta = v/c$ and $\gamma = (1 - \beta^2)^{-1/2}$. Of course this is just a more succinct way of writing down the result (6.13). The inverse transformation is

$$\begin{pmatrix} x \\ y \\ z \\ ict \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & -i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ i\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \\ ict' \end{pmatrix}. \quad (6.25)$$

Any four-component entity which transforms like x_μ is called a *four-vector*. Apart from the position four-vector $x_\mu = (x, y, z, ict) = (\mathbf{r}, ict)$ of a particle, the important four-vectors for describing relativistic motion are four-vectors associated with the velocity, momentum and acceleration of a particle. It is clear that the generalization of the concepts in Special Relativity has to be done with one eye on the transformation property (6.23). Otherwise we will run into trouble trying to satisfy the second part of Einstein's Principle of Relativity, viz. that all laws of motion be identical in all inertial frames. Consider the concept of particle velocity. An observer in Σ watches a particle change its position $\mathbf{r}(t)$ by $d\mathbf{r}$ in a time dt as measured by his clock and assigns it a velocity $\mathbf{v} = d\mathbf{r}/dt$. The change in four-vector position in this time is

$$dx_\mu = [d\mathbf{r}, ict], \quad (6.26)$$

which is also a four-vector. However, the quotient dx_μ/dt is *not*, since both dx_μ and dt change when viewed from different reference frames. Therefore, how the time-dependence of x_μ is to be characterized requires some care. We have seen that the proper time interval ds/c , i.e. the time interval as measured in a reference frame attached to the particle and

moving with it, is the same in all inertial frames. We distinguish the time interval by the symbol $d\tau$ from now on, $d\tau = ds/c$. Then, not dx_μ/dt , but rather

$$v_\mu = \frac{dx_\mu}{d\tau} = \frac{dx_\mu}{dt} \frac{dt}{d\tau} = \left[\frac{\mathbf{v}}{\sqrt{1-\beta^2}}, \frac{ic}{\sqrt{1-\beta^2}} \right] \quad (6.27)$$

transforms like a four-vector if we regard the position four-vector as a function of the proper time of the particle, $x_\mu = x_\mu(\tau)$. The construct v_μ is called the *four-velocity*. The last step in (6.27) follows from the time-dilation between observer-time dt and proper-time $d\tau = dt\sqrt{1-\beta^2}$, where $v = \beta c$ is the particle velocity. A second differentiation with respect to τ produces another four-vector,

$$f_\mu = \frac{dv_\mu}{d\tau} = \left[\frac{1}{\sqrt{1-\beta^2}} \frac{d}{dt} \left(\frac{\mathbf{v}}{\sqrt{1-\beta^2}} \right), \frac{ic}{\sqrt{1-\beta^2}} \frac{d}{dt} \left(\frac{1}{\sqrt{1-\beta^2}} \right) \right], \quad (6.28)$$

called the *four-acceleration*.

Rotations in real three-dimensional space leave the length, or norm, of a vector unchanged. The same is true for the norm of a four-vector, which is unchanged by the transformation (6.23). This is called the *Lorentz-invariance* of the norm. Similarly, the scalar product $\sum_\mu A_\mu B_\mu$ of two four-vectors A_μ and B_μ is a Lorentz invariant. The property of Lorentz invariance is a very useful one, since it allows us to calculate norms and scalar products of four-vectors in any convenient frame. For example, the norm of the four-velocity v_μ in the frame attached to the particle, or *rest frame* (we denote all such quantities by a prime) is simply $\sum v_\mu'^2 = -c^2$. Hence, by Lorentz invariance

$$\sum_\mu v_\mu^2 = -c^2 \quad (6.29)$$

holds for v_μ in any inertial frame. Likewise, the scalar product $\sum_\mu f_\mu v_\mu$ vanishes in any inertial frame, since

$$\sum_\mu f_\mu v_\mu = \frac{1}{2} \frac{d}{d\tau} \left(\sum_\mu v_\mu^2 \right) = \frac{1}{2} \frac{d}{d\tau} \left(\sum_\mu v_\mu'^2 \right) = 0 \quad (6.30)$$

and $\sum_\mu v_\mu'^2$ is constant in the rest frame of the particle.

The construction of v_μ and f_μ , as well as their properties, only depended on the *form* of the transformation (6.23) and the relation between proper time and observer time. The actual values of the $A_{\mu\nu}$ were not relevant. The developments in the following paragraphs will also be of this nature. Examination of the behavior of entities under transformations that respect the fundamental invariance of the infinitesimal interval $ds = cd\tau$ will become the guiding element when we discuss particle kinematics and particle dynamics. Such a point of departure has a much more dependable mathematical "feel" than the repeated introduction of

"Gedanken" experiments involving clocks and measuring rods. However, we have to use the Lorentz transformation explicitly whenever the values of a four-vector are required in two different frames. The question of how to combine velocities in relativity is a case in point. An observer in a frame Σ' , moving with velocity V along the x -axis of a second frame Σ (see Fig. 6.1 again), records the velocity of a particle as v' . What velocity does this particle have with respect to Σ ? From (6.14), the space and time increments as seen from Σ and Σ' are related by

$$dx = \frac{dx' + Vdt'}{\sqrt{1 - V^2/c^2}}, \quad dy = dy', \quad dz = dz', \quad dt = \frac{dt' + \frac{V}{c^2}dx'}{\sqrt{1 - V^2/c^2}}, \quad (6.31)$$

so that, since $\mathbf{v} = d\mathbf{r}/dt$ is the velocity recorded in Σ , $\mathbf{v}' = d\mathbf{r}'/dt'$ that recorded in Σ' ,

$$v_x = \frac{dx' + Vdt'}{dt' + \frac{V}{c^2}dx'} = \frac{v'_x + V}{1 + Vv'_x/c^2} \quad (6.32)$$

$$v_y = \frac{dy'}{dt' + \frac{V}{c^2}dx'} \sqrt{1 - \frac{V^2}{c^2}} = \frac{v'_y \sqrt{1 - \frac{V^2}{c^2}}}{1 + Vv'_x/c^2} \quad (6.33)$$

$$v_z = \frac{dz'}{dt' + \frac{V}{c^2}dx'} \sqrt{1 - \frac{V^2}{c^2}} = \frac{v'_z \sqrt{1 - \frac{V^2}{c^2}}}{1 + Vv'_x/c^2}. \quad (6.34)$$

The curious structure of these equations runs contrary to intuition as with so many results in relativistic physics. However, the common denominator $1 + Vv'_x/c^2$ is essential if the velocity c is not to be exceeded. For consider a light ray moving along the x -axis. Its velocity is recorded as c in Σ' . Now let Σ' be moving with velocity c relative to Σ . The velocity Σ records is $v_x = c$ according to (6.32). However, the *direction* that a light ray travels relative to observers in Σ and Σ' is affected by their relative motion (abberation of light, see Problems).

6-6 Momentum and Energy

We now wish to study the motion of a free particle of mass m moving with velocity v , by means of the Principle of Least Action that was discussed in Chapter 1. We do so by constructing an integral for the action function S that satisfies the following two conditions. (i) The integral must be Lorentz-invariant in order to satisfy the Principle of Relativity and (ii) its integrand must be a differential of the first order. The only Lorentz-invariant differential of first order that refers to a single particle is the invariant interval ds introduced in (6.4). In addition the integrand for S must have the dimensions of energy \times time. The only combination of this nature that is available for a free particle is

$mc^2 d\tau = mc ds$. Hence,

$$S_{\text{free}} = -mc \int_{s_1}^{s_2} ds \quad (6.35)$$

is an expression for S with the required attributes. The integral is taken along a world line of the particle that connects the two events s_1 and s_2 in its history. The minus sign ensures that S has a *minimum* value if the world line connecting s_1 and s_2 is a straight line, corresponding to the actual motion of the particle along a straight world line in free space (A little reflection will show that the distance $\int_{s_1}^{s_2} ds$ is a maximum when the integration is along a straight world line. This contrasts with the case in real space where integration along a straight line gives the shortest distance in free space. The difference arises because of the imaginary time coordinate in ds).

A useful expression for S is obtained by introducing the observer's time interval dt via $ds = c\sqrt{1 - \beta^2} dt$ for ds in (6.35). Then

$$S_{\text{free}} = -mc^2 \int_{t_1}^{t_2} \sqrt{1 - \frac{v^2}{c^2}} dt, \quad (6.36)$$

which in turn identifies a possible Lagrange function for a free particle.

$$L_{\text{free}} = -mc^2 \sqrt{1 - \frac{v^2}{c^2}}. \quad (6.37)$$

For small velocities, $L_{\text{free}} \simeq -mc^2 + \frac{1}{2}mv^2$, which differs from the nonrelativistic Lagrangian $\frac{1}{2}mv^2$ by the constant $-mc^2$. However, we have seen that L is not unique. An equally suitable function for a free particle would be

$$L'_{\text{free}} = L_{\text{free}} + mc^2 = -mc^2 \left(\sqrt{1 - \frac{v^2}{c^2}} - 1 \right). \quad (6.38)$$

The momentum \mathbf{p} is given by

$$\mathbf{p} = \frac{\partial L_{\text{free}}}{\partial \mathbf{v}} = \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad (6.39)$$

irrespective of which form for L_{free} is used. However, the first form, (6.37) is preferable. This preference has to do with the expression for the energy of a free particle. From (1.78) of Chapter 1, the energy associated with the system described by L_{free} is

$$\mathcal{E} = \mathbf{p} \cdot \mathbf{v} - L_{\text{free}} = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (6.40)$$

Thus, \mathcal{E} reduces, not to zero, but to the so-called *rest mass energy* mc^2 when $\mathbf{v} = 0$. By contrast, L'_{free} describes a system with energy $E_k =$

$\mathcal{E} - mc^2$ measured relative to the rest mass energy. The quantity

$$E_k = \mathcal{E} - mc^2 = mc^2 \left[\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} - 1 \right] \quad (6.41)$$

is called the *kinetic energy*.

The fact that the energy \mathcal{E} is more fundamental than E_k emerges from its transformation properties under Lorentz transformations. The fact that \mathbf{p} , as given by (6.39) and \mathcal{E} together constitute the components of the following four-vector,

$$p_\mu = [\mathbf{p}, \frac{i}{c}\mathcal{E}] \quad (6.42)$$

called the *four-momentum*, or momentum-energy four-vector, means the \mathbf{p} and \mathcal{E} are connected with the corresponding values \mathbf{p}' and \mathcal{E}' in a moving frame Σ' via (6.25), i.e.

$$\begin{pmatrix} p_x \\ p_y \\ p_z \\ i\mathcal{E}/c \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & -i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ i\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} p'_x \\ p'_y \\ p'_z \\ i\mathcal{E}'/c \end{pmatrix}. \quad (6.43)$$

Since $\mathbf{p}' = 0$ and $\mathcal{E} = mc^2$ in the rest frame of the particle, the Lorentz-invariance of $\sum_\mu p_\mu^2$ provides a particularly convenient way of determining the energy-momentum relation for a free particle. The rest frame value is $\sum_\mu p_\mu'^2 = -(mc)^2$. Therefore by Lorentz invariance $\sum_\mu p_\mu^2 = -(mc)^2$, or

$$p^2 - \frac{\mathcal{E}^2}{c^2} = -(mc)^2, \quad (6.44)$$

where \mathbf{p} and \mathcal{E} refer to the observer's frame. Elimination of $(1 - \frac{v^2}{c^2})^{\frac{1}{2}}$ between (6.39) and (6.40) provides us with the particle velocity

$$\mathbf{v} = \mathbf{p} \frac{c^2}{\mathcal{E}} = \frac{\partial \mathcal{E}}{\partial \mathbf{p}}. \quad (6.45)$$

The energy-momentum relation (6.44) for a relativistic particle leads to a hyperbolic, rather than a parabolic relation between energy and momentum. We illustrate this in Fig. 6.2 after rewriting (6.44) as

$$\left(\frac{p}{mc}\right)^2 + \left(\frac{\mathcal{E}}{mc^2}\right)^2 = -1 \quad (6.46)$$

which is the equation for a hyperbola with semi-axes mc and mc^2 .

Since the gradient $\partial \mathcal{E} / \partial \mathbf{p}$ determines the particle velocity at each point according to (6.45), the asymptotes $\mathcal{E} = \pm cp$ in Fig. 6.2 describe the energy-momentum relation of "particles" moving with the speed of light. Such "particles" must necessarily have zero rest mass. Otherwise the energy of such particles would always be infinite according to (6.40).

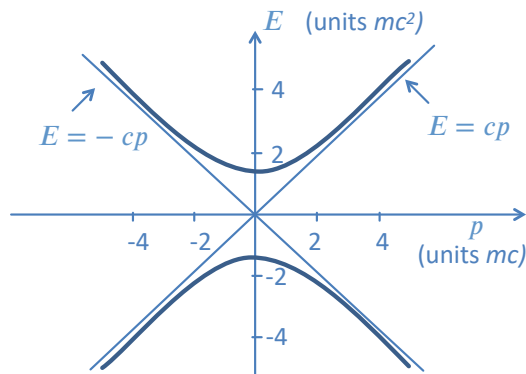


Figure 6.2: The energy-momentum relation for a relativistic free particle.

Photons, or "particles" or light are an example of entities satisfying the energy-momentum relation $\mathcal{E} = cp$. In fact, the four-vector character of $[\mathbf{p}, i\mathcal{E}/c]$, together with the photon picture of light $\mathcal{E} = cp$, where $\mathcal{E} = h\nu$, $h = \text{Planck's constant} = 6.63 \times 10^{-34}$ joule sec, is the energy carried by the photons constituting a light wave of frequency ν , leads to the formulae describing the Doppler shift and aberration of light in a natural way (see Problems.)

Returning to the relation (6.44) for material particles, we note that Fig. 6.2 provides visual assurances that the particle speed v cannot exceed c , since the gradient $\partial\mathcal{E}/\partial\mathbf{p}$ of the hyperbola is always *less* than the gradient of the relevant asymptote. The other aspect to notice about Fig. 6.2 is that the energy-momentum relation is two-valued, there being two possible values $\pm\mathcal{E}$ of the energy for each value of \mathbf{p} . These negative energy values, also shown in Fig. 6.2 are properly dismissed as unphysical in classical physics. However, all this changes in quantum physics, where it can be shown that the negative energy solutions do have a physical interpretation⁷³ in terms of a "sea" of occupied "antiparticle" states.

The inclusion of the rest mass energy mc^2 in \mathcal{E} is an essential feature of (6.44). This has the consequence that the separate conservation of mass and energy of classical physics is amalgamated into a single mass-energy conservation law expressed by (6.44). Mass has to be considered a form of energy that is not distinct from other forms of energy and is not conserved separately. Consider an atomic nucleus by way of a concrete example. Its mass is M say, taken as a whole, and so the nucleus has energy Mc^2 at rest. However, the constituent nucleons, A in number, possess energy by virtue of their kinetic motion and mutual interaction, in addition to their rest energies $m_i c^2$. Consequently the sum $\sum_{i=1}^A m_i c^2 \neq Mc^2$. A part of the rest energy Mc^2 of the nucleus resides in the nucleon interactions. In fact, the rest mass energy Mc^2 of a *stable* nucleus is always smaller than the sum of the masses of its constituent

⁷³ See for example, J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics*, (McGraw-Hill Book Company, New York and London, 1964), Chapter 5.

nucleons by an amount

$$-B = (Mc^2 - \sum_{i=1}^A m_i c^2), \quad (6.47)$$

where $B > 0$ is called the binding energy. An energy B has to be supplied to break up M into its constituents m_i . Conversely, if $B < 0$, the system is unstable and will disintegrate with emission of energy, usually in the form of kinetic energy of the decay products. The advent of nuclear power sources is made possible by the conversion of mass-energy into some other (usable) form of energy. In the fission of uranium nuclei by neutron bombardment for example, the mass energies of the uranium and its fission products differ by about 200 MeV. This amounts to an energy release of nearly 10^{14} joules/kg of fissile material!

The conversion of a part of the rest mass energy into other forms of energy during fission is certainly one graphic illustration of the validity of (6.40) for describing the energy of a free particle. The other aspect of this relation, i.e. how the energy of a relativistic free particle depends on its speed and the existence of a limiting speed, has been beautifully illustrated by examining the speed and energy gain of electrons after passing down the accelerating tube of a linear accelerator (W. Bertozzi, *Am. J. Phys.* **32**, 551 (1964)). The experiment measures the speed v and kinetic energy E_k *independently* that the electrons attain in a given accelerating potential. A plot of the square of the speed attained for a given kinetic energy is shown in Fig. 6.3.

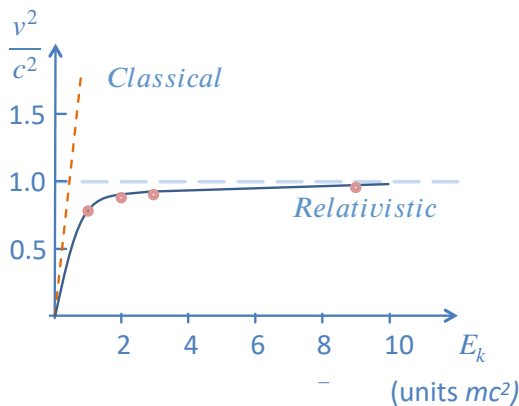


Figure 6.3: Experimental confirmation of the relativistic relation between the speed and the kinetic energy of relativistic electrons. The open circles denote the experimental points.

The deviation of the experimental points from the non-relativistic relation $v^2/c^2 = 2E_k/mc^2$ in favour of the relativistic prediction

$$\frac{v^2}{c^2} = 1 - \frac{1}{\left(1 + \frac{E_k}{mc^2}\right)^2}, \quad (6.48)$$

which follows from (6.41), gives a convincing illustration of the correctness of the latter over the former. The figure also suggests that no

appreciable increase in speed is brought about by increasing the energy supplied to the electron, once this energy exceeds the rest mass energy $mc^2 = 0.511$ MeV of the electron.

6-7 Particles and Fields

So far, we have avoided the question of how to introduce the idea of interactions into Special Relativity. We now consider this question, again from the point of view of constructing a suitable action function S for the interacting system. A prototype of the sort of problem one faces is the relativistic motion of a charge e in an electromagnetic field. This problem is relatively simple because of the known fact that the vector and scalar potentials \mathbf{A} and ϕ describing the electromagnetic field in free space form the components of a four-vector

$$A_\mu = [\mathbf{A}, i\frac{\phi}{c}], \quad (6.49)$$

called the *four-potential* of the field. This result is a consequence of the *automatic* invariance of Maxwell's equations under Lorentz transformations. It is thus a simple matter to construct a Lorentz-invariant expression representing the particle-field interaction term in S . We write

$$S = S_{\text{free}} + S_{\text{int}}, \quad (6.50)$$

where S_{free} refers to a free particle. As for S_{int} , describing the particle-field interaction, we know that this term must (i) be a Lorentz-invariant, and (ii) contain quantities that refer to both the particle as well as field variables in order to describe a particle-field interaction. The simplest expression meeting these requirements is

$$S_{\text{int}} = e \int_{x_1}^{x_2} \sum_{\mu} A_{\mu} dx_{\mu}, \quad (6.51)$$

where $dx_{\mu} = v_{\mu}d\tau$ is the four-displacement in the proper time interval $d\tau$ and A_{μ} is evaluated at the particle-position $x_{\mu} = x_{\mu}(\tau)$. The electric charge e , which is a Lorentz-invariant property of the particle, serves as the "coupling constant" between particle and field. Multiplying S_{int} by additional constant amounts to setting the units of charge and electromagnetic field. The choice (6.51) is equivalent to using MKS units throughout. Writing $dx = (v_{\mu}/c) ds$ and adding the resulting expression for S_{int} to S_{free} , one has

$$S = \int_{s_1}^{s_2} [-mc + \frac{e}{c} \sum_{\mu} v_{\mu} A_{\mu}] ds. \quad (6.52)$$

Introducing the observer's time interval dt in place of ds again, this becomes

$$S = \int_{t_1}^{t_2} [-mc^2 \sqrt{1 - \frac{v^2}{c^2}} - e\phi + e(\mathbf{v} \cdot \mathbf{A})] dt, \quad (6.53)$$

since $\sum_{\mu} A_{\mu} ds = (-\phi + \mathbf{v} \cdot \mathbf{A}) dt$. The Lagrange function for a charged particle in an electromagnetic field can now be read off as

$$L = -mc^2 \sqrt{1 - \frac{v^2}{c^2}} - e\phi + e(\mathbf{v} \cdot \mathbf{A}). \quad (6.54)$$

This expression only differs from its non-relativistic analogue (2.73) of Chapter 2, in the term describing the free particle. Since the electromagnetic interaction piece is already Lorentz-invariant (although we had no way of ascertaining this in Chapter 2), it comes through unchanged in form. We can now "turn the crank" once more to obtain the momentum and energy,

$$\mathbf{p} = \frac{\partial L}{\partial \mathbf{v}} = \frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}} + e\mathbf{A} \quad (6.55)$$

and

$$\mathcal{E} = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} + e\phi, \quad (6.56)$$

for a charged particle in an electromagnetic field.

Since the contribution of the interaction terms with the electromagnetic field are the same in these results as in the non-relativistic case of Chapter 2, the equation of motion of the charge is still given by (2.77), but with the mechanical momentum modified by the factor $(1 - v^2/c^2)^{1/2}$. Hence,

$$\frac{d}{dt} \left(\frac{m\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}} \right) = e\mathbf{E} + e(\mathbf{v} \times \mathbf{B}) \quad (6.57)$$

replaces the classical equation of motion for a charge in an electromagnetic field ($\mathbf{E} \cdot \mathbf{B}$).

To find the energy-momentum relation in the presence of an electromagnetic field, we observe that $[\mathbf{p} - e\mathbf{A}, \frac{i}{c}(\mathcal{E} - e\phi)]$ is a four-vector. Lorentz invariance of the norm of this four-vector yields

$$(\mathbf{p} - e\mathbf{A})^2 - \frac{1}{c^2}(\mathcal{E} - e\phi)^2 = -mc^2. \quad (6.58)$$

6-8 A Lorentz-Invariant Lagrangian

Although the action integrals in (6.51) and (6.52) are Lorentz-invariant, the Lagrange functions they lead to are not. Other formulations of the relativistic particle problem exist which use Lagrange functions that are manifestly Lorentz-invariant. By analogy with the arguments of the non-relativistic case, consider an action function given by the integral

$$S = \int_{\tau_1}^{\tau_2} \mathcal{L}[x_{\mu}(\tau), v_{\mu}(\tau)] d\tau \quad (6.59)$$

over the proper time interval $\tau_2 - \tau_1$ associated with the particle. The function \mathcal{L} depends on the position and velocity four-vectors $x_{\mu}(\tau)$ and

$v_\mu(\tau)$. Note that the four-velocity components v_μ are not independent, since $\sum_\mu v_\mu^2 = -c^2$. Introducing a suitable Lagrange-multiplier to take care of this, one finds that the variation problem $\delta S = 0$ with $\delta x_\mu(\tau)$ vanishing at the end points is satisfied, provided that

$$\frac{d}{d\tau} \left(\frac{\partial \mathcal{L}}{\partial v_\mu} \right) = \frac{\partial \mathcal{L}}{\partial x_\mu}, \quad \mu = 1, 2, 3, 4, \quad (6.60)$$

a by now familiar equation. A suitable Lagrange function of this type for a particle in an electromagnetic field is

$$\mathcal{L} = \frac{1}{2} \sum_\mu m v_\mu^2 + e \sum_\mu v_\mu A_\mu. \quad (6.61)$$

Notice that the "kinetic" term $\sum m v_\mu^2 = -mc^2$ is actually a constant. However, as stressed in Chapter 1, it is the *functional form* of \mathcal{L} that matters. The required manipulations on the function \mathcal{L} are simple. One has

$$\frac{d}{d\tau} (m v_\mu + e A_\mu) = e \sum_\nu \frac{\partial A_\nu}{\partial x_\mu} v_\nu, \quad (6.62)$$

a result which may be written as

$$\frac{d}{d\tau} (m v_\mu) = e \sum_\nu F_{\mu\nu} v_\nu \quad (6.63)$$

if we transpose $dA_\mu/d\tau = \sum_\nu (\partial A_\mu / \partial x_\nu) v_\nu$ to the right hand side and introduce the abbreviation

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x_\mu} - \frac{\partial A_\mu}{\partial x_\nu} = -F_{\nu\mu}. \quad (6.64)$$

The entity $F_{\mu\nu}$ is by construction an antisymmetric tensor of rank two and therefore has six independent entries. For this reason $F_{\mu\nu}$ often goes by the name *electromagnetic six-vector*, which terminology becomes clear on writing out its components in terms of the components of the electromagnetic fields \mathbf{E} and \mathbf{B} ,

$$\mathbf{E} = -\text{grad } \phi - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \text{curl } \mathbf{A}. \quad (6.65)$$

Then,

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & B_z & -B_y & -\frac{i}{c} E_x \\ -B_z & 0 & B_x & -\frac{i}{c} E_y \\ B_y & -B_x & 0 & -\frac{i}{c} E_z \\ \frac{i}{c} E_x & \frac{i}{c} E_y & \frac{i}{c} E_z & 0 \end{pmatrix}. \quad (6.66)$$

The equation of motion given by (6.63) is manifestly invariant in form (or as we say, covariant) under Lorentz transformations. This circumstance is brought about because it is written explicitly in terms of four-vectors and/or derivatives thereof with respect to proper time that is a Lorentz-invariant concept. In terms of ordinary three-vectors and derivatives

with respect to the observer's time, the equation of motion splits into two parts. The space part of (6.63) just duplicates (6.57). For the fourth, or "time" component, one finds

$$\frac{d}{dt} \left(\frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} \right) = e\mathbf{E} \cdot \mathbf{v}, \quad (6.67)$$

expressing the fact that the increase in energy in time dt is supplied by the work done by the electric field on the charge in this interval. The covariant form of the equation of motion for the charge thus combines Newton's law of motion plus the rate of working of the applied forces (a consequence of Newton's law of motion) into a single, compact equation.

The results given in (6.57) and (6.67) suggest a way of introducing the idea of "force" into Special Relativity. When written in terms of the proper time of the particle, these equations become

$$\frac{d\mathbf{p}}{d\tau} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \mathbf{F}^{(l)} \quad (6.68)$$

and

$$\frac{d}{d\tau} \left(\frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} \right) = \frac{\mathbf{v} \cdot \mathbf{F}^{(l)}}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (6.69)$$

in terms of the Lorentz force $\mathbf{F}^{(l)} = e\mathbf{E} + e(\mathbf{v} \times \mathbf{B})$ acting on the charge. Multiplying the lower equation by i/c shows that the set (6.68) and (6.69) is equivalent to the single equation

$$\frac{d\mathbf{p}}{d\tau} = f_\mu, \quad \mu = 1, 2, 3, 4, \quad (6.70)$$

where the *four-force* f_μ is given by

$$f_\mu = \left[\frac{\mathbf{F}}{\sqrt{1 - \frac{v^2}{c^2}}}, \frac{i}{c} \frac{\mathbf{v} \cdot \mathbf{F}}{\sqrt{1 - \frac{v^2}{c^2}}} \right]. \quad (6.71)$$

Here F (now not necessarily $F^{(l)}$) is the force on the particle as it would appear in the non-relativistic equation of motion. Alternatively, one simply can *define* f_μ by (6.70). In either case, the four-vector nature of f_μ guarantees the invariance in form, covariance of (6.70) under Lorentz transformations. Note that $\sum_\mu v_\mu f_\mu = 0$, i.e. the four-force is always perpendicular to the four-velocity, in view of (6.30).

The form and transformation properties of (6.70) suggest that it is a suitable generalization of Newton's second law of motion for relativistic particles. The definition of "force" embodied in this equation also synthesises the two fundamental conservation laws of classical physics, those of energy and momentum, into a single law of the conservation of four-momentum when the four-force is zero. That is, if $f_\mu = 0$, then

$$p_\mu = \left[\mathbf{p}, \frac{i}{c} \mathcal{E} \right] = \text{constant}, \quad (6.72)$$

which can only hold if both the space and time components \mathbf{p} and \mathcal{E} of p_μ are constant.

The special theory of relativity, providing as it does a fertile field for new insights and gross misconceptions, remains a somewhat aloof, if essential element of classical physics. For this reason, we have attempted to keep the presentation on a reasonably elementary level and close to the basic principles of relativity. The most natural field of application for most of the material in this chapter concerns the motion of charged particles in given electromagnetic fields, the generation of such fields by charges in arbitrary motion, and finally the "feedback" of the field generated by a charge in motion, on that motion. Such problems require a simultaneous study, which is not attempted here, of the equations of motion governing both particle motion and field motion in the presence of particle-field coupling terms⁷⁴. Neither have we made any mention (beyond the name!) of the principles underlying the General Theory of Relativity⁷⁵.

⁷⁴ See for example J.D. Jackson, *Classical Electrodynamics* (John Wiley and Sons, Inc., New York, 1962), Chapter 17.

⁷⁵ See for example, R. Adler, M. Bazin and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill Book Company, New York and London, 1965).

Problems

6-1. Some stars, called white dwarfs, have nearly the mass of the sun, but the size of the earth. Assume that (6.22) describes the frequency shift experienced by light emitted from a white dwarf and calculate a value for $\Delta\nu/\nu_0$. How do Doppler shifts, as per (6.20) enter into the picture? Can you suggest a procedure for untangling the two effects?

6-2. Doppler shift. A light source of frequency ν_0 moves towards an observer with speed $V = \beta c$. Show that the formula (6.20) for the frequency ν_D that is observed follows from the transformation properties of the associated energy $\mathcal{E}' = h\nu_0$ and momentum $p' = \mathcal{E}'/c$ of the photons emitted by the source.

6-3. The photons referred to in problem 6-2 have to be emitted by a light source, e.g. an excited atom or nucleus. Examine, via the conservation of four-momentum how the recoil of the emitting source influences the previous result for ν_D .

6-4. Aberration of light. A light source, moving with velocity $V = \beta c$ along the x -axis of a stationary reference frame, emits photons in a direction θ' with respect to its direction of travel, as observed in the rest frame of the source. Show that, according to an observer at rest, these photons are emitted in a direction θ with respect to the direction of travel, where

$$\cos \theta = \frac{\cos \theta' + \beta}{1 + \beta \cos \theta'}. \quad (6.73)$$

This phenomenon is known as the *aberration of light*.

6-5. Consider reference frames Σ , Σ' and Σ'' in relative motion along a common x -axis, such that Σ'' has a velocity $V'' = \beta''c$ relative to Σ' , which in turn moves with velocity $V' = \beta'c$ relative to Σ . Prove that the velocity of Σ'' as seen by Σ is $V = \beta c$, where

$$\beta = \frac{\beta' + \beta''}{1 + \beta'\beta''} \quad (6.74)$$

by considering the successive Lorentz transformation $\Sigma \rightarrow \Sigma' + \Sigma''$ as constituting the single transformation $\Sigma \rightarrow \Sigma''$. Hint: the transformations $\Sigma \rightarrow \Sigma'$ and $\Sigma' \rightarrow \Sigma''$ are "rotations" through angles ϕ' and ϕ'' , where $\tan \phi' = i\beta'$ etc., that are equivalent to the single "rotation" $\phi = \phi' + \phi''$. Set $\tan \phi = i\beta$ and use the addition theorem for the tangent function.

6-6. Photons of sufficiently high energy can lose energy in matter by the process of pair production, where an electron-positron pair is created. Show that this process cannot take place without the presence of a third particle (an atomic nucleus) to take care of momentum and energy conservation.

6-7. There has been some speculation on the literature (see e.g. O.M. Bilaniuk und E.C.G. Sudershan, *Physics Today* **22**, 43 (1966) for an elementary exposition) regarding the possible existence and properties of particles ("tachyons") that always move with a velocity greater than that of light. Accepting this hypothesis, show that the every-velocity relation for tachyons implies that the rest mass parameter m in (6.40) must be replaced by an imaginary parameter im . Find the corresponding energy-momentum and energy-momentum-velocity relations for tachyons. From these show that the faster a tachyon goes, the *lower* is its energy.

6-8. The fact that interactions propagate with a finite speed implies that the concept of a "rigid" body has to be examined with care from the point of view of Relativity. A paradox that illustrates the sort of problems one runs into is the following (E.M. Dewan, *Am. J. Phys.* **31**, 342 (1963)): A polevaulter runs into a barn with his pole held horizontally. The rest length of the pole is such that it will not fit into the barn, but according to an observer in the barn, its Lorentz contracted length will fit in. Once the polevaulter is inside the barn, the barn door is slammed shut. Can the polevaulter "explain" the fact that the door can be shut behind him, since according to him, it is the barn that has contracted so that the pole cannot possibly fit into it?

Chapter 7 Hamiltonian Mechanics

7.1 Hamilton's Equations

We saw in Chapter 1 that the Lagrange formulation of dynamics enjoyed many advantages in terms of the freedom of choice of coordinates, the ready recognition of conservation laws, and finally in the presentation of the laws of dynamics themselves in terms of the action principle

$$\delta \int_{t_1}^{t_2} L(q, \dot{q}, t) = 0. \quad (7.1)$$

In this Chapter, we will study yet another reformulation of the equations of motion of a system, whereby the n Lagrange equations

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) = \frac{\partial L}{\partial q_k}, \quad k = 1, 2, \dots, n, \quad (7.2)$$

for a system having n degrees of freedom, and which are second-order differential equations in the time variable t , are replaced by $2n$ first-order differential equations, known as *Hamilton's equations of motion*. The simplest way of accomplishing this end is to introduce n auxiliary variables r_k , where

$$\dot{q}_k = r_k, \quad k = 1, 2, \dots, n. \quad (7.3)$$

Then the Lagrange equations (7.2) revert trivially to a set of n first order differential equations

$$\frac{d}{dt} \left(\frac{\partial L}{\partial r_k} \right) = \frac{\partial L}{\partial q_k}, \quad (7.4)$$

which, together with (7.3), make up a set of $2n$ first-order differential equations that are equivalent to the n original Lagrange equations (7.2) that are of second order. However, instead of using the r_k , it proves much more convenient to use the canonical momenta

$$p_k = \frac{\partial L}{\partial \dot{q}_k} = \frac{\partial L}{\partial r_k}, \quad (7.5)$$

defined in (1.72) of Chapter 1. Assuming that this relation can be turned "inside out" to solve for the r_k in terms of q_k and p_k , we can pass from the set of $2n$ variables (q_k, r_k) to the set (q_k, p_k) .

Let us retrace the steps leading from the extremum condition (7.1) to the Lagrange equations, but where L is now considered as a function of q_k , and r_k as per (7.3), $L = L(q, r, t)$. Therefore the variation indicated in (7.1) has now to be carried out under the restriction that (7.3) must hold, or that

$$\delta\dot{q}_k - \delta r_k = 0, \quad (7.6)$$

in terms of the variations in $\delta\dot{q}_k = (d/dt)\delta q_k$ and δr_k . Therefore, the $2n$ variables q_k and r_k are not independent. But from the discussion of constraints in Section 1-6 of Chapter 1, we are already prepared to deal with this complication by introducing Lagrange multipliers λ_k , where

$$\sum_k \lambda_k (\dot{q}_k - r_k) = 0. \quad (7.7)$$

Adding the variation of this equation, i.e. the identity

$$\delta \sum_k \lambda_k (\dot{q}_k - r_k) = 0 \quad (7.8)$$

to L in (7.1), we find that

$$\delta \int_{t_1}^{t_2} [L + \sum_k \lambda_k (\dot{q}_k - r_k)] dt = 0, \quad (7.9)$$

where, as previously, $\delta\dot{q}_k = \frac{d}{dt}(\delta q_k)$. The variations δq_k and δr_k can now be considered to be independent, and the coefficients of each δq_k and δr_k must vanish separately. In particular, the coefficients of the δr_k determine the λ_k according to

$$\frac{\partial L}{\partial r_k} = \lambda_k. \quad (7.10)$$

Returning to (7.9) with this information, we discover that the extremum condition is equivalent to

$$\delta \int_{t_1}^{t_2} [L + \sum_k \frac{\partial L}{\partial r_k} (\dot{q}_k - r_k)] dt = 0, \quad (7.11)$$

or

$$\delta \int_{t_1}^{t_2} [L - \sum_k p_k \dot{q}_k + \sum_k p_k \dot{q}_k] dt = 0, \quad (7.12)$$

if we make use of $r_k = \dot{q}_k$ and the definition of p_k . The last relation may be written as

$$\delta \int_{t_1}^{t_2} [-H + \sum_k p_k \dot{q}_k] dt = 0 \quad (7.13)$$

in terms of the *Hamilton function*, or *Hamiltonian*

$$H = -L + \sum_k p_k \dot{q}_k. \quad (7.14)$$

We have met this expression previously in (1.78) of Chapter 1, there as a result of investigating the conservation properties of a system. We see

from (7.13) that the principle of least action can be expressed in terms of H as well, so that H is indeed as fundamental as the Lagrange function for describing the dynamics of a system. We note that by agreement L was to be considered a function of q_k and r_k or equivalently q_k and p_k (courtesy of (7.5) in the above derivation). Consequently the same holds true for H : The Hamilton function is to be considered a function of the $2n$ canonical variables (q_k, p_k) , i.e.

$$H = H(q, p, t) \quad (7.15)$$

using the shorthand notation $q = (q_1, q_2, \dots, q_n)$, etc.

With this proviso in mind we can use the action principle in the form (7.13) to find the equations of motion for the variables q_k and p_k . Carrying out the indicated variations, one sees that

$$\int_{t_1}^{t_2} \sum_k \left[\left(-\frac{\partial H}{\partial q_k} - \dot{p}_k \right) \delta q_k + \left(-\frac{\partial H}{\partial p_k} + \dot{q}_k \right) \delta p_k \right] dt = 0, \quad (7.16)$$

after employing a partial integration to shift the time derivative of $\delta \dot{q}_k = (d/dt)\delta q_k$ onto its cofactor p_k , and the boundary conditions $\delta q_k(t_1) = \delta q_k(t_2) = 0$. Since the variations δq_k and δp_k are both arbitrary and independent, we conclude that

$$\dot{p}_k = -\frac{\partial H}{\partial q_k}; \quad \dot{q}_k = \frac{\partial H}{\partial p_k}, \quad k = 1, 2, \dots, n. \quad (7.17)$$

This set of $2n$ first order differential equations constitute Hamilton's equations of motion for a system with n degrees of freedom. Note in passing that if a coordinate q_s is absent (cyclic) in L , it is also absent in H . Hence from the first equation,

$$\dot{p}_s = -\frac{\partial H}{\partial q_s} = 0, \quad \text{or} \quad p_s = \text{constant}, \quad (7.18)$$

or that each momentum that is canonical to a cyclic coordinate is conserved, as in (1.75) of Chapter 1. The Hamilton function H is itself conserved if it does not depend explicitly on t . We see this by calculating dH/dt ,

$$\frac{dH}{dt} = \sum_k \left(\frac{\partial H}{\partial q_k} \dot{q}_k + \frac{\partial H}{\partial p_k} \dot{p}_k \right) + \frac{\partial H}{\partial t} = \frac{\partial H}{\partial t} \quad (7.19)$$

and using Hamilton's equations. Hence, if

$$\frac{\partial H}{\partial t} = 0, \quad \text{then} \quad H = \text{constant}. \quad (7.20)$$

H may or may not represent the total energy, depending on the functional form of T and V in the Lagrange function, see discussion leading to (1.81) of Chapter 1: If L has the simple form $L = T - V$, with T being a homogeneous quadratic function of the q_k , and V only a function of the q_k , then

$$H = T + V = E \quad (7.21)$$

is constant and equal to the total energy E of the system.

7-2 Some Examples of Hamilton's Equations

Experience has shown that it is usually as easy or as difficult to solve Hamilton's equations of motion as it is to solve Lagrange's equations for a given physical problem. The utility of Hamilton's equations actually lies in a different direction as will become clear presently. For the moment we content ourselves with a few examples to gain familiarity with Hamilton's method.

(i) *One-dimensional harmonic oscillator*: Calling the displacement of the particle away from equilibrium q instead of x , one has

$$L = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}m\omega_0^2q^2 \quad (7.22)$$

if the mass m oscillates with frequency ω_0 . The Hamilton function (7.14) becomes

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega_0^2q^2 \quad (7.23)$$

in this case, since

$$p = \frac{\partial L}{\partial \dot{q}} = m\dot{q} \quad (7.24)$$

can be used to eliminate \dot{q} in favour of p . Hamilton's equations of motion for an oscillator are thus

$$\dot{p} = -m\omega_0^2q; \quad \dot{q} = \frac{p}{m}. \quad (7.25)$$

The second of these just reconfirms the momentum-velocity relation (7.24). If we use this relation to eliminate p from the first equation, then

$$\ddot{q} + \omega_0^2q = 0, \quad (7.26)$$

which is just the equation of motion for an oscillator.

(ii) *Central Motion*

Consider again the motion of a particle, mass m in a central potential field $V(r)$. In plane polar coordinates (r, θ) the Lagrange function is

$$L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - V(r), \quad (7.27)$$

leading to the canonical momenta

$$p_r = \frac{\partial L}{\partial \dot{r}} = m\dot{r}; \quad p_\theta = \frac{\partial L}{\partial \dot{\theta}} = mr^2\dot{\theta}. \quad (7.28)$$

The Hamiltonian, which is also the total energy in this and the previous problem, reads

$$H = \frac{1}{2m}(p_r^2 + \frac{p_\theta^2}{r^2}) + V(r), \quad (7.29)$$

and the equations of motion are

$$\dot{p}_r = \frac{p_\theta^2}{mr^3} - \frac{\partial V}{\partial r}; \quad \dot{r} = \frac{p_r}{m} \quad (7.30)$$

and

$$\dot{p}_\theta = 0; \quad \dot{\theta} = \frac{p_\theta}{mr^2}. \quad (7.31)$$

The last set of relations confirms that the angular momentum $p_\theta = mr^2\dot{\theta}$ is conserved, while p_r and p_θ can be eliminated in the first partner of (7.30) to give the usual radial equation of motion

$$m(\ddot{r} - r\dot{\theta}^2) = F(r) \quad (7.32)$$

in a force-field $F(r) = -\partial V/\partial r$.

(iii) *Motion of a Charged Particle:*

The Lagrange function for this problem has been given before in (2.73) of Chapter 2, i.e.

$$L = \frac{1}{2}mv^2 - e\phi + e(\mathbf{v} \cdot \mathbf{A}), \quad (7.33)$$

where m is the mass of the charge e moving with velocity \mathbf{v} . If we locate e in cartesian coordinates at $\mathbf{r} = (x, y, z)$, then the canonical momentum $\mathbf{p} = (p_x, p_y, p_z)$ is given as

$$\mathbf{p} = m\mathbf{v} + e\mathbf{A}, \quad (7.34)$$

so that

$$H = \frac{1}{2m}(\mathbf{p} - e\mathbf{A})^2 + e\phi. \quad (7.35)$$

A straight-forward calculation, using

$$\dot{\mathbf{p}} = -\text{grad}_r H; \quad \mathbf{v} = \dot{\mathbf{r}} = \text{grad}_p H \quad (7.36)$$

leads to the equation of motion established previously, i.e.

$$\frac{d}{dt}(m\mathbf{v}) = e\mathbf{E} + e(\mathbf{v} \times \mathbf{B}). \quad (7.37)$$

7-3 Canonical Transformations

We saw in Chapter 1 how a reformulation of Newton's second law of motion as Lagrange's equations led to a particularly efficient way of looking at dynamical problems. But the real bonus was the freedom we gained in the choice of coordinates that could be employed. This freedom was guaranteed in the derivation of the Lagrange equations from an action principle, and showed up explicitly in their invariance in form under *point transformations* of the type

$$q_k = f_k(q'_1, q'_2, \dots, q'_n; t) \quad (7.38)$$

from one set of n coordinates q_k to another set q'_k . The particular advantage of Hamilton's formulation of the equations of motion lies in the fact that, in addition to retaining their form under point transformations like (7.38), they admit a much wider class of transformations for which this

invariance also holds. This class of transformations, called *canonical* or *contact transformations* provides for the transformation of the old coordinates *and* momenta to new coordinates and momenta Q_k and P_k , via relations of the form

$$\begin{aligned} q_k &= f_k(Q_1, \dots, Q_n; P_1, \dots, P_n; t) \\ p_k &= g_k(Q_1, \dots, Q_n; P_1, \dots, P_n; t) \end{aligned} \quad (7.39)$$

such that the new canonical variables also satisfy Hamilton's equations

$$\dot{P}_k = -\frac{\partial K}{\partial Q_k}; \quad \dot{Q}_k = \frac{\partial K}{\partial P_k} \quad (7.40)$$

in terms of a *new* Hamilton function $K = K(Q, P; t)$. The construction of transformations like (7.39) that turn Hamilton's equations for the old variables into mirror relationships for the new variables seems like a tall order. Actually the problem is quite simple. The hint as to how one should view the problem comes from the variational principle in the form (7.13). There we indicated how Hamilton's equations followed from this principle. Conversely, one can show that if Hamilton's equations are valid, then (7.13) must necessarily hold. With this in mind it follows therefore that if (7.40) are to hold, K must satisfy the following variational principle

$$\delta \int_{t_1}^{t_2} [-K + \sum_k P_k \dot{Q}_k] dt = 0, \quad (7.41)$$

with $\delta Q(t_1) = \delta Q(t_2) = 0$. The discovery of the class of transformations that are canonical therefore boils down to the problem of introducing transformations which *automatically* guarantee that (7.41) will hold, provided that (7.13), i.e.

$$\delta \int_{t_1}^{t_2} [-H + \sum_k p_k \dot{q}_k] dt = 0 \quad (7.42)$$

holds in the old variables, with the familiar boundary conditions $\delta q_k(t_1) = \delta q_k(t_2) = 0$. That (7.41) should be an automatic consequence of (7.42) does not mean that their integrands are equal, however. For we can add the total time derivative dG/dt of any arbitrary function G to the integrand of either equation without changing the value of the variation. This fact follows from the nature of the δ -variation. Let $G = G(q, Q, t)$ be some function of the old and new coordinates. Then

$$\int_{t_1}^{t_2} \frac{dG}{dt} dt = [\delta G]_{t_1}^{t_2} = \sum_n \left[\frac{\partial G}{\partial q_k} \delta q_k + \frac{\partial G}{\partial Q_k} \delta Q_k \right]_{t_1}^{t_2} = 0, \quad (7.43)$$

due to the boundary conditions on the δq_k and δQ_k at t_1 and t_2 . Consequently, (7.41) will be a consequence of (7.42) provided that

$$-H(q, p, t) + \sum_k p_k \dot{q}_k = -K(Q, P, t) + \sum_k P_k \dot{Q}_k + \frac{dG}{dt}(q, Q, t). \quad (7.44)$$

The function G is called the *generating function* of the canonical transformation. Canonical transformations are thus characterized by specifying G instead of the relations (7.39). However, these relations tell us that G can only depend on any two of the sets of variables (q_k, p_k) and (Q_k, P_k) , and possibly the time t . It proves to be convenient to specify G as depending on one of the following four sets of variables that are taken to be independent: (i) (q_k, Q_k) , or (ii) (q_k, P_k) , or (iii) (p_k, Q_k) , or (iv) (p_k, P_k) , so that G has "one foot in each camp" so as to speak. The missing variables in each case are then to be calculated in terms of any of these sets from a knowledge of the generating function itself. To see how this works out in practice, let us look at Case (i) $G = G(q, Q, t)$. Calculating the time derivative in (7.44) and equating coefficients of q_k and Q_k , one finds the set of equations

$$p_k = \frac{\partial G}{\partial q_k} \quad (7.45)$$

$$P_k = -\frac{\partial G}{\partial Q_k} \quad (7.46)$$

$$K = H + \frac{\partial G}{\partial t} \quad (7.47)$$

for each value of k . Equations (7.45) and (7.46) give $p_k = g'_k(q, Q, t)$ and $P_k = f'_k(q, Q, t)$. If we can "solve" the latter equation in the form $q_k = f_k(Q, P, t)$ and use this information to eliminate the q_k from the former, $p_k = g'_k(f_k(Q, P, t), Q, t) = g_k(Q, P, t)$, we have rendered the canonical transformation $(q_k, p_k) \rightarrow (Q_k, P_k)$ in the form (7.39).

Case (ii): Since q_k and P_k are considered independent, we must eliminate the $\sum_k P_k \dot{Q}_k$ in favour of $\sum_k \dot{P}_k Q_k$ in (7.44). This can be accomplished by writing

$$G(q, Q, t) = S(q, P, t) - \sum_k P_k Q_k, \quad (7.48)$$

so that

$$\frac{dG}{dt} = \sum_k \frac{\partial S}{\partial q_k} \dot{q}_k + \sum_k \frac{\partial S}{\partial P_k} \dot{P}_k + \frac{\partial S}{\partial t} - \sum_k \dot{P}_k Q_k - \sum_k P_k \dot{Q}_k. \quad (7.49)$$

Using this information to (7.44) once more, leads to another set of relations in terms of the function S ,

$$p_k = \frac{\partial S}{\partial q_k} \quad (7.50)$$

$$Q_k = \frac{\partial S}{\partial P_k} \quad (7.51)$$

$$K = H + \frac{\partial S}{\partial t}. \quad (7.52)$$

Once more, the first two equations can be solved in principle to provide the (q_k, p_k) as functions of the (Q_k, P_k) .

The relation (7.48) is an example of a *Legendre transformation* that is so useful in shifting to new independent variables in thermodynamics. Cases (iii) and (iv) above can be dealt with in like manner via the Legendre transformations $G = T(p, Q, t) + \sum_k p_k q_k$ for (iii) or $G = U(p, P, t) + \sum(p_k q_k - P_k Q_k)$ for (iv) respectively. The dependent variables are given by

$$q_k = -\frac{\partial T}{\partial p_k} \quad (7.53)$$

$$P_k = -\frac{\partial T}{\partial Q_k} \quad (7.54)$$

$$K = H + \frac{\partial T}{\partial t} \quad (7.55)$$

in case (iii), $T = T(p, Q, t)$, and

$$q_k = -\frac{\partial U}{\partial p_k} \quad (7.56)$$

$$Q_k = \frac{\partial U}{\partial P_k} \quad (7.57)$$

$$K = H + \frac{\partial U}{\partial t} \quad (7.58)$$

in case (iv), $U = U(p, P, t)$.

Some examples of familiar transformations "in canonical clothing" are useful to bear in mind. The identity transformation, $q_k = Q_k$, and $p_k = P_k$ is obviously generated by setting $G = 0$. In terms of a type (ii) or type (iii) generating function, this means that either of

$$S = \sum_k q_k P_k, \quad \text{or} \quad T = \sum_k p_k Q_k \quad (7.59)$$

generate an identity transformation, as may be verified by direct calculation. The "interchange of names" transformation $q_k = -P_k$ and $p_k = Q_k$ is obviously canonical. It is generated by either of

$$G = \sum_k q_k Q_k, \quad \text{or} \quad U = \sum_k p_k P_k. \quad (7.60)$$

These last two examples show rather clearly that which variable is termed a "coordinate" and which a "momentum" is a matter of semantics in Hamiltonian mechanics, and the term "canonical variables" for each pair (q_k, p_k) is preferable. However, it is well to bear in mind that this semantic freedom has to be tempered by the realization that the physical significance of the canonical variables will eventually have to be identified in making actual calculations.

The next transformation in ascending order of complexity, is the point transformation between coordinates,

$$q_k = f_k(Q_1, Q_2, \dots, Q_n; t), \quad \text{or} \quad Q_k = F_k(q_1, q_2, \dots, q_n; t). \quad (7.61)$$

The first one is generated directly by

$$T = - \sum_l f_l(Q_1, Q_2, \dots, Q_n; t) p_l, \quad (7.62)$$

or indirectly by

$$S = \sum_l F_l(q_1, q_2, \dots, q_n; t) P_l, \quad (7.63)$$

which generating function also gives the second form of the point transformation directly. Writing q and Q for the set of variables in f and F , one has

$$\begin{aligned} q_k &= f_k(Q, t) \\ P_k &= \sum_l \frac{\partial f_l}{\partial Q_k} p_l \end{aligned} \quad (7.64)$$

or

$$\begin{aligned} Q_k &= F_k(q, t) \\ p_k &= \sum_l \frac{\partial F_l}{\partial q_k} P_l \end{aligned} \quad (7.65)$$

upon using (7.53)-(7.55) or (7.50)-(??). One can show directly that each of these transformations is canonical (see Problems). In a similar fashion the choices $S = \sum_l q_l g_l(P, t)$, or $T = - \sum_l G_l(p, t) Q_l$ generate "point" transformations between momenta, $p_k = g_k(P, t)$ or $P_k = G_k(p, t)$, while $G = \sum_l q_l W_l(Q, t)$ and $U = - \sum_l p_l W_l(P, t)$ would produce "mixed" point transformations like $p_k = W_k(Q, t)$, or $q_k = W_k(P, t)$ respectively. None of these results are really surprising when one remembers that the q 's and p 's are always treated on an equal footing in Hamilton theory. Instead of arbitrarily constructing more complicated canonical transformations at will, we rather ask at this stage whether there are perhaps canonical transformations that have a particular significance in dynamics. This question is examined in the following pages.

7-4 Special Canonical Transformations and Hamilton-Jacobi Theory

Our considerations thus far have ignored the fact that the Hamilton function itself is altered to

$$K = H + \frac{\partial G}{\partial t} \quad (7.66)$$

under a canonical transformation of variables. We have also seen that G may be replaced by any one of the functions S , T or U . Since K is to be a function of the new variables (Q_k, P_k) , a knowledge of the associated canonical transformation in the guise of (7.39) is a prerequisite in order to eliminate the old canonical variables on the right hand side of (7.66)

in favour of the new. However, there is another way of looking at (7.66): It tells us that we may "fiddle" with the functional dependence and/or value of the new Hamiltonian at will by changing the function G . Therefore, the idea presents itself of looking for those generating functions (and thus canonical transformations) that render the new Hamilton function simpler than the old one, when looked at from the point of view of solving Hamilton's equations. Thus (7.66) is now to be regarded as a condition determining G (or whatever alternate generating function is considered), rather than a prescription for calculating K .

The question as to what is the "simplest" problem to solve in terms of the new canonical variables, opens up a host of possibilities. Perhaps the simplest imaginable dynamical problem is no problem at all. $K = 0$! By (7.40) this would mean that all the new canonical variables are constant in time, $\dot{P}_k = 0$ and $\dot{Q}_k = 0$ for all k , or

$$P_k = \alpha_k, \quad Q_k = \beta_k, \quad (7.67)$$

where α_k and β_k are constants. The generating function that accomplishes this transformation is determined by (7.66) with K set equal to zero, i.e.

$$H(q, p, t) + \frac{\partial G}{\partial t} = 0. \quad (7.68)$$

Equation (7.68) is called the *Hamilton-Jacobi* equation. It determines the generating function that performs the canonical transformation to constant canonical variables⁷⁶. Once G (or S , or T , or U) has been determined, one can construct from it the associated canonical transformation

$$q_k = f_k(\beta_1, \dots, \beta_n; \alpha_1, \dots, \alpha_n; t) \quad (7.69)$$

$$p_k = g_k(\beta_1, \dots, \beta_n; \alpha_1, \dots, \alpha_n; t) \quad (7.70)$$

in terms of the $2n$ constants α_k and β_k and the time t . But at $t = 0$ (or any other convenient time instant), (7.69) and (7.70) allow one to relate the $2n$ constants (α_k, β_k) to the $2n$ initial values of the canonical variables (q_k, p_k), and thus allow one to calculate the q_k and p_k at any subsequent time in terms of their initial values, i.e. to solve the equations of motion in the original variables. Consequently, in solving the Hamilton-Jacobi equation, one is at the same time solving the associated dynamical problem.

To see how this works out, we must choose a particular set of independent variables to work with. The standard choice in discussing the Hamilton-Jacobi equation is to seek a generating function of the type $S = S(q, P, t)$, although any of the remaining three will do just as well. If we choose $G \rightarrow S$ as our unknown generating function in (7.68), then we may eliminate all the p 's in $H(q, p, t)$ with the help of (7.50) to find, in detail,

$$H(q_1, \dots, q_n; \frac{\partial S}{\partial q_1}, \dots, \frac{\partial S}{\partial q_n}; t) + \frac{\partial S}{\partial t} = 0, \quad (7.71)$$

⁷⁶ Sometimes the name Hamilton-Jacobi equation is reserved for this equation only when $S = S(q, P, t)$ is considered as the generating function. S itself is then called *Hamilton's principal function*.

the *Hamilton-Jacobi partial differential equation* in $(n + 1)$ variables for S . We have already met this equation as (1.111) of Chapter 1. The function S , designated Hamilton's principal function, depends on the n q 's and the time t , as well as the n constants, $P_k = \alpha_k$, that give the constant values of the new canonical momenta. The physical significance of these constant momenta is unspecified at this stage, and we will choose them later to suit our convenience. The important point to notice, however, is that none of the constants α_k in S are additive since S and S plus an arbitrary additive constant are both solutions of (7.71) (since this equation only involved derivatives of S). Now, the complete integral of (7.71) must contain as many constants of integration as there are variables, i.e. $(n + 1)$ constants. According to the above observation one of these is necessarily additive and to be discarded. We are then left with a solution $S(q_1, \dots, q_n, \alpha_1, \dots, \alpha_n; t)$ involving n non-additive constants. The second equation (7.51) now enters to relate the n constant values of the new coordinates, $Q_k = \beta_k$ to the q_k and the time,

$$\beta_k = \frac{\partial S}{\partial \alpha_k} = F_k(q_1, \dots, q_n; \alpha_1, \dots, \alpha_n; t). \quad (7.72)$$

These n relations can in principle be turned "inside out" to give the old coordinates as a function of the α_k, β_k and the time t ,

$$q_k = f_k(\alpha_1, \dots, \alpha_n; \beta_1, \dots, \beta_n; t). \quad (7.73)$$

Finally, with the help of this result and the relation

$$p_k = \frac{\partial S}{\partial q_k} \quad (7.74)$$

we can solve for the p_k in terms of the constants (α_k, β_k) and t in the form

$$p_k = g_k(\alpha_1, \dots, \alpha_n; \beta_1, \dots, \beta_n; t), \quad (7.75)$$

which brings us to the point envisaged in (7.69) and (7.70), and allows one to solve for the motion.

The procedure is best appreciated via an example. For the one-dimensional harmonic oscillator, H is given in (7.23). The Hamilton-Jacobi equation (7.71) for S reads in this instance

$$\frac{1}{2m} \left(\frac{\partial S}{\partial q} \right)^2 + \frac{1}{2} m \omega_0^2 q^2 + \frac{\partial S}{\partial t} = 0, \quad (7.76)$$

where $S = S(q, \alpha, t)$ contains a single non-additive constant α equal to the new canonical momentum. This partial differential equation may be solved by the method of separation of variables. Writing $S = W(q) + f(t)$ in (7.76) and using (7.52), we find that $f(t) = -Et$, where $H = E$ is the constant total energy of the oscillator. Hence,

$$S = W(q) - Et, \quad (7.77)$$

where the function $W(q)$ satisfies the ordinary differential equation

$$\frac{1}{2m} \left(\frac{dW}{dq} \right)^2 + \frac{1}{2} m \omega_0^2 q^2 = E, \quad (7.78)$$

with solution

$$W = \sqrt{2m} \int \sqrt{E - \frac{1}{2} m \omega_0^2 q^2} dq, \quad (7.79)$$

after leaving off the trivial additive constant of integration. While this expression is simple to integrate directly, we refrain from doing so until after using (7.72) to find q as a function of time and the constant new momentum α . We note that

$$S = W(q, E) - Et \quad (7.80)$$

contains a single non-additive constant, the total energy E in this case. What then is α ? The point is simply that we are now at liberty to identify the new canonical momentum (which is constant) with E , so that

$$P = \alpha = E, \quad (7.81)$$

or any function of E . Since the last term in (7.80) contains E linearly, the above identification is a natural choice. Then (7.72) reads

$$\beta = \frac{\partial S}{\partial E} = \frac{1}{\omega_0} \sin^{-1} \left\{ \sqrt{\frac{m\omega_0^2}{2E}} q \right\} - t, \quad (7.82)$$

or

$$q = \sqrt{\frac{2E}{m\omega_0^2}} \sin \omega_0 [(t + \beta)]. \quad (7.83)$$

The remaining relation, $p = \partial S / \partial q$ at $t = 0$, relates the canonical variables $(Q, P) = (\beta, E)$ to the initial momentum. But since we know $p = m\dot{q}$ from Hamilton's equations, it is simpler to calculate p directly from (7.83) as

$$p = \sqrt{2mE} \cos \omega_0 (t + \beta). \quad (7.84)$$

Now both β and E may be determined in terms of the initial values of q and p at $t = 0$, say. Thus S given by (7.79) and (7.80) generates a canonical transformation to new canonical variables that are respectively the time $-\beta$ when $q(t)$ vanishes and the total energy E of the oscillator.

7-5 Hamilton-Jacobi Theory for Conservative Systems

For systems where the Hamilton function is conserved, $H = \text{constant}$, it is always possible to effect the separation of the time variable as in (7.77). We only consider such systems from now on where H , when conserved, also represents the total energy, $H = E$. Then, in view of (7.80) we may write generally

$$S = W(q_1, \dots, q_n; E, \alpha_2, \dots, \alpha_n) - Et \quad (7.85)$$

for a system with n degrees of freedom. Here we have identified the "first" new canonical momentum α_1 with E , as we are at liberty to do. Then W satisfies the following partial differential equation if S is to satisfy the Hamilton-Jacobi equation,

$$H(q_1, \dots, q_n; \frac{\partial W}{\partial q_1}, \dots, \frac{\partial W}{\partial q_n}) = E. \quad (7.86)$$

The function W , which depends on the n coordinates q_k , and n constants ($E, \alpha_2, \dots, \alpha_n$) is called *Hamilton's characteristic function* to distinguish it from S . We also met (7.86) in Chapter 1 as (1.112), with W there called S_0 for uniformity of notation. Once (7.86) has been solved for W , we return to (7.72) to learn that $\beta_1 = \partial S / \partial E = \partial W / \partial E - t$, or that

$$t + \beta_1 = \frac{\partial W}{\partial E} \quad (7.87)$$

and

$$\beta_k = \frac{\partial S}{\partial \alpha_k} = \frac{\partial W}{\partial \alpha_k}, \quad k = 2, 3, \dots, n. \quad (7.88)$$

The second set of relations is particularly interesting since the time t appears nowhere explicitly. Thinking in terms of a system with two degrees of freedom for simplicity, (7.88) reads in detail

$$\beta = \frac{\partial W}{\partial \alpha} = F(q_1, q_2; E, \alpha) \quad (7.89)$$

if we call $\beta_2 = \beta$ and $\alpha_2 = \alpha$. This relation determines the relation between q_1 and q_2 at all times for given values of α , β and E , i.e. it determines the *orbit*. By contrast, (7.87) determines the transit time of the system from $-\beta_1$ to t , as we already observed in (1.109) of Chapter 1. A combination of all n relations (two in number for two degrees of freedom) give both the "orbit" and the q_k as a function of time for specified initial conditions.

(i) Central motion in two dimensions

Again, the procedure is best appreciated via an example. We return to the case of particle motion in a central field $V(r)$ and write down (7.86) in plane polar coordinates, $q_1 = r$ and $q_2 = \theta$. Then, in view of (7.29), one has

$$\frac{1}{2m} \left(\frac{\partial W}{\partial r} \right)^2 + \frac{1}{2mr^2} \left(\frac{\partial W}{\partial \theta} \right)^2 + V(r) = E. \quad (7.90)$$

However, we know that, in addition to E , the angular momentum $p_\theta = mr^2\dot{\theta}$ canonical to θ is also conserved. Being constant, p_θ , which is one of the old canonical coordinates, can serve just as well as one of the new canonical momenta, $\alpha_2 = p_\theta$, say. We give expression to this fact by separating W into a sum of a function $f(r)$ of r and an identity generator in θ , thus

$$W = f(r) + \theta \alpha_2. \quad (7.91)$$

It follows at once that the new momentum $P_\theta = \alpha_2$ is identical with p_θ since $p_\theta = \partial W / \partial \theta = \alpha_2$. The function $f(r)$ is determined by

$$\frac{1}{2m} \left(\frac{df}{dr} \right)^2 = E - U(r), \tag{7.92}$$

where $U = V(r) + p_\theta^2 / 2mr^2$ is the effective potential introduced previously in Chapter 2. Hence,

$$W = \theta p_\theta + \sqrt{2m} \int \sqrt{E - U} dr, \tag{7.93}$$

if we take the positive square root in evaluating f . Application of (7.88) with $\alpha_2 = p_\theta$ and $\beta_2 = \theta_0$ gives back (2.22) of Chapter 2 again,

$$\theta_0 = \frac{\partial W}{\partial p_\theta} = \theta - \int_{r_{\min}}^r \frac{p_\theta}{\sqrt{2m(E - U)}} \frac{dr}{r^2}, \tag{7.94}$$

if we identify θ_0 as the angle where $r = r_{\min}$, the closest approach distance. The time dependence of r is provided by the companion equation, (7.87). Calling $t_0 = -\beta_1$, the time when $r = r_{\min}$, one has

$$t - t_0 = \frac{\partial W}{\partial E} = \int_{r_{\min}}^r \frac{dr}{\sqrt{\frac{2}{m}(E - U)}}, \tag{7.95}$$

which is just (2.23). We have thus extracted the two relevant equations for describing central orbits from Hamilton's principal function in a rather systematic way. However, there is a price for this convenience: We must be able to solve (7.86) for W . The trick that was used in solving the central field problem by writing W as a *sum* of functions, with the coordinates distributed each to a function, is called "separating" variables.⁷⁷ The procedure is certainly successful whenever all the old canonical momenta bar one are constant. Examples of this kind are central-field motion in two or three dimensions, or the motion of a top under gravity. But this proviso is too restrictive; however we do not go into further detail and refer the interested reader to further literature research.

(ii) A non-separable problem

Before leaving the practical applications of Hamilton-Jacobi theory, it is well to point out that all is not necessarily lost if the Hamilton-Jacobi equation fails to separate. Often the resulting non-separable partial differential equation can still be solved by appealing to the vast literature on the theory of partial differential equations of the Hamilton-Jacobi type for guidance. An example of this nature is provided by the problem of the motion of a charge e in crossed electric and magnetic fields \mathbf{F} and \mathbf{B} . We use cartesian coordinates (x, y) to locate the charge and point $\mathbf{F} = F\hat{y}$ along y and $\mathbf{B} = B\hat{z}$ along z so that the motion is confined to the x, y plane. This electromagnetic field is described by scalar and vector potentials

$$\phi = -Fy, \quad \mathbf{A} = \frac{1}{2}B[-y\hat{x} + x\hat{y}] \tag{7.96}$$

⁷⁷ It is useful to note that the separation of variables for the corresponding quantum mechanical problem involves choosing a *product*, rather than a *sum* to represent the unknown function. The reason for this difference becomes clear if we recall the connection between the solutions of the wave equation in the quasi-classical limit $\hbar \rightarrow 0$ and Hamilton's characteristic function W ($\hbar =$ Planck's constant divided by 2π). Writing $\psi \sim \exp(iW/\hbar)$ for the wave function, one finds that the Schroedinger energy operator gives

$$\begin{aligned} & \left[-\frac{\hbar^2}{2m} \nabla^2 + V \right] \psi \\ & \simeq \left[\frac{1}{2m} (\nabla W)^2 + V \right] \\ & \times \exp\left(\frac{i}{\hbar} W\right) \\ & = E \exp\left(\frac{i}{\hbar} W\right) \end{aligned}$$

in the limit $\hbar \rightarrow 0$. Thus W is a solution of (7.86) for the corresponding classical problem. Now, the quantum mechanical solution $\psi(t, \theta)$ for the central field problem in two dimensions separates as $\psi(r, \theta) = F(r) \exp\left[\frac{i p_\theta}{\hbar} \theta\right] \sim \exp\left[\frac{i}{\hbar} f(r)\right] \exp\left[\frac{i p_\theta}{\hbar} \theta\right] = \exp\left[\frac{i}{\hbar} (f(r) + \theta p_\theta)\right]$ showing the genesis of $W = f + \theta p_\theta$ quite clearly.

so that (7.86) reads

$$\frac{1}{2m} \left[\left(\frac{\partial W}{\partial x} + \frac{1}{2} eBy \right)^2 + \left(\frac{\partial W}{\partial y} - \frac{1}{2} eBx \right)^2 \right] - eFy = E, \quad (7.97)$$

with the help of the expression (7.35) for H . This equation is not separable. However, by using the method of Charpit, one can show that

$$\frac{\partial W}{\partial x} = \frac{1}{2} eBy + \alpha, \quad (7.98)$$

where α is a constant of integration⁷⁸. Combining this result with the original partial differential equation, we can solve for $\partial W/\partial y$ and thus construct the differential $dW = (\partial W/\partial x)dx + (\partial W/\partial y)dy$. Hence,

$$W = \int \sqrt{2m(E + eFy) - (eBy + \alpha)^2} dy + \alpha x + \frac{1}{2} eBxy. \quad (7.99)$$

We take E and α as the new canonical momenta. Then (7.87) and (7.88) read

$$\beta_1 + t = \int_{y_0}^{y(t)} \frac{dy}{\sqrt{\frac{2}{m}(E + eFy) - \left(\frac{eB}{m}y + \frac{\alpha}{m}\right)^2}} \quad (7.100)$$

and

$$\beta_2 = x - \int_{y_0}^{y(x)} \frac{\left(\frac{eB}{m}y + \frac{\alpha}{m}\right)}{\sqrt{\frac{2}{m}(E + eFy) - \left(\frac{eB}{m}y + \frac{\alpha}{m}\right)^2}} dy, \quad (7.101)$$

where $y = y_0$ at $t = -\beta_1$, or $x = \beta_2$. Integration of these expressions is elementary and gives the path of the charge as a *trochoid* in general. In the special case that the charge starts out from rest at the origin at $t = 0$, one has both $E = 0$ and $\alpha = 0$ so that

$$\begin{aligned} y(t) &= \frac{eF}{\omega B} (1 - \cos \omega t), \\ x(t) &= \frac{eF}{\omega B} (\omega t - \sin \omega t), \quad \omega = \frac{eB}{m}, \end{aligned} \quad (7.102)$$

which are the parametric equations of a cycloid. Note that ω is just twice the Larmor frequency $\Omega = eB/2m$.

(iii) Motion in a non-inertial frame

As a final example of the utility of canonical transformations, we rederive the result given in (2.100) of Chapter 2 for the equation of motion of a particle relative to a rotating frame of reference. Let Σ' refer to a set of axes $Ox'y'z'$ that are rotating with angular velocity $\Omega = \Omega(t)$ about an axis passing through the common origin of $Ox'y'z'$ and a set of axes $Oxyz$ that constitute an inertial frame Σ . The motion of a particle in Σ is described by the Hamiltonian

$$H = \frac{p^2}{2m} + V(r), \quad (7.103)$$

⁷⁸ A. R. Forsyth, *ibid.*, pp420.

leading to the equation of motion $\dot{p} = -\text{grad}V = \mathbf{F}$. The coordinates \mathbf{r} and \mathbf{r}' of the particle in Σ and Σ' are related by the point transformation

$$\mathbf{r}' = A(t)\mathbf{r}, \quad (7.104)$$

representing the rotation of Σ' relative to Σ . A is the operator that performs this rotation, and is represented by an orthogonal matrix $AA^T = A^T A = I$ as we already discussed in Chapter 3. The elements $A_{ij} = A_{ij}(t)$ are time dependent. This has been indicated by the blanket notation $A(t)$. The generator of the point transformation $\mathbf{r}' = A(t)\mathbf{r}$ is

$$S = \mathbf{p}' \cdot (A(t)\mathbf{r}) = (A^T(t)\mathbf{p}') \cdot \mathbf{r}, \quad (7.105)$$

as may readily be proven from matrix algebra. Here \mathbf{p}' is the new canonical momentum that goes along with \mathbf{r}' . We have

$$\mathbf{p} = \text{grad}_{\mathbf{r}} S = A^T(t)\mathbf{p}'. \quad (7.106)$$

The new Hamiltonian determining the motion in Σ' is given by (??), or

$$K = \frac{p'^2}{2m} + U(\mathbf{r}') + \mathbf{p}' \cdot \left(\frac{\partial A}{\partial t} A^T(t) \right) \mathbf{r}', \quad (7.107)$$

after using (7.106), and the inverse transformation $\mathbf{r} = A^T(t)\mathbf{r}'$ to eliminate the old canonical variables, and calling $V(\mathbf{r}) = V(A^T\mathbf{r}') = U(\mathbf{r}')$. The effect of the operation $(\partial A/\partial t)A^T(t)$ follows on examining the effect of $A(t)$ on the position vector $\mathbf{R}_p(0)$ at $t = 0$ of a point P rigidly attached to Σ' . At time t this point is moved to $\mathbf{R}_p(t) = A(t)\mathbf{R}_p(0)$ by $A(t)$. The change in position of P in time dt is therefore

$$\mathbf{R}_p(t + dt) - \mathbf{R}_p(t) \simeq dt \frac{\partial A}{\partial t} \mathbf{R}_p(0) = dt \frac{\partial A}{\partial t} A^T(t) \mathbf{R}_p(t). \quad (7.108)$$

But this shift also equals $-dt(\Omega \times \mathbf{R}_p(t))$, where Ω is the angular velocity of Σ' at time t . Notice the *minus* sign. This is necessary because as we have seen in Chapter 3, $A(t)$ rotates the vector $\mathbf{R}_p(0)$ in the *opposite sense* to the rotation of axes envisaged previously. Thus, since

$$\frac{\partial A}{\partial t} A^T(t) \mathbf{R}_p(t) = -\Omega \times \mathbf{R}_p(t) \quad (7.109)$$

holds for any \mathbf{R}_p at any instant of time, we conclude that $(\partial A/\partial t)A^T(t)$ is equivalent to the "operation" $-\Omega \times$, i.e.

$$\frac{\partial A}{\partial t} A^T \rightarrow -\Omega \times. \quad (7.110)$$

Employing this result, we finally have

$$K = \frac{p'^2}{2m} + U(\mathbf{r}') - \mathbf{p}' \cdot (\Omega \times \mathbf{r}'). \quad (7.111)$$

The Hamilton equations in the new variables \mathbf{p}' and \mathbf{r}' become

$$\dot{\mathbf{p}}' = -\text{grad}_{\mathbf{r}'} K = \mathbf{F} + \mathbf{p}' \times \Omega \quad (7.112)$$

and

$$\dot{\mathbf{r}}' = \mathbf{v}' = \text{grad}_{\mathbf{p}'} K = \frac{1}{m} \mathbf{p}' + \mathbf{r}' \times \Omega. \quad (7.113)$$

We solve for the canonical momentum \mathbf{p}' from the above equation, $\mathbf{p}' = m\mathbf{v}' + m(\Omega \times \mathbf{r}')$, and find

$$\frac{d}{dt}(m\mathbf{v}') = \mathbf{F} + 2(m\mathbf{v}' \times \Omega) + \Omega \times (m\mathbf{r}' \times \Omega) + (m\mathbf{r}' \times \dot{\Omega}), \quad (7.114)$$

as in (2.100) of Chapter 2.

7-6 Periodic Systems and Action-Angle Variables

Periodic motion has, as we have seen, a rather special place in mechanics. We have devoted considerable space to discussing central orbits that are periodic, and a whole chapter to small vibrations. In preparation for a discussion of periodic systems in Hamilton-Jacobi theory, let us first point out that Hamilton's characteristic function W , which was originally introduced as the time-independent piece of S in (7.85), induces a canonical transformation of its own that is quite different from that introduced by S . Consider a conservative system, $H = E$. Let us enquire into the properties of the generating function, which we provisionally call $W = W(q_1, \dots, q_n, P_1, \dots, P_n)$, that transforms H into a new Hamilton function K which *only* depends on one half of the new canonical variables. Thus, all the Q_k (say) are cyclic, all the P_k are constants, α_k . From (7.50)- (??) one has that

$$p_k = \frac{\partial W}{\partial q_k} \quad (7.115)$$

$$Q_k = \frac{\partial W}{\partial \alpha_k} \quad (7.116)$$

$$H = H(\alpha_1, \dots, \alpha_n). \quad (7.117)$$

For a conservative system, $H = E$, so that W is determined by (7.86) again,

$$H(q_1, \dots, q_n; \frac{\partial W}{\partial q_1}, \dots, \frac{\partial W}{\partial q_n}) = E. \quad (7.118)$$

The solution for W will contain the constant $\alpha_1 = E$, and $n - 1$ additional constant of integration $\alpha_2, \dots, \alpha_n$, none of which are additive. Thus, as previously,

$$W = W(q_1, \dots, q_n; \alpha_1, \dots, \alpha_n), \quad (7.119)$$

where the α_k are identified with the new constant momenta P_k .

We now return to systems performing a finite motion that are in addition described by a separable Hamilton-Jacobi equation. The form of the generating function (7.119) is then

$$W = \sum_{i=1}^n W_i(q_i, \alpha_1, \dots, \alpha_n), \quad (7.120)$$

where each W_i only depends on a single coordinate q_i . The conjugate momenta, p_k ,

$$p_k = \frac{\partial W_k}{\partial q_k} = p_k(q_k, \alpha_1, \dots, \alpha_n) \quad (7.121)$$

are functions of the single canonical coordinate q_k and the set of constants $\alpha_1, \alpha_2, \dots, \alpha_n$. This circumstance, which only occurs in separable systems, allows one to classify the motion of each (q_k, p_k) pair into one of two types. If we plot p_k as a function of q_k as given by (7.121), the result is a curve in (q, p) -space, or the *phase-space* of the (q_k, p_k) variables. Then (i) if this curve is a closed one, q_k must oscillate back and forth between two turning points as the point (q_k, p_k) moves around the curve. Thus q_k returns to its original value if (q_k, p_k) moves once around the closed curve, and q_k is said to have completed its full *cycle*. This motion in q_k is called a *vibration* (or *libration*) with q_k returning to its original value after each cycle. (ii) If on the other hand q_k advances by ϕ_k every time the system returns to its original state, the motion is termed *rotational*, with a full cycle in q_k being represented by its changing by ϕ_k .

Systems performing a vibrational or rotational motion in the sense described above have special properties that are best exhibited in terms of a special set of canonical variables called *action angle* variables. The action variables J_k are a new set of constants that replace the constants α_k in (7.115) - (7.121). They are defined as (the 2π is arbitrary, but convenient)

$$2\pi J_k = \oint p_k dq_k, \quad (7.122)$$

where the integral $\oint dq_k \dots$ means that q_k is taken over its full cycle. Because of the separability displayed in (7.120), the J_k are indeed constants. Using the expression for p_k in (7.121),

$$J_k = \frac{1}{2\pi} \oint \frac{\partial W_k}{\partial q_k} dq_k = J_k(\alpha_1, \dots, \alpha_n), \quad (7.123)$$

since the single q_k appearing in W_k integrates away. By inverting this equation we can eliminate the α_k in favour of the J_k in (7.115) - (7.121).

Then the generating function becomes

$$W = \sum_{i=1}^n W_i(q_i, J_1, \dots, J_n), \quad (7.124)$$

giving rise to the following equations for the associated canonical transformation it induces:

$$p_k = \frac{\partial W_k}{\partial q_k} = p_k(q_k; J_1, \dots, J_n) \quad (7.125)$$

$$\psi_k = \frac{\partial W}{\partial J_k} = \psi_k(q_1, \dots, q_n; J_1, \dots, J_n) \quad (7.126)$$

$$K = H = E(J_1, \dots, J_n). \quad (7.127)$$

We have designated by ψ_k the variable conjugate to J_k . It is called an *angle* variable for reasons which will become clear presently.

The equations of motion and their solutions, for the ψ_k and J_k are, respectively,

$$\dot{\psi}_k = \frac{\partial K}{\partial J_k} = \frac{\partial E}{\partial J_k} = \omega_k(J_1, \dots, J_n) \quad (7.128)$$

with solutions

$$\psi_k(t) = \omega_k t + \psi_k(0), \quad k = 1, 2, \dots, n \quad (7.129)$$

and

$$\dot{J}_k = -\frac{\partial K}{\partial \psi_k} = 0, \quad (7.130)$$

with solutions

$$J_k = \text{constant}, \quad k = 1, 2, \dots, n. \quad (7.131)$$

The constants $\omega_k = \omega_k(J_1, \dots, J_n)$ are called the *fundamental frequencies* of the separable system. We will also have more to say about them in a moment.

The canonical pairs (ϕ_k, J_k) are known as *action-angle* variables. They are specific to separable systems performing a periodic motion (in order to define a q_k cycle) and, as we have seen, have the property that each ψ_k increases linearly with time, while the J_k are constants. Geometrically, $2\pi J_k$ represents the "area" in phase space, enclosed by the orbit described by the point (q_k, p_k) in a vibrational motion, or the area under this curve for one cycle of q_k , if the motion is of the rotational type. Consequently,

$$J_k = J_k(q_1, \dots, q_n; p_1, \dots, p_n) \quad (7.132)$$

is a single-valued function of the old coordinates so that the J_k provide for n single-valued constants of motion of a separable system. In contrast with the J_k , the ψ_k increase by 2π for every cycle in q_k , the remaining q_l 's ($q_l \neq q_k$) and all the J 's being held fixed. We see this by calculating the change $\Delta_l \psi_k$ due to a full cycle in q_l : From (7.126),

$$\begin{aligned} \Delta_l \psi_k &= \oint \frac{\partial^2 W}{\partial q_l \partial J_k} dq_l = \frac{\partial}{\partial J_k} \oint \frac{\partial W_l}{\partial q_l} dq_l \\ &= 2\pi \frac{\partial J_l}{\partial J_k} = 2\pi \delta_{kl}. \end{aligned} \quad (7.133)$$

Notice that (7.133) does *not* mean ψ_k changes by 2π during q_k 's cycle in the actual motion, because other q 's are also changing in this case. However, it does mean that each q_k must be a *periodic* function in the ψ_k , with a fundamental period of 2π . Thus, if we solve (7.125) or (7.126) for the q_k in terms of the action-angle variables,

$$q_k = q_k(\psi_1, \psi_2, \dots; J_1, J_2, \dots), \quad (7.134)$$

these solutions must have the property that (the n 's are integers)

$$q_k(\psi_1 + 2\pi n_1, \psi_2 + 2\pi n_2, \dots; J_1, J_2, \dots) = q_k(\psi_1, \psi_2, \dots; J_1, J_2, \dots). \quad (7.135)$$

Hence any *single-valued* q_k can be written as a multiple *Fourier series*

$$q_k = \sum_{s_1=-\infty}^{\infty} \cdots \sum_{s_n=-\infty}^{\infty} A_{s_1 s_2 \dots s_n}^k \exp \left[i \sum_{l=1}^n s_l \psi_l \right], \quad (7.136)$$

where the s_l are integers running over the indicated ranges. Introducing the explicit form of the ψ_k from (7.129), one obtains the time-dependence

$$\exp \left[it \sum_{l=1}^n s_l \omega_l \right] \quad (7.137)$$

for each term of given $(s_1 s_2 \dots s_n)$ in (7.136). Thus, each term of that sum oscillates periodically with a frequency $\sum_{l=1}^n s_l \omega_l$, but the sum of such terms is not periodic in general. One then speaks of a *conditionally* periodic motion of the $q_k = q_k(t)$ as a function of time. However, it often happens that two or more fundamental frequencies are commensurable, i.e. are connected by relations of the form

$$\frac{\omega_k}{n_k} = \frac{\omega_l}{n_l} = \dots = \omega, \quad (7.138)$$

where the n_k and n_l are integers. Such frequencies are said to be *degenerate*. If all n frequencies are degenerate, i.e. if there are n entries on the left hand side of (7.138), the degeneracy is termed *complete*. Completely degenerate systems display a simple periodic motion in all their coordinates with a common period $T = 2\pi/\omega$. We see this immediately upon replacing the ω_k in (7.137) by the degenerate frequencies $n_k \omega$. Then,

$$\exp \left[i\omega t \sum_l s_l n_l \right] \quad (7.139)$$

will return to its initial value after a time $T = 2\pi/\omega$ for any set of s_l . Consequently, the q_k in (7.136) are all periodic, $q_k(t + T) = q_k(t)$ with a common period T .

The occurrence of degeneracies is more the rule than the exception for many dynamical systems of interest. One important example is the degeneracy in particle motion in a spherically symmetric potential $V(r)$.

This degeneracy, which is due to the central nature of $V(r)$ shows up in the dependence of the particle energy

$$E = E(J_\phi + J_\theta, J_r) \quad (7.140)$$

on the sum of the first two action variables, where

$$2\pi J_\phi = \oint p_\phi d\phi, \quad 2\pi J_\theta = \oint p_\theta d\theta, \quad 2\pi J_r = \oint p_r dr \quad (7.141)$$

in spherical polar coordinates. Hence ω_ϕ and ω_θ are degenerate, since

$$\omega_\phi = \frac{\partial E}{\partial J_\phi} = \frac{\partial E}{\partial J_\theta} = \omega_\theta. \quad (7.142)$$

Particularizing $V(r)$ still further to be the inverse field $V(r) = -\alpha/r$ introduces an additional degeneracy,

$$E = E(J_\phi + J_\theta + J_r) = -\frac{1}{2} \frac{m\alpha^2}{(J_\phi + J_\theta + J_r)^2} \quad (7.143)$$

that is peculiar to this potential. Then all three frequencies ω_ϕ , ω_θ and ω_r are degenerate, and equal to the common frequency of motion

$$\omega = \frac{\partial E}{\partial J_\phi} = \frac{\partial E}{\partial J_\theta} = \frac{\partial E}{\partial J_r} = \frac{2}{\alpha} \sqrt{\frac{2|E|^3}{m}} \quad (7.144)$$

for all three variables. The motion is truly periodic, the orbit being a closed one as we already know from Chapter 2.

We have thus found the orbiting frequency from a knowledge of the function $E = E(J_r, J_\theta, J_\phi)$ *without* first solving the equations of motion. This result shows the power of using action-angle variables, and is typical of the general situation: If the energy $E = E(J_1, \dots, J_n)$ is known as a function of the action variables, then so are the fundamental frequencies ω_k . Often the function $E = E(J_1, \dots, J_n)$ can be constructed without solving the equations of motion first. However, whether or not the resulting ω_k represent actual frequencies of motion has to be decided on the basis of what degeneracies are present, as we have already discussed.

The advent of such degeneracies has a further consequence. The number of (single-valued) constants of motion then exceeds the number n represented by the n action variables $J_k = J_k(q, p)$ found in (7.67) for a finite, separable motion. Referring by way of an example to central motion in three dimensions one has the energy E , the total angular momentum l , and its projection l_z along an arbitrary axis as the three single-valued integrals of motion. If, however, $V(r) = -\alpha/r$, leading to the additional degeneracy recorded above there is an additional integral of motion,

$$(\mathbf{v} \times \mathbf{l}) - \frac{\alpha}{r} \mathbf{r} \quad (7.145)$$

by construction. We see from (7.137) how such additional single-valued integrals of motion can arise. For, apart from the n constants J_k , the $n - 1$ constructs

$$\psi_k \omega_l - \psi_l \omega_k \quad (7.146)$$

are also constant in time. However, they are not single-valued. But when $\omega_k/n_k = \omega_l/n_l = \omega$, the difference

$$\psi_k n_k - \psi_l n_l \quad (7.147)$$

only changes by multiples of 2π when the ψ_k change by 2π as per (7.133). Any trigonometric function of $\psi_k n_k - \psi_l n_l$ will therefore be an additional single-valued constant of motion.

The existence of more single-valued constants of motion than degrees of freedom in separable degenerate systems⁷⁹ also means that the choice of coordinates in which to represent the n action variables cannot be unique, i.e. that the associated Hamilton-Jacobi equation must be separable in more than one set of coordinates. For example, the Hamilton-Jacobi equation with $V(r) = -\alpha/r$ separates in both spherical polar and parabolic coordinates. The occurrence of degeneracies in $E(J_1, \dots, J_n)$, as exemplified by (7.143) also suggest a means for their removal. In the case of any central motion we can dispose of one action variable (J_ϕ say, with its attendant angle variable ψ_ϕ) by using our prior knowledge that the orbit lies in a plane. Using polar coordinates (r, θ) , in this plane, the energy will in general be a function of J_θ and J_r separately, $E = E(J_\theta, J_r)$. However, if $V(r) = -\alpha/r$, one has a further degeneracy, since

$$E = -\frac{1}{2} \frac{m\alpha^2}{(J_\theta + J_r)^2}. \quad (7.148)$$

If we now consider E in (7.148) to be the "old" Hamiltonian and introduce a second canonical transformation

$$G = (\psi_\theta - \psi_r)J' + \psi_r J \quad (7.149)$$

to new action-angle variables (ψ', J') and (ψ, J) where

$$J_\theta = \frac{\partial G}{\partial \psi_\theta} = J', \quad J_r = \frac{\partial G}{\partial \psi_r} = -J' + J \quad (7.150)$$

and

$$\psi' = \frac{\partial G}{\partial J'} = \psi_\theta - \psi_r, \quad \psi = \frac{\partial G}{\partial J} = \psi_r, \quad (7.151)$$

then the "new" Hamiltonian

$$J = E(J) = -\frac{m\alpha^2}{2J^2} \quad (7.152)$$

is only a function of J . As a result, the equations of motion for the new angle variables are

$$\dot{\psi}' = 0, \quad \dot{\psi} = \frac{\partial E}{\partial J} = \omega, \quad (7.153)$$

⁷⁹ The reader will recognise that much of this commentary has a counterpart in the properties of the associated Schroedinger wave equation, especially in view of what was said a few pages back about the connection between the solutions of the wave equation and the Hamilton-Jacobi equation. L.P. Eisenhart, Phys. Rev. 74, 87 (1948), has identified potentials for which the Schroedinger equation is separable. The same considerations obviously apply for the separation of the Hamilton-Jacobi equation.

showing that $\psi' = \text{constant}$, while ψ increases linearly with time as before,

$$\psi(t) = \omega t + \psi(0). \quad (7.154)$$

Referring back to (7.93), it is interesting to construct the canonical transformation from the old to the new variables explicitly. We have for ψ in particular,

$$\psi = \psi_{r'} = \frac{\partial W}{\partial J_r} = \omega \frac{\partial W}{\partial E} = \xi - e \sin \xi, \quad (7.155)$$

either by direct integration, or by comparison with (2.34) of Chapter 2, if we set $\psi = 0$ at the perihelion. The geometric constants of this orbit can also be expressed in terms of J and J' . Noting that $J' = J_\theta = p_\theta$, the constant angular momentum of the motion, one finds

$$a = \frac{J^2}{m\alpha}, \quad e = \sqrt{1 - \frac{J'^2}{J^2}} \quad (7.156)$$

for the semi-major axis and the eccentricity of the elliptic orbit (refer to (2.32) of Chapter 2).

7-7 Perturbations

The action-angle variables introduced in the previous section turn out to be very suitable variables in which to study the effect of perturbations of the Hamiltonian K in (7.127). We do not intend to present a systematic development of such *canonical perturbation theory*, on account of its complexity in relation to the sort of problem we want to discuss, but rather wish to illustrate some salient features by way of an example. For this purpose consider the motion of a particle in the central field $V(r) = -\alpha/r$, upon which a constant force field \mathbf{F} has been superposed. The potential energy for the perturbed problem is thus

$$V(r) = -\frac{\alpha}{r} - (\mathbf{r} \cdot \mathbf{F}). \quad (7.157)$$

We wish to investigate how the motion in the inverse field, i.e. the unperturbed motion, is affected by the additional interaction $(\mathbf{r} \cdot \mathbf{F})$, assuming this to be "small". The philosophy of perturbation theory runs as follows: We know the geometric constants of the unperturbed elliptic orbit of the particle in terms of J and J' . The presence of a small perturbation will therefore have the effect of introducing a slow time dependence (also relative to the period of elliptic motion) into J and J' . Thus, the geometric constants (and orientation) of the unperturbed ellipse are expected to change slowly with time. The problem is therefore to find out how the J 's vary with time. But since the variation in the shape and orientation of the ellipse is a slow one, we can as a first approximation replace $(\mathbf{r} \cdot \mathbf{F})$ by its time-average $\langle \mathbf{r} \cdot \mathbf{F} \rangle$ over one period of unperturbed motion. Now,

the time-average $\langle \mathbf{r} \rangle$ of the position vector \mathbf{r} of the particle relative to the center of force obviously lies along the major axis of the ellipse by symmetry. Its value is

$$\langle x \rangle = \frac{1}{r} \int_0^r r \cos \theta dt = \frac{\alpha}{2\pi} \int_0^{2\pi} (\cos \zeta - e)(1 - e \cos \zeta) d\zeta = -\frac{3}{2}ae, \quad (7.158)$$

after appealing to (2.34) and (2.43) of Chapter 2. Hence, the *average* perturbing interaction is

$$H' = -\langle \mathbf{r} \cdot \mathbf{F} \rangle = \frac{3}{2}aeF \cos \psi', \quad (7.159)$$

where ψ' is the angle between \mathbf{F} and the semi-major axis of the ellipse. If we now assume that \mathbf{F} lies in the plane of the orbit, then ψ' is just the constant angle variable conjugate to J' that gives the orientation of the unperturbed ellipse relative to some arbitrary direction that we now choose to be along \mathbf{F} . We see this by noting that $\psi' = \partial W / \partial J'$, leads to (7.94) again, with $\psi' = \theta_0$. The perturbed Hamiltonian $K' = K + H'$ is thus

$$K' = E(J) + \frac{3}{2}F\left(\frac{J^2}{m\alpha}\right)\sqrt{1 - \frac{J'^2}{J^2}} \cos \psi', \quad (7.160)$$

when expressed in the canonical variables J, J' , and ψ' . K' determines the evolution in time of these variables to first order in F . We have

$$\dot{j} = -\frac{\partial K'}{\partial \psi} = 0 \quad (7.161)$$

$$\dot{j}' = -\frac{\partial K'}{\partial \psi'} = \frac{3aF}{2J} \sqrt{J^2 - J'^2} \sin \psi' \quad (7.162)$$

$$\dot{\psi}' = \frac{\partial K'}{\partial J'} = -\frac{3aF}{2J} \frac{J'}{\sqrt{J^2 - J'^2}} \cos \psi'. \quad (7.163)$$

The first equation shows that J , and therefore the semi-major axis a , is not changed by H' . We have therefore re-introduced a into (7.162) and (7.163), since it is constant in time. Noting that

$$\frac{d}{dt}(\sqrt{1 - J'^2/J^2} \cos \psi') = 0, \quad (7.164)$$

in view of (7.162) and (7.163), or that

$$(\sqrt{1 - J'^2/J^2} \cos \psi') = \text{constant} = e_0, \quad (7.165)$$

(e_0 is the eccentricity of the unperturbed motion), we can eliminate ψ' between (7.162) and (7.163) to find

$$\ddot{j}' + \left(\frac{3aF}{2J}\right)^2 J' = 0. \quad (7.166)$$

Hence,

$$J'(t) = p_\theta \cos\left(\frac{3aF}{2J}t\right), \quad (7.167)$$

where p_θ is the unperturbed value of J' . The eccentricity $e(t)$ and orientation $\psi'(t)$ of the major axis oscillate with the same frequency

$$\omega' = \frac{3aF}{2J} \quad (7.168)$$

as $J'(t)$. From (7.156) and (7.165),

$$\begin{aligned} e(t) &= \sqrt{1 - (1 - e_0^2) \cos^2 \omega' t} \\ \cos \psi'(t) &= \frac{e_0}{e(t)}. \end{aligned} \quad (7.169)$$

These results show that the effect of H' in (7.159) is to cause the semi-major axis of the ellipse to oscillate with frequency ω' about the direction of \mathbf{F} while maintaining its original length. The eccentricity $e(t)$ waxes and wanes with the same frequency. It is interesting to note that $J'(t)$ changes sign periodically, indicating that the particle motion is reversed periodically (assuming that the perturbation theory holds for sufficiently long times for this to happen). The frequency of oscillation ω' is expected to be much slower than the frequency ω of elliptic motion. From (7.156),

$$\frac{1}{\omega'} \frac{\partial E}{\partial J} \simeq \frac{2m\alpha^2}{3aFJ^2} = \frac{4}{3} \frac{|E|}{aF}, \quad (7.170)$$

a ratio that is large if the size of the perturbing potential (as measured by aF) is small relative to the total energy in the elliptic motion.

The above calculations apply in particular to the case of a hydrogen atom placed in an external electric field (classical theory of the Stark effect). Actually, the Hamilton-Jacobi equation with $V(r)$ given by (7.157) is separable in parabolic coordinates, so that the problem is exactly solvable for arbitrary external electric fields. One can thus investigate the range of validity of our approximate solutions, as well as questions like whether the bounded elliptic motion can become unbounded if the external field is strong enough (see Problems).

7-8 Further Aspects of Canonical Transformations

We return to (7.52) and consider the case that the generating function $S = S(q, P)$ is (a) independent of time, and (b) differs infinitesimally from the identity transformation $\sum_k q_k P_k$ thus:

$$S = \sum_k q_k P_k + \eta f(q, P). \quad (7.171)$$

Here f is an arbitrary function, and η a small parameter characterizing the difference between S and the identity transformation. The equations

of transformation read

$$\begin{aligned} p_k &= \frac{\partial S}{\partial q_k} = P_k + \eta \frac{\partial f}{\partial q_k} \\ Q_k &= \frac{\partial S}{\partial P_k} = q_k + \eta \frac{\partial f}{\partial P_k}. \end{aligned} \quad (7.172)$$

If η is small however, we obtain the differences

$$\begin{aligned} \delta p_k &= P_k - p_k \simeq -\eta \frac{\partial}{\partial q_k} f(q, p) \\ \delta q_k &= Q_k - q_k \simeq \eta \frac{\partial}{\partial p_k} f(q, p), \end{aligned} \quad (7.173)$$

correct to first order in η , upon replacing P_k by p_k everywhere in the last terms on the right of (7.172). The generator S is then said to generate the *infinitesimal canonical transformation* given in (7.173).

Consider now any function $F = F(q, p)$ of the canonical variables but not an explicit function of the time. If q_k and p_k are shifted by the amounts given by (7.173) in an infinitesimal canonical transformation, the change in F is

$$\delta F = F(q + \delta q, p + \delta p) = \eta \sum_k \left(\frac{\partial F}{\partial q_k} \frac{\partial f}{\partial p_k} - \frac{\partial F}{\partial p_k} \frac{\partial f}{\partial q_k} \right). \quad (7.174)$$

The combination of derivatives under the sum appears often enough to merit a name and a special symbol. The expression

$$\sum_k \left(\frac{\partial F}{\partial q_k} \frac{\partial f}{\partial p_k} - \frac{\partial F}{\partial p_k} \frac{\partial f}{\partial q_k} \right) = [F, f]_{q,p} \quad (7.175)$$

is called a *Poisson bracket*, and is denoted by the square bracket symbol on the right. Thus the change in F induced by the infinitesimal generator ηf appears as

$$\delta F = \eta [F, f]_{q,p}. \quad (7.176)$$

We will discover later that the subscripts (q, p) are not necessary. Poisson brackets are invariant under canonical transformations and may thus be evaluated in any convenient set of canonical variables. Let us investigate some special cases of (7.176). Since we are allowed to choose both f and the infinitesimal parameter η , take $f = H(q, p)$ a time-independent Hamilton function of some system, and $\eta = \delta t$, an infinitesimal element of time. Then,

$$\delta F = \delta t [F, H], \quad (7.177)$$

or, writing $\delta F = \dot{F} \delta t$ as $\delta t = 0$,

$$\dot{F} = [F, H]. \quad (7.178)$$

The time rate of change of F equals its Poisson bracket with the Hamiltonian responsible for changing q_k and p_k by δq_k and δp_k in time δt . In fact, setting $F = p_k$ and q_k respectively one finds

$$\begin{aligned}\delta p_k &= \dot{p}_k \delta t = \delta t [p_k, H] = -\delta t \frac{\partial H}{\partial q_k} \\ \delta q_k &= \dot{q}_k \delta t = \delta t [q_k, H] = \delta t \frac{\partial H}{\partial p_k}.\end{aligned}\quad (7.179)$$

In a very real sense then, the Hamilton function is the generator of the infinitesimal transformation that shifts q_k and p_k from their values at time t , to their values that obtain at time $t + \delta t$ during the actual motion in the time δt .

Two further views of (7.178) and its special case are important. Firstly, if $F = F(q, p)$ happens to be a constant of the motion, $\dot{F} = 0$, then

$$[F, H] = 0, \quad (7.180)$$

i.e. all constants of motion have a vanishing Poisson bracket with the Hamiltonian⁸⁰. Secondly if we go back to (7.176) and calculate the change in H itself under the infinitesimal transformation ηf , then

$$\delta H = \eta [H, f]. \quad (7.181)$$

Comparing this with the preceding equation, we may say: The generators f of those infinitesimal canonical transformations that leave the Hamiltonian invariant, $\delta H = 0$; are constants of the motion.

The symmetry properties of H , and the constants of motion of the dynamical system it describes, are brought to the fore once again. For example if H is translationally invariant, the total momentum is conserved. To see this in the context of the present discussion, consider an N -particle system with a translationally invariant H . Then $\delta H = 0$ upon shifting each particle position vector \mathbf{r}_i by a small common amount $\delta \mathbf{a} = \delta \mathbf{r}_i$. The generator for this infinitesimal transformation is

$$\eta f = \sum_{i=1}^N \delta \mathbf{a} \cdot \mathbf{p}_i = \delta \mathbf{a} \cdot \mathbf{P}, \quad (7.182)$$

where \mathbf{P} is the total momentum. If $\delta H = 0$ then by (7.181) the quantity $\delta \mathbf{a} \cdot \mathbf{P}$ is a constant of motion. Thus, \mathbf{P} itself must be constant since $\delta \mathbf{a}$ is an arbitrary vector.

Likewise, a rotationally invariant Hamiltonian conserves the total angular momentum of the system it describes. A common rotation of all particle coordinates through an angle $\delta \theta$ about any axis $\hat{\mathbf{n}}$ shifts the coordinates and associated momenta by

$$\begin{aligned}\delta \mathbf{r}_i &= \delta \theta (\hat{\mathbf{n}} \times \mathbf{r}_i) \\ \delta \mathbf{p}_i &= \delta \theta (\hat{\mathbf{n}} \times \mathbf{p}_i).\end{aligned}$$

⁸⁰ An identical statement holds in quantum mechanics if the Poisson bracket is replaced by the commutator bracket.

The generator for this rotation is found to be

$$\eta f = \delta\theta \sum_{i=1}^N \hat{\mathbf{n}} \cdot (\mathbf{r}_i \times \mathbf{p}_i) = \delta\theta(\hat{\mathbf{n}} \cdot \mathbf{L}), \quad (7.183)$$

where $\mathbf{L} = \sum_i(\mathbf{r}_i \times \mathbf{p}_i)$ is the total angular momentum. If H is left invariant by this rotation, then $\delta\theta(\hat{\mathbf{n}} \cdot \mathbf{L})$, and therefore \mathbf{L} itself is a constant of motion.

7-9 Liouville's Equation

There is another way of looking at Hamilton's equations of motion,

$$\dot{q}_k = \frac{\partial H}{\partial p_k}, \quad \dot{p}_k = -\frac{\partial H}{\partial q_k}, \quad k = 1, 2, \dots, n, \quad (7.184)$$

that is associated with the idea of the *phase space* of the system. For, if instead of thinking of the $2n$ variables (q_k, p_k) and how each varies with time, we follow the motion of the single representative point with "coordinates" $(q_1, \dots, q_n; p_1, p_2, \dots, p_n)$ in the $2n$ -dimensional space with axes labelled by the q 's and p 's, then Hamilton's equations give us the "velocity" components of this point directly. The $2n$ -dimensional space spanned by the $2n$ canonical variables is called the *phase space* of the system, and the point $(q_1, \dots, q_n; p_1, p_2, \dots, p_n)$ occupied by the system at time t its representative point, or *system point* in phase space. As the system moves according to the dictates of its Hamilton function and initial conditions, its system point moves along some curve, called a phase path, in phase space. Notice that the initial conditions (specification of the q 's and p 's at some initial time) tells us where the system point starts out in phase space, while Hamilton's equations determine how it proceeds away from this initial configuration.

(i) Liouville's Theorem

One important property of phase space is that the volume element

$$d\Omega = dq_1 \dots dq_n dp_1 \dots dp_n, \quad (7.185)$$

and therefore also any finite volume $\int d\Omega$ of phase space is invariant under a canonical transformation of variables. For consider by way of illustration a system with one degree of freedom, and introduce new canonical variables (Q, P) in place of (q, p) . Then

$$d\Omega = dqdp = JdQdP = Jd\Omega', \quad (7.186)$$

where

$$J = \frac{\partial(q, p)}{\partial(Q, P)} = [q, p]_{Q, P} \quad (7.187)$$

is the Jacobian for the transformation $[q, p] \rightarrow [Q, P]$. The anticipated invariance $d\Omega = dqdp = d\Omega'$ therefore amounts to showing that the

Jacobian of all canonical transformations is unity. This brings in the idea of the fundamental Poisson brackets $[q, p]$, $[q, q]$ and $[p, p]$ and raises the question of how Poisson brackets transform under canonical transformations. Consider any Poisson bracket $[F, G]_{q,p}$, where F and G are functions of the canonical variables q and p . We assert that

$$[F, G]_{q,p} = [F, G]_{Q,P}, \quad (7.188)$$

where (Q, P) are new canonical variables, i.e. that the Poisson bracket is invariant under canonical transformations, and that both sets of subscripts are unnecessary in (7.188). The proof of (7.188) involves straightforward but tedious algebra using the properties of canonical transformations (see Problems). A direct argument for proving the invariance has been suggested by Landau and Lifshitz⁸¹. If we suppose G is the Hamiltonian of some fictitious system then, in view of (7.178), the Poisson bracket $[F, G]_{q,p}$ gives the time rate of change of F . But this must be given equally by $[F, G]_{Q,P}$, since the time rate of change can only depend on the properties of the system, not the choice of coordinates. Therefore, (7.188) must be true. Returning to our problem, we compute the invariant values

$$[q, p]_{Q,P} = [q, p]_{q,p} = 1 \quad (7.189)$$

$$[q, q]_{Q,P} = [q, q]_{q,p} = 0 \quad (7.190)$$

$$[p, p]_{Q,P} = [p, p]_{q,p} = 0 \quad (7.191)$$

of the fundamental Poisson brackets. Consequently $J = 1$ in (7.186), confirming the invariance $dq dp = dQ dP$. The proof for the $2n$ -dimensional volume element in (7.185) is similar (see Problems). Therefore, it also follows that

$$\int dq_1 \dots dq_n dp_1 \dots dp_n = \int dQ_1 \dots dQ_n dP_1 \dots dP_n, \quad (7.192)$$

where the integral is taken over some volume of phase space.

The invariance of the volume in phase space under a canonical transformation leads to the following conclusion, known as *Liouville's Theorem*. We consider a collection of system points in a given volume of phase space. Such a collection might arise for example from considering a large number of identical mechanical systems with differing starting conditions. As time progresses each system point moves in accordance with Hamilton's equations, so that the boundaries of the phase space under consideration moves as well. However, as we saw in (7.179), the change in each q_k and p_k during the actual motion is made up of a series of infinitesimal canonical transformations generated by the Hamiltonian.

⁸¹ L. Landau and E. Lifshitz, *Mechanics*, (Addison-Wesley Inc., Cambridge Mass., 1960) p. 145.

Therefore the volume occupied by the system points under consideration cannot change, or

$$\int d\Omega = \text{constant}, \quad (7.193)$$

a result known as Liouville's theorem.

(ii) Liouville's Equation

We can follow up one consequence of Liouville's theorem a little further by introducing the concept of density $\rho(q_1, \dots, q_n; p_1, \dots, p_n; t)$ of system points in phase space. Then

$$\rho d\Omega \quad (7.194)$$

gives the number of system points in the volume element $d\Omega = dq_1 \dots dq_n \times dp_1 \dots dp_n$ at time t . As time marches on, the system points move in accordance with Hamilton's equations, to occupy a new volume $d\Omega'$ at time $t + \delta t$. However, the *number* of system points in $d\Omega$ at time t must be the same as the number in $d\Omega'$ at time $t + \delta t$ since no system points can leave or enter the volume under consideration. Now, each representative system point changes position and momentum by $\dot{q}_k \delta t$ and $\dot{p}_k \delta t$ in time δt . Therefore the conservation of the number of system points means that

$$\rho(q_1 + \dot{q}_1 \delta t, \dots; p_1 + \dot{p}_1 \delta t, \dots; t + \delta t) d\Omega' = \rho(q_1, \dots; p_1, \dots; t) d\Omega', \quad (7.195)$$

or, since $d\Omega = d\Omega'$ by Liouville's theorem, that

$$\sum_k \left(\dot{q}_k \frac{\partial \rho}{\partial q_k} + \dot{p}_k \frac{\partial \rho}{\partial p_k} \right) + \frac{\partial \rho}{\partial t} = 0. \quad (7.196)$$

We insert Hamilton's equation for q_k and p_k at this point to find Liouville's equation,

$$[\rho, H] + \frac{\partial \rho}{\partial t} = 0 \quad (7.197)$$

for the density of system points in phase space. Obviously, the left hand side of either equation gives the total time derivative of ρ , so that Liouville's equation may also be written simply as $\dot{\rho} = 0$.

Further developments along these lines lead into statistical mechanics, an aspect that lies beyond the purpose of this chapter. Nevertheless, we cannot resist presenting a problem, due to Max Born⁸², that bears on the question of in what sense classical mechanics may be regarded as deterministic. It is obvious from the mathematical problem posed by Hamilton's equations that the values of the canonical variables q_k and p_k can be calculated *exactly* at any later time in terms of their initial values $q_k(0)$ and $p_k(0)$ at time $t = 0$ say. That is to say, Hamilton's equations are fully deterministic. However, this statement tacitly presupposes that such initial values, or starting conditions, are known with arbitrary precision. We now pose, with Max Born, the following question: Suppose the

⁸² M. Born, *Physikalische Blätter* **15**, 342 (1959).

starting conditions are known only imprecisely. What then can be said about q_k and p_k at later times? Consider the motion of a free particle in one dimension by way of an example. Its position and momentum at any time t are given by

$$q(t) = q_0 + \frac{p_0}{m}t, \quad p = p_0 \quad (7.198)$$

in terms of their initial values q_0 and p_0 at $t = 0$. If, however, these initial values are imprecise, we can still assign a *probability*

$$f(q_0, p_0)d\Omega_0 \quad (7.199)$$

to the occurrence of a given q_0 and p_0 in the phase volume $d\Omega_0 = dq_0dp_0$ at $t = 0$. This is equivalent to considering N identical particles that start off with different values of (q_0, p_0) such that the density of system points at $t = 0$ is

$$\rho(q_0, p_0, 0) = Nf(q_0, p_0). \quad (7.200)$$

By Liouville's equation, this must also equal the density at all later times,

$$\rho(q, p, t) = \rho(q_0, p_0, 0), \quad (7.201)$$

or

$$f(q, p, t) = f(q_0, p_0). \quad (7.202)$$

Born considers the example where q_0 and p_0 have independent Gaussian distributions about their respective means, so that

$$f(q_0, p_0) = \frac{1}{2\pi\sigma\tau} \exp\left\{-\left(\frac{q_0 - \bar{q}}{\sigma}\right)^2 - \left(\frac{p_0 - \bar{p}}{\tau}\right)^2\right\}, \quad (7.203)$$

where σ and τ measure the scatter in q_0 and p_0 about their means \bar{q} and \bar{p} . At all later time, therefore

$$f(q, p, t) = f\left(q - \frac{p}{m}t, p\right) = \frac{1}{2\pi\sigma\tau} \exp\left\{-\left(\frac{q - \bar{q} - pt/m}{\sigma}\right)^2 - \left(\frac{p - \bar{p}}{\tau}\right)^2\right\} \quad (7.204)$$

in view of (7.202) and (7.198). Notice that the q and p distributions interlock at later times, $f(q, p, t)$ no longer being a product of a function of q times a function of p . Curves of equal probability are therefore ellipses

$$\left(\frac{q - \bar{q} - pt/m}{\sigma}\right)^2 + \left(\frac{p - \bar{p}}{\tau}\right)^2 = C \quad (7.205)$$

that enclose an area in phase space within which $f(q, p, t) >$ a constant determined by C . We examine the representative case $C = 1$. Setting

$$q - \bar{q} - \frac{\bar{p}}{m}t = \sigma\xi, \quad p - \bar{p} = \tau\eta, \quad (7.206)$$

we obtain

$$(\xi - \alpha t\eta)^2 + \eta^2 = 1. \quad (7.207)$$

This is an ellipse in the variables ζ and η , referred to its center. Its semi-major axis a makes an angle

$$\chi = \frac{1}{2} \cot^{-1}\left(\frac{1}{2}\alpha t\right) \simeq \frac{1}{\alpha t}, \quad \alpha t \gg 1, \quad (7.208)$$

with the positive ζ -axis and has the value

$$a = \cot \chi \simeq \alpha t. \quad (7.209)$$

At $t = 0$, this ellipse reduces to a unit circle in the variables (ζ, η) , the center of which lies at (\bar{q}, \bar{p}) . The area, π , of this circle must by Liouville's theorem equal the area of the ellipse (7.207) at all later times. Therefore, the axes a and b of the ellipse are reciprocal,

$$b = \frac{1}{a} = \tan \chi \simeq \frac{1}{\alpha t}. \quad (7.210)$$

So, as time progresses, two things happen: From (7.207) we see that the center of the circle moves with the mean speed $\bar{v} = \bar{p}/m$ of the distribution, while the circle itself deforms into an ellipse that becomes more and more "emaciated" as its major axis dips down towards the ζ -axis, see Fig. 7.1.

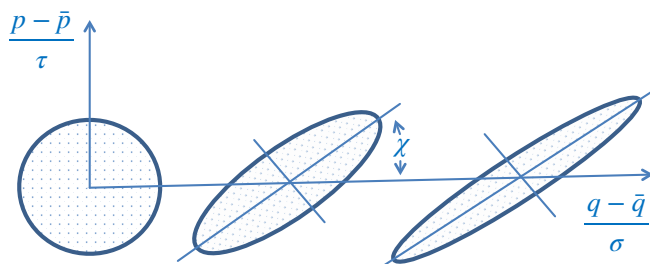


Figure 7.1: Area-preserving ellipses in the phase space of a free particle, shown at three equal intervals of time. System points inside the circular boundary at $t = 0$ remain inside the corresponding elliptic boundary as time increases.

Thus, the definition of the one canonical variable, p gets sharper at the expense of the other, q , a result we can understand by noting again that $f(q, p, t)$ is *not* a simple product of q and p distributions for $t > 0$,

The sharpening up of one canonical variable at the expense of the other has an interesting consequence for periodic systems. We have seen that such systems can always be described in terms of action-angle variables. Thus, if we introduce action-angle variables (ψ, J) for a one-dimensional system with Hamiltonian $K = E(J)$, then

$$\begin{aligned} \psi(t) &= \psi_0 + \omega(J)t \\ J(t) &= J_0 \\ \omega(J) &= \frac{\partial E}{\partial J}. \end{aligned} \quad (7.211)$$

We suppose that ψ_0 and J_0 also have a Gaussian distribution as in (7.203). If we further assume that the J -distribution is sharply peaked at \bar{J} so that

the first two terms in

$$\omega(J) = \omega(\bar{J}) + \frac{\partial^2 E}{\partial \bar{J}^2} (J - \bar{J}) + \dots \quad (7.212)$$

give an adequate representation of $\omega(J)$, we regain (7.207) once more, with

$$\psi - \bar{\psi} - \omega(\bar{J})t = \sigma \xi, \quad J - \bar{J} = \tau \eta \quad (7.213)$$

and

$$\alpha = \frac{\mu \sigma}{\tau}, \quad \mu = \frac{\partial^2 E}{\partial \bar{J}^2}. \quad (7.214)$$

Consequently, the action variable becomes better defined at the expense of the angle variable as time passes. In fact, since ψ increases by 2π during each cycle of the system, we will have lost all information of ψ after a time t , where the semi-major axis of the ellipse, $a \simeq \alpha t$, exceeds $2\pi/\sigma$, i.e. when

$$t \gg t_c = \frac{2\pi}{\mu \tau}. \quad (7.215)$$

Thus, t_c represents a critical time for the periodic system. After this time we have lost all information on the angle variable. Notice that the time t_c is inversely proportional to the spread τ of the action variable, provided that $\mu \neq 0$. The latter circumstance is a special feature of a linear harmonic oscillator for which $E(J) = \omega J$ and $\mu = 0$. Thus, $t_c \rightarrow \infty$ in this case, a result that is immediately understood by noting that the ξ and η (or ψ and J) distributions remain uncoupled for all times if $\mu = 0$.

Similar considerations follow for a system with n degrees of freedom, described by a Hamiltonian $K = E(J_1, \dots, J_n)$. Equations (7.211) and (7.212) are replaced by

$$\begin{aligned} \psi_k(t) &= \psi_{0,k} + \omega_k t \\ J_k(t) &= J_{0,k} \\ \omega_k &= \frac{\partial E}{\partial J_k} \end{aligned} \quad (7.216)$$

and

$$\begin{aligned} \omega_k &= \bar{\omega}_k + \sum_l \frac{\partial^2 E}{\partial \bar{J}_k \partial \bar{J}_l} (J_l - \bar{J}_l) + \dots \\ \bar{\omega}_k &= \omega_k(\bar{J}_1, \dots, \bar{J}_n) \end{aligned} \quad (7.217)$$

for each value of k . Assuming Gaussian distributions in $\psi_{0,k}$ and $J_{0,k}$ with widths σ_k and τ_k ,

$$\sum_{k=1}^n \left\{ (\xi_k - t \sum_{l=1}^n \alpha_{kl} \eta_l)^2 + \eta_k^2 \right\} = 1 \quad (7.218)$$

replaces (7.207), if we set

$$\psi_k - \bar{\psi}_k - \bar{\omega}_k t = \sigma_k \xi_k, \quad J_k - \bar{J}_k = \tau_k \eta_k, \quad (7.219)$$

and call

$$\alpha_{kl} = \frac{\mu_{kl}\tau_l}{\sigma_k}, \quad \mu_{kl} = \frac{\partial^2 E}{\partial J_k \partial J_l}. \quad (7.220)$$

We take $\sigma_k \tau_k = \sigma_l \tau_l \neq 0$ for simplicity and exclude systems (harmonic oscillators) for which $\mu_{kl} = 0$. Then α_{kl} is a real, symmetric matrix which can be diagonalized by a real orthogonal matrix U . Call its eigenvalues α_i , and transform ξ and η to new variables $U\xi'$ and $U\eta'$. Then (7.218) is replaced by

$$\sum_{i=1}^n \{(\xi'_i - \alpha_i t \eta'_i)^2 + \eta_i'^2\} = 1. \quad (7.221)$$

Being a sum of terms like those in (7.207), we can analyze this equation for a longest critical time (α is the smallest eigenvalue of α_{kl})

$$t_c = \frac{2\pi}{\alpha} \left[\sum_{k=1}^n \frac{1}{\sigma_k^2} \right]^{\frac{1}{2}}, \quad (7.222)$$

after which time we lack all information about one half of the action-angle variables: The remarkable feature about this result is that it holds for any system for which action-angle variables can be defined. The Hamiltonian $K = E(J_1, \dots, J_n)$ itself has remained unspecified and incidental to the whole discussion.

Problems

7-1. Show that the generating functions $T(p, Q, t)$ and $U(p, P, t)$ lead to the canonical transformations given by (7.53) - (7.58).

7-2. Prove directly that the point transformation $q_k = f_k(Q_1, \dots, Q_n; t)$ where $k = 1, 2, \dots, n$ is canonical.

7-3. Set up and solve the Hamilton-Jacobi equation for the motion of a projectile near the surface of the earth. Find both the equation for the path, as well as the coordinates as a function of the time.

7-4. Set up and solve the Hamilton-Jacobi equations for the motion of a top spinning about a fixed point. Take the analysis as far as you can.

7-5. Confirm the results given in paragraph 7-5(ii) of the text concerning the motion of a charged particle in crossed electric and magnetic fields.

7-6. Obtain the result

$$E = -\frac{1}{2} \frac{m\alpha^2}{(J_\phi + J_\theta + J_r)^2} \quad (7.223)$$

for the energy of a particle, mass m , moving in the central potential $V(r) = -\alpha/r$. (Remark: The action variable $2\pi J_r = \oint p_r dr$ can be evaluated in a straightforward way by solving the energy equation for p_r . The resulting integral over r can then be done either pedantically, using integration tables, or elegantly using contour integration.)

7-7. Using parabolic coordinates, set up and solve the Hamilton-Jacobi equation for a hydrogen atom in an external electric field E_0 . Introduce action-angle variables and check the conclusions reached in paragraph 7-7 of the text. Address the question of whether there is a critical electric field strength for which the orbit is rendered unbound. Estimate a typical value for ea_0E_0 (e = electron charge, a_0 = Bohr radius) and decide whether or not perturbation theory is applicable to hydrogen for the sorts of electric fields usually available in the laboratory.

7-8. A hydrogen atom is placed in a constant magnetic field \mathbf{B} , pointing perpendicular to the plane of the orbiting electron. Assuming \mathbf{B} is "small", carry out a similar analysis to that in paragraph 7-7, to study how the electron orbit is changed by the external field.

7-9. It is known from electrodynamics that an accelerating charge e radiates energy at the rate

$$P = \frac{2}{3} \frac{e^2}{c^2} |\dot{\mathbf{v}}|^2, \quad (7.224)$$

where $\dot{\mathbf{v}}$ is the acceleration and c the velocity of light, and the charge e is measured in esu. Now consider the electron in the lowest state of a hydrogen atom. Classically its orbit is a circle of radius $e_0 = \hbar^2/mc^2 = 0.53 \times 10^{-8}$ cm (the Bohr radius). Assuming that this orbit remains approximately circular while the electron is syphoning off energy at the rate P , show that the radius of the orbit will contract to

$$r(t) = a_0 \left[1 - \frac{8}{3} \frac{r_0^3}{a_0^3} \frac{t}{\tau} \right]^{\frac{1}{3}}. \quad (7.225)$$

Here the constants are the classical electron radius $r_0 = e^2/mc^2$ ($= 2.8 \times 10^{-13}$ cm) and τ , where $c\tau = (2/3)r_0$ is the characteristic time for light to travel across the classical electron. Assuming the above formula to be valid at all distances, calculate the time for a classical hydrogen atom to contract to a point. How many orbits does the electron complete in this time? Why do real hydrogen atoms not behave in this way? (Don't just say "quantum mechanics". Try to understand the reason in detail).

7-10. The orientation in space and geometric constants of an orbit of a planet are sometimes described in terms of its *Delaunay elements*. Look

up what these are and how they relate to the action-angle variables. Then study the problem posed by (7.157) again, that is particle motion in the field $V(r) = -\alpha/r - (\mathbf{r} \cdot \mathbf{F})$, but where the force \mathbf{F} now has an *arbitrary* orientation relative to the plane of the elliptic orbit. Generalize problem 7-8 in the same way.

7-11. Prove that the Jacobian

$$J = \frac{\partial(q_1, \dots, q_n, p_1, \dots, p_n)}{\partial(Q_1, \dots, Q_n, P_1, \dots, P_n)} = 1 \quad (7.226)$$

if the transformation from the (q_k, p_k) to the (Q_k, P_k) is a canonical one.

7-12. Show that the fundamental Poisson brackets of $2n$ canonical variables q_k and p_k are

$$[q_k, p_k] = \delta_{kl}, \quad [q_k, q_l] = 0, \quad [p_k, p_l] = 0. \quad (7.227)$$

We saw in the text that Poisson brackets are canonical invariants. Prove the converse of this statement for the fundamental Poisson brackets, that is to say, transform to new (not necessarily canonical) variables Q_k and P_k , and show that these variables also satisfy the canonical equations of motion if q_k and p_k do, provided that

$$[Q_k, P_k]_{q,p} = \delta_{kl}, \quad [Q_k, Q_l]_{q,p} = 0, \quad [P_k, P_l]_{q,p} = 0. \quad (7.228)$$

Assume the transformation $(q, p) \rightarrow (Q, P)$ to be time-independent for simplicity. Thus the canonical invariance of the fundamental Poisson brackets serves as both a necessary and a sufficient condition for a transformation to be canonical.

7-13. Prove directly that the Poisson bracket $[F, G]_{q,p}$ of any two functions $F(q, p)$ and $G(q, p)$ is a canonical invariant. Hint: Start by proving the statement for the fundamental brackets and then use these as building blocks for the general proof.

7-14. Show that the action variables J_k defined in (7.123) also have the property of being *adiabatic invariants*. This means that if the J_k are constructed for a Hamiltonian $H(q, p, \varepsilon)$ that depends on some time-dependent external parameter $\varepsilon(t)$, then the J_k remain constant for slow changes in $\varepsilon(t)$ (slow means $\dot{\varepsilon} \ll \varepsilon/T$, where T is a typical period of motion of the system if ε is constant). Start off your proof by observing that the generating function W inducing the transformation $(q_k, p_k) \rightarrow (\psi_k, J_k)$ will depend on $\varepsilon(t)$ in this case, and thus become time-dependent. Therefore, the new Hamiltonian will be given by

$$K = E(J_1, \dots, J_n) + \dot{\varepsilon} \frac{\partial W}{\partial \varepsilon}, \quad (7.229)$$

where $W = W(q, J, \varepsilon) = \sum_{i=1}^n W_i(q_i, J, \varepsilon)$. Now write down the equation of motion for J_k and average over a time large relative to T but small relative to time over which $\varepsilon(t)$ changes appreciably. Complete the proof by showing that $\partial^2 W / \partial \psi_k \partial \varepsilon$ is a one-valued function of the q_k and thus expressible as a periodic function in the ψ_k . The time-average of such a function vanishes.

The Lagrange formalism forms the basis of modern theoretical physics. Its usefulness in solving mechanical problems in particular lies in the fact that it is based on a scalar quantity, the so-called *action*, and therefore offers a vector-free formulation. The realm of applications is myriad. In this book we concentrate on its usage in classical mechanics, starting with basic principles, discussing particle dynamics, then rigid bodies. Thereafter, we devote a full chapter to small oscillations. We touch on fluid motion and the theory of relativity, before, in the final chapter, elucidating the connection to Hamilton's theory. Each chapter is more or less self-contained. We have presented many example problems, but also include a list of (unsolved) problems at the end of each chapter.

About the Authors

Sandra P. Klevansky is a retired Director of the Heidelberg Graduate School for Physics and Department of Physics and Astronomy, and adjunct Professor for Theoretical Physics at Heidelberg University. Her interests have ranged from theoretical nuclear and particle physics to non-Hermitian formulations of quantum field theory.

Richard H. Lemmer (deceased) was Professor of Mathematical Physics at the University of the Witwatersrand in Johannesburg. His international career had many stations, one of the longer being 10 years at the Massachusetts Institute of Technology (MIT), USA, before he returned to his country of birth, South Africa. He has made many significant contributions to nuclear and particle physics.

