Newtonian Dynamics

Richard Fitzpatrick

Professor of Physics

The University of Texas at Austin

Contents

1 Introduction

1.1 Intended Audience

This book presents a single semester course on Newtonian dynamics that is intended primarily for upper-division (*i.e.*, junior and senior) undergraduate students majoring in physics. A thorough understanding of physics at the lower-division level, including a basic working knowledge of the laws of mechanics, is assumed. It is also taken for granted that students are familiar with the fundamentals of integral and differential calculus, complex analysis, ordinary differential equations, and linear algebra. On the other hand, vector analysis plays such a central role in the study of Newtonian dynamics that a brief, but fairly comprehensive, review of this subject area is provided in Appendix A. Likewise, those results in matrix eigenvalue theory which are helpful in the analysis of rigid body motion and coupled oscillations are discussed in Sections 8.5 and 11.4, respectively. Finally, the calculus of variations, an area of mathematics that is central to Hamiltonian dynamics, is outlined in Section 10.2.

1.2 Scope of Book

The scope of this book is indicated by its title, "Newtonian Dynamics". "Dynamics" is the study of the *motions* of the various objects in the world around us. Furthermore, by "Newtonian", we understand that the theory which we are actually going to employ in our investigation of dynamics is that which was first formulated by Sir Isaac Newton in 1687.

For the sake of simplicity, and brevity, we shall restrict our investigations to the motions of idealized *point particles* and idealized *rigid bodies*. To be more exact, we shall exclude from consideration any discussion of statics, the strength of materials, and the non-rigid motions of continuous media. We shall also concentrate, for the most part, on motions which take place under the influence of *conservative forces*, such as gravity, which can be accurately represented in terms of simple mathematical formulae. Finally, with one major exception, we shall only consider that subset of dynamical problems that can be solved by means of conventional mathematical analysis.

Newtonian dynamics was originally developed in order to predict the motions of the objects which make up the *Solar System*. It turns out that this is an ideal application of the theory, since the objects in question can be modeled as being rigid to a fair degree of accuracy, and the motions take place under the action of a single conservative force—namely, gravity—that has a simple mathematical form. In particular, the frictional forces which greatly complicate the application of Newtonian dynamics to the motions of everyday objects close to the Earth's surface are completely absent. Consequently, in this book we shall make a particular effort to describe how Newtonian dynamics can successfully account for a wide variety of different solar system phenomena. For example, during the course of this

book, we shall explain the origins of Kepler's laws of planetary motion (see Chapter 5), the rotational flattening of the Earth, the tides, the Roche radius (*i.e.*, the minimum radius at which a moon can orbit a planet without being destroyed by tidal forces), the forced precession and nutation of the Earth's axis of rotation, and the forced perihelion precession of the planets (see Chapter 12). We shall also derive the Tisserand criterion used to re-identify comets whose orbits have been modified by close encounters with massive planets, account for the existence of the so-called Trojan asteroids which share the orbit of Jupiter (see Chapter 13), and analyze the motion of the Moon (see Chapter 14).

Virtually all of the results described in this book were first obtained—either by Newton himself, or by scientists living in the 150, or so, years immediately following the initial publication of his theory—by means of conventional mathematical analysis. Indeed, scientists at the beginning of the 20th century generally assumed that they knew everything that there was to known about Newtonian dynamics. However, they were mistaken. The advent of fast electronic computers, in the latter half of the 20th century, allowed scientists to solve *nonlinear* equations of motion, for the first time, via numerical techniques. In general, such equations are insoluble using standard analytic methods. The numerical investigation of dynamical systems with nonlinear equations of motion revealed the existence of a previously unknown type of motion known as *deterministic chaos*. Such motion is quasi-random (despite being derived from deterministic equations of motion), aperiodic, and exhibits extreme sensitivity to initial conditions. The discovery of chaotic motion lead to a renaissance in the study of Newtonian dynamics which started in the late 20th century and is still ongoing. It is therefore appropriate that the last chapter in this book is devoted to an in-depth numerical investigation of a particular dynamical system that exhibits chaotic motion (see Chapter 15).

1.3 Major Sources

The material appearing in Appendix A is largely based on the author's recollections of a vector analysis course given by Dr. Stephen Gull at the University of Cambridge. Major sources for the material appearing in Chapters 2–13 include *Dynamics*, H. Lamb, 2nd Edition (Cambridge University Press, Cambridge UK, 1923); *Analytical Mechanics*, G.R. Fowles (Holt, Rinehart, and Winston, New York NY, 1977); *Solar System Dynamics*, C.D. Murray, and S.F. Dermott (Cambridge University Press, Cambridge UK, 1999); *Classical Mechanics*, 3rd Edition, H. Goldstein, C. Poole, and J. Safko (Addison-Wesley, San Fransisco CA, 2002); *Classical Dynamics of Particles and Systems*, 5th Edition, S.T. Thornton, and J.B. Marion (Brooks/Cole—Thomson Learning, Belmont CA, 2004); and *Analytical Mechanics*, 7th Edition, G.R. Fowles, and G.L. Cassiday (Brooks/Cole—Thomson Learning, Belmont CA, 2005). The various sources for the material appearing in Chapters 14 and 15 are identified in footnotes.

2 Newton's Laws of Motion

2.1 Introduction

This chapter discusses Newton's laws of motion.

2.2 Newtonian Dynamics

Newtonian dynamics is a mathematical model whose purpose is to predict the motions of the various objects that we encounter in the world around us. The general principles of this model were first enunciated by Sir Isaac Newton in a work entitled *Philosophiae Naturalis Principia Mathematica* (Mathematical Principles of Natural Philosophy). This work, which was first published in 1687, is nowadays more commonly referred to as the *Principa*.[1]

Up until the beginning of the 20th century, Newton's theory of motion was thought to constitute a *complete* description of all types of motion occurring in the Universe. We now know that this is not the case. The modern view is that Newton's theory is only an *approximation* which is valid under certain circumstances. The theory breaks down when the velocities of the objects under investigation approach the speed of light in vacuum, and must be modified in accordance with Einstein's *special theory of relativity*. The theory also fails in regions of space which are sufficiently curved that the propositions of Euclidean geometry do not hold to a good approximation, and must be augmented by Einstein's *general theory of relativity*. Finally, the theory breaks down on atomic and subatomic length-scales, and must be replaced by *quantum mechanics*. In this book, we shall neglect relativistic and quantum effects all together. It follows that we must restrict our investigations to the motions of *large* (compared to an atom) *slow* (compared to the speed of light) objects moving in *Euclidean* space. Fortunately, virtually all of the motions which we commonly observe in the world around us fall into this category.

Newton very deliberately modeled his approach in the Principia on that taken in *Euclid's Elements*. Indeed, Newton's theory of motion has much in common with a conventional *axiomatic system* such as Euclidean geometry. Like all such systems, Newtonian dynamics starts from a set of terms that are *undefined* within the system. In this case, the fundamental terms are *mass*, *position*, *time*, and *force*. It is taken for granted that we understand what these terms mean, and, furthermore, that they correspond to *measurable* quantities which can be ascribed to, or associated with, objects in the world around us. In particular, it is assumed that the ideas of position in space, distance in space, and position as a function of time in space, are correctly described by the Euclidean vector algebra and

[1] An excellent discussion of the historical development of Newtonian dynamics, as well as the physical and philosophical assumptions which underpin this theory, is given in *The Discovery of Dynamics: A Study from a Machian Point of View of the Discovery and the Structure of Dynamical Theories*, J.B. Barbour (Oxford University Press, Oxford UK, 2001).

vector calculus discussed in Appendix A. The next component of an axiomatic system is a set of *axioms*. These are a set of *unproven* propositions, involving the undefined terms, from which all other propositions in the system can be derived via logic and mathematical analysis. In the present case, the axioms are called *Newton's laws of motion*, and can only be justified via experimental observation. Note, incidentally, that Newton's laws, in their primitive form, are only applicable to *point objects*. However, these laws can be applied to extended objects by treating them as collections of point objects.

One difference between an axiomatic system and a physical theory is that, in the latter case, even if a given prediction has been shown to follow necessarily from the axioms of the theory, it is still incumbent upon us to test the prediction against experimental observations. Lack of agreement might indicate faulty experimental data, faulty application of the theory (for instance, in the case of Newtonian dynamics, there might be forces at work which we have not identified), or, as a last resort, incorrectness of the theory. Fortunately, Newtonian dynamics has been found to give predictions which are in excellent agreement with experimental observations in all situations in which it is expected to be valid.

In the following, it is assumed that we know how to set up a rigid *Cartesian frame of reference*, and how to measure the positions of point objects as functions of time within that frame. It is also taken for granted that we have some basic familiarity with the laws of mechanics, and standard mathematics up to, and including, calculus, as well as the vector analysis outlined in Appendix A.

2.3 Newton's Laws of Motion

Newton's laws of motion, in the rather obscure language of the Principia, take the following form:

1. Every body continues in its state of rest, or uniform motion in a straight-line, unless compelled to change that state by forces impressed upon it.

2. The change of motion (*i.e.*, momentum) of an object is proportional to the force impressed upon it, and is made in the direction of the straight-line in which the force is impressed.

3. To every action there is always opposed an equal reaction; or, the mutual actions of two bodies upon each other are always equal and directed to contrary parts.

Let us now examine how these laws can be applied to dynamical systems.

2.4 Newton's First Law of Motion

Newton's first law of motion essentially states that a point object subject to zero net external force moves in a straight-line with a constant speed (*i.e.*, it does not accelerate). However, this is only true in special frames of reference called *inertial frames*. Indeed, we

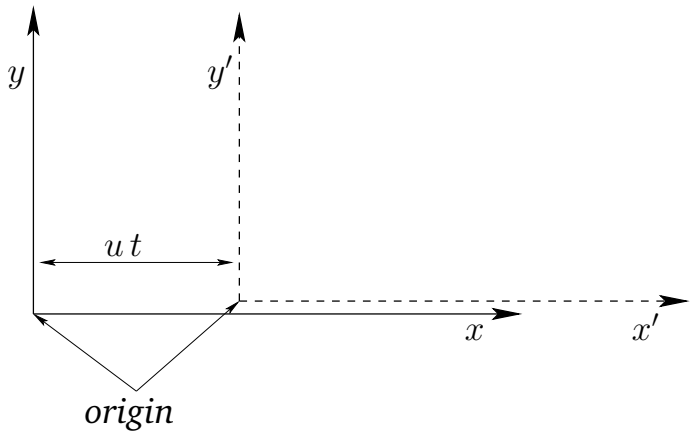


Figure 2.1: *A Galilean coordinate transformation.*

can think of Newton's first law as the *definition* of an inertial frame: *i.e.*, an inertial frame of reference is one in which a point object subject to zero net external force moves in a straight-line with constant speed.

Suppose that we have found an inertial frame of reference. Let us set up a Cartesian coordinate system in this frame. The motion of a point object can now be specified by giving its position vector, $\mathbf{r} \equiv (x, y, z)$, with respect to the origin of the coordinate system, as a function of time, t. Consider a second frame of reference moving with some *constant* velocity $\mathbf{u}$ with respect to the first frame. Without loss of generality, we can suppose that the Cartesian axes in the second frame are parallel to the corresponding axes in the first frame, that $\mathbf{u} \equiv (u, 0, 0)$, and, finally, that the origins of the two frames instantaneously coincide at $t = 0$—see Figure 2.1. Suppose that the position vector of our point object is $\mathbf{r}' \equiv (x', y', z')$ in the second frame of reference. It is evident, from Figure 2.1, that at any given time, t, the coordinates of the object in the two reference frames satisfy

$$x' = x - u\,t, \tag{2.1}$$

$$y' = y, \tag{2.2}$$

$$z' = z. \tag{2.3}$$

This coordinate transformation was first discovered by Galileo Galilei, and is known as a *Galilean transformation* in his honor.

By definition, the instantaneous velocity of the object in our first reference frame is given by $\mathbf{v} = d\mathbf{r}/dt \equiv (dx/dt, dy/dt, dz/dt)$, with an analogous expression for the velocity, $\mathbf{v}'$, in the second frame. It follows from differentiation of Equations (2.1)–(2.3) with respect to time that the velocity components in the two frames satisfy

$$v_x' = v_x - u, \tag{2.4}$$

$$v_y' = v_y, \tag{2.5}$$

$$v_z' = v_z. \tag{2.6}$$

These equations can be written more succinctly as

$$\mathbf{v}' = \mathbf{v} - \mathbf{u}. \tag{2.7}$$

Finally, by definition, the instantaneous acceleration of the object in our first reference frame is given by $\mathbf{a} = d\mathbf{v}/dt \equiv (dv_x/dt,\ dv_y/dt,\ dv_z/dt)$, with an analogous expression for the acceleration, $\mathbf{a}'$, in the second frame. It follows from differentiation of Equations (2.4)–(2.6) with respect to time that the acceleration components in the two frames satisfy

$$a'_x = a_x, \tag{2.8}$$

$$a'_y = a_y, \tag{2.9}$$

$$a'_z = a_z. \tag{2.10}$$

These equations can be written more succinctly as

$$\mathbf{a}' = \mathbf{a}. \tag{2.11}$$

According to Equations (2.7) and (2.11), if an object is moving in a straight-line with constant speed in our original inertial frame (*i.e.*, if $\mathbf{a} = \mathbf{0}$) then it also moves in a (different) straight-line with (a different) constant speed in the second frame of reference (*i.e.*, $\mathbf{a}' = \mathbf{0}$). Hence, we conclude that the second frame of reference is *also* an inertial frame.

A simple extension of the above argument allows us to conclude that there are an *infinite* number of different inertial frames moving with *constant velocities* with respect to one another. Newton thought that one of these inertial frames was special, and defined an absolute standard of rest: *i.e.*, a static object in this frame was in a state of absolute rest. However, Einstein showed that this is not the case. In fact, there is no absolute standard of rest: *i.e.*, all motion is relative—hence, the name "relativity" for Einstein's theory. Consequently, one inertial frame is just as good as another as far as Newtonian dynamics is concerned.

But, what happens if the second frame of reference *accelerates* with respect to the first? In this case, it is not hard to see that Equation (2.11) generalizes to

$$\mathbf{a}' = \mathbf{a} - \frac{d\mathbf{u}}{dt}, \tag{2.12}$$

where $\mathbf{u}(t)$ is the instantaneous velocity of the second frame with respect to the first. According to the above formula, if an object is moving in a straight-line with constant speed in the first frame (*i.e.*, if $\mathbf{a} = \mathbf{0}$) then it does not move in a straight-line with constant speed in the second frame (*i.e.*, $\mathbf{a}' \neq \mathbf{0}$). Hence, if the first frame is an inertial frame then the second is *not*.

A simple extension of the above argument allows us to conclude that any frame of reference which accelerates with respect to a given inertial frame is not itself an inertial frame.

For most practical purposes, when studying the motions of objects close to the Earth's surface, a reference frame which is fixed with respect to this surface is approximately

inertial. However, if the trajectory of a projectile within such a frame is measured to high precision then it will be found to deviate slightly from the predictions of Newtonian dynamics—see Chapter 7. This deviation is due to the fact that the Earth is rotating, and its surface is therefore accelerating towards its axis of rotation. When studying the motions of objects in orbit around the Earth, a reference frame whose origin is the center of the Earth, and whose coordinate axes are fixed with respect to distant stars, is approximately inertial. However, if such orbits are measured to extremely high precision then they will again be found to deviate very slightly from the predictions of Newtonian dynamics. In this case, the deviation is due to the Earth's orbital motion about the Sun. When studying the orbits of the planets in the Solar System, a reference frame whose origin is the center of the Sun, and whose coordinate axes are fixed with respect to distant stars, is approximately inertial. In this case, any deviations of the orbits from the predictions of Newtonian dynamics due to the orbital motion of the Sun about the galactic center are far too small to be measurable. It should be noted that it is impossible to identify an *absolute* inertial frame—the best approximation to such a frame would be one in which the cosmic microwave background appears to be (approximately) isotropic. However, for a given dynamical problem, it is always possible to identify an *approximate* inertial frame. Furthermore, any deviations of such a frame from a true inertial frame can be incorporated into the framework of Newtonian dynamics via the introduction of so-called fictitious forces—see Chapter 7.

2.5 Newton's Second Law of Motion

Newton's second law of motion essentially states that if a point object is subject to an external force, $\mathbf{f}$, then its equation of motion is given by

$$\frac{d\mathbf{p}}{dt} = \mathbf{f}, \tag{2.13}$$

where the momentum, $\mathbf{p}$, is the product of the object's inertial mass, m, and its velocity, $\mathbf{v}$. If m is not a function of time then the above expression reduces to the familiar equation

$$m\,\frac{d\mathbf{v}}{dt} = \mathbf{f}. \tag{2.14}$$

Note that this equation is only valid in a *inertial frame*. Clearly, the inertial mass of an object measures its reluctance to deviate from its preferred state of uniform motion in a straight-line (in an inertial frame). Of course, the above equation of motion can only be solved if we have an independent expression for the force, $\mathbf{f}$ (*i.e.*, a law of force). Let us suppose that this is the case.

An important corollary of Newton's second law is that force is a *vector quantity*. This must be the case, since the law equates force to the product of a scalar (mass) and a vector (acceleration). Note that acceleration is obviously a vector because it is directly related to displacement, which is the prototype of all vectors—see Appendix A. One consequence of force being a vector is that two forces, $\mathbf{f}_1$ and $\mathbf{f}_2$, both acting at a given point, have the

same effect as a single force, $\mathbf{f} = \mathbf{f}_1 + \mathbf{f}_2$, acting at the same point, where the summation is performed according to the laws of vector addition—see Appendix A. Likewise, a single force, $\mathbf{f}$, acting at a given point, has the same effect as two forces, $\mathbf{f}_1$ and $\mathbf{f}_2$, acting at the same point, provided that $\mathbf{f}_1 + \mathbf{f}_2 = \mathbf{f}$. This method of combining and splitting forces is known as the *resolution of forces*, and lies at the heart of many calculations in Newtonian dynamics.

Taking the scalar product of Equation (2.14) with the velocity, $\mathbf{v}$, we obtain

$$m\,\mathbf{v}\cdot\frac{d\mathbf{v}}{dt} = \frac{m}{2}\frac{d(\mathbf{v}\cdot\mathbf{v})}{dt} = \frac{m}{2}\frac{dv^2}{dt} = \mathbf{f}\cdot\mathbf{v}. \tag{2.15}$$

This can be written

$$\frac{dK}{dt} = \mathbf{f}\cdot\mathbf{v}. \tag{2.16}$$

where

$$K = \frac{1}{2}\,m\,v^2. \tag{2.17}$$

The right-hand side of Equation (2.16) represents the rate at which the force does work on the object: *i.e.*, the rate at which the force transfers energy to the object. The quantity K represents the energy the object possesses by virtue of its motion. This type of energy is generally known as *kinetic energy*. Thus, Equation (2.16) states that any work done on point object by an external force goes to increase the object's kinetic energy.

Suppose that, under the action of the force, $\mathbf{f}$, our object moves from point P at time t_1 to point Q at time t_2. The net change in the object's kinetic energy is obtained by integrating Equation (2.16):

$$\Delta K = \int_{t_1}^{t_2} \mathbf{f}\cdot\mathbf{v}\,dt = \int_P^Q \mathbf{f}\cdot d\mathbf{r}, \tag{2.18}$$

since $\mathbf{v} = d\mathbf{r}/dt$. Here, $d\mathbf{r}$ is an element of the object's path between points P and Q, and the integral in $\mathbf{r}$ represents the net work done by the force as the objects moves along the path from P to Q.

As described in Section A.15, there are basically two kinds of forces in nature. Firstly, those for which line integrals of the type $\int_P^Q \mathbf{f}\cdot d\mathbf{r}$ depend on the end points, but not on the path taken between these points. Secondly, those for which line integrals of the type $\int_P^Q \mathbf{f}\cdot d\mathbf{r}$ depend both on the end points, and the path taken between these points. The first kind of force is termed *conservative*, whereas the second kind is termed *non-conservative*. It was also demonstrated in Section A.15 that if the line integral $\int_P^Q \mathbf{f}\cdot d\mathbf{r}$ is *path independent* then the force $\mathbf{f}$ can always be written as the gradient of a scalar field. In other words, all conservative forces satisfy

$$\mathbf{f} = -\nabla U, \tag{2.19}$$

for some scalar field $U(\mathbf{r})$. Note that

$$\int_P^Q \nabla U\cdot d\mathbf{r} \equiv \Delta U = U(Q) - U(P), \tag{2.20}$$

irrespective of the path taken between P and Q. Hence, it follows from Equation (2.18) that

$$\Delta K = -\Delta U \tag{2.21}$$

for conservative forces. Another way of writing this is

$$E = K + U = \text{constant}. \tag{2.22}$$

Of course, we recognize this as an *energy conservation equation*: E is the object's total energy, which is conserved; K is the energy the object has by virtue of its motion, otherwise know as its *kinetic energy*; and U is the energy the object has by virtue of its position, otherwise known as its *potential energy*. Note, however, that we can only write such energy conservation equations for conservative forces. Gravity is a good example of a conservative force. Non-conservative forces, on the other hand, do not conserve energy. In general, this is because of some sort of frictional energy loss which drains energy from the dynamical system whilst it remains in motion. Note that potential energy is undefined to an arbitrary additive constant. In fact, it is only the *difference* in potential energy between different points in space that is well-defined.

2.6 Newton's Third Law of Motion

Consider a system of N mutually interacting point objects. Let the ith object, whose mass is m_i, be located at vector displacement $\mathbf{r}_i$. Suppose that this object exerts a force $\mathbf{f}_{ji}$ on the jth object. Likewise, suppose that the jth object exerts a force $\mathbf{f}_{ij}$ on the ith object. Newton's third law of motion essentially states that these two forces are equal and opposite, irrespective of their nature. In other words,

$$\mathbf{f}_{ij} = -\mathbf{f}_{ji}. \tag{2.23}$$

One corollary of Newton's third law is that an object cannot exert a force on itself. Another corollary is that all forces in the Universe have corresponding reactions. The only exceptions to this rule are the fictitious forces which arise in non-inertial reference frames (*e.g.*, the centrifugal and Coriolis forces which appear in rotating reference frames—see Chapter 7). Fictitious forces do not possess reactions.

It should be noted that Newton's third law implies *action at a distance*. In other words, if the force that object i exerts on object j suddenly changes then Newton's third law demands that there must be an *immediate* change in the force that object j exerts on object i. Moreover, this must be true irrespective of the distance between the two objects. However, we now know that Einstein's theory of relativity forbids information from traveling through the Universe faster than the velocity of light in vacuum. Hence, action at a distance is also forbidden. In other words, if the force that object i exerts on object j suddenly changes then there must be a *time delay*, which is at least as long as it takes a light ray to propagate between the two objects, before the force that object j exerts on object i can respond. Of course, this means that Newton's third law is not, strictly speaking, correct.

However, as long as we restrict our investigations to the motions of dynamical systems on time-scales that are long compared to the time required for light-rays to traverse these systems, Newton's third law can be regarded as being approximately correct.

In an inertial frame, Newton's second law of motion applied to the ith object yields

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \sum_{j=1,N}^{j\neq i} \mathbf{f}_{ij}. \tag{2.24}$$

Note that the summation on the right-hand side of the above equation excludes the case $j = i$, since the ith object cannot exert a force on itself. Let us now take the above equation and sum it over all objects. We obtain

$$\sum_{i=1,N} m_i \frac{d^2\mathbf{r}_i}{dt^2} = \sum_{i,j=1,N}^{j\neq i} \mathbf{f}_{ij}. \tag{2.25}$$

Consider the sum over forces on the right-hand side of the above equation. Each element of this sum—$\mathbf{f}_{ij}$, say—can be paired with another element—$\mathbf{f}_{ji}$, in this case—which is equal and opposite, according to Newton's third law. In other words, the elements of the sum all cancel out in pairs. Thus, the net value of the sum is *zero*. It follows that the above equation can be written

$$M \frac{d^2\mathbf{r}_{cm}}{dt^2} = \mathbf{0}, \tag{2.26}$$

where $M = \sum_{i=1}^{N} m_i$ is the total mass. The quantity $\mathbf{r}_{cm}$ is the vector displacement of the *center of mass* of the system, which is an imaginary point whose coordinates are the mass weighted averages of the coordinates of the objects that constitute the system: *i.e.*,

$$\mathbf{r}_{cm} = \frac{\sum_{i=1}^{N} m_i \mathbf{r}_i}{\sum_{i=1}^{N} m_i}. \tag{2.27}$$

According to Equation (2.26), the center of mass of the system moves in a uniform straight-line, in accordance with Newton's first law of motion, irrespective of the nature of the forces acting between the various components of the system.

Now, if the center of mass moves in a uniform straight-line then the center of mass velocity,

$$\frac{d\mathbf{r}_{cm}}{dt} = \frac{\sum_{i=1}^{N} m_i \, d\mathbf{r}_i/dt}{\sum_{i=1}^{N} m_i}, \tag{2.28}$$

is a constant of the motion. However, the momentum of the ith object takes the form $\mathbf{p}_i = m_i \, d\mathbf{r}_i/dt$. Hence, the total momentum of the system is written

$$\mathbf{P} = \sum_{i=1}^{N} m_i \frac{d\mathbf{r}_i}{dt}. \tag{2.29}$$

A comparison of Equations (2.28) and (2.29) suggests that $\mathbf{P}$ is also a constant of the motion. In other words, the total momentum of the system is a *conserved* quantity, irrespective of the nature of the forces acting between the various components of the system. This result (which only holds if there is no net external force acting on the system) is a direct consequence of Newton's third law of motion.

Taking the vector product of Equation (2.24) with the position vector $\mathbf{r}_i$, we obtain

$$m_i\,\mathbf{r}_i\times\frac{d^2\mathbf{r}_i}{dt^2}=\sum_{j=1,N}^{j\neq i}\mathbf{r}_i\times\mathbf{f}_{ij}. \tag{2.30}$$

However, it is easily seen that

$$m_i\,\mathbf{r}_i\times\frac{d^2\mathbf{r}_i}{dt^2}=\frac{d}{dt}\left(m_i\,\mathbf{r}_i\times\frac{d\mathbf{r}_i}{dt}\right)=\frac{d\mathbf{l}_i}{dt}, \tag{2.31}$$

where

$$\mathbf{l}_i=m_i\,\mathbf{r}_i\times\frac{d\mathbf{r}_i}{dt} \tag{2.32}$$

is the *angular momentum* of the ith particle about the origin of our coordinate system. The total angular momentum of the system (about the origin) takes the form

$$\mathbf{L}=\sum_{i=1,N}\mathbf{l}_i \tag{2.33}$$

Hence, summing Equation (2.30) over all particles, we obtain

$$\frac{d\mathbf{L}}{dt}=\sum_{i,j=1,N}^{i\neq j}\mathbf{r}_i\times\mathbf{f}_{ij}. \tag{2.34}$$

Consider the sum on the right-hand side of the above equation. A general term, $\mathbf{r}_i\times\mathbf{f}_{ij}$, in this sum can always be paired with a matching term, $\mathbf{r}_j\times\mathbf{f}_{ji}$, in which the indices have been swapped. Making use of Equation (2.23), the sum of a general matched pair can be written

$$\mathbf{r}_i\times\mathbf{f}_{ij}+\mathbf{r}_j\times\mathbf{f}_{ji}=(\mathbf{r}_i-\mathbf{r}_j)\times\mathbf{f}_{ij}. \tag{2.35}$$

Let us assume that the forces acting between the various components of the system are *central* in nature, so that $\mathbf{f}_{ij}$ is parallel to $\mathbf{r}_i-\mathbf{r}_j$. In other words, the force exerted on object j by object i either points directly toward, or directly away from, object i, and *vice versa*. This is a reasonable assumption, since most of the forces which we encounter in the world around us are of this type (*e.g.*, gravity). It follows that if the forces are central in nature then the vector product on the right-hand side of the above expression is zero. We conclude that

$$\mathbf{r}_i\times\mathbf{f}_{ij}+\mathbf{r}_j\times\mathbf{f}_{ji}=\mathbf{0}, \tag{2.36}$$

for all values of i and j. Thus, the sum on the right-hand side of Equation (2.34) is zero for any kind of central force. We are left with

$$\frac{d\mathbf{L}}{dt} = \mathbf{0}. \tag{2.37}$$

In other words, the total angular momentum of the system is a *conserved* quantity, provided that the different components of the system interact via *central* forces (and there is no net external torque acting on the system).

2.7 Non-Isolated Systems

Up to now, we have only considered *isolated* dynamical systems, in which all of the forces acting on the system originate within the system itself. Let us now generalize our approach to deal with *non-isolated* dynamical systems, in which some of the forces originate outside the system. Consider a system of N mutually interacting point objects. Let m_i and $\mathbf{r}_i$ be the mass and position vector of the ith object, respectively. Suppose that the ith object is subject to two forces. First, an *internal force* which originates from the other objects in the system, and second an *external force* which originates outside the system. In other words, let the force acting on the ith object take the form

$$\mathbf{f}_i = \sum_{j=1,N}^{j\neq i} \mathbf{f}_{ij} + \mathbf{F}_i, \tag{2.38}$$

where $\mathbf{f}_{ij}$ is the internal force exerted by object j on object i, and $\mathbf{F}_i$ the net external force acting on object i.

The equation of motion of the ith object is

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{f}_i = \sum_{j=1,N}^{j\neq i} \mathbf{f}_{ij} + \mathbf{F}_i. \tag{2.39}$$

Summing over all objects, we obtain

$$\sum_{i=1,N} m_i \frac{d^2\mathbf{r}_i}{dt^2} = \sum_{i,j=1,N}^{j\neq i} \mathbf{f}_{ij} + \sum_{i=1,N} \mathbf{F}_i, \tag{2.40}$$

which reduces to

$$\frac{d\mathbf{P}}{dt} = \mathbf{F}, \tag{2.41}$$

where

$$\mathbf{F} = \sum_{i=1,N} \mathbf{F}_i \tag{2.42}$$

is the net external force acting on the system. Here, the sum over the internal forces has cancelled out in pairs due to Newton's third law of motion. We conclude that the total system momentum evolves in time according to the simple equation (2.41) when there is a net external force acting on the system, but is completely unaffected by the internal forces. The fact that Equation (2.41) is similar in form to Equation (2.13) suggests that the center of mass of a system of many point objects has analogous dynamics to a point object.

Taking $\mathbf{r}_i\times$ Equation (2.39), and summing over all objects, we obtain

$$\frac{d\mathbf{L}}{dt} = \mathbf{T}, \tag{2.43}$$

where

$$\mathbf{T} = \sum_{i=1,N} \mathbf{r}_i \times \mathbf{F}_i \tag{2.44}$$

is the net external torque acting on the system. Here, the sum over the internal torques has cancelled out in pairs, assuming that the internal forces are central in nature. We conclude that the total system angular momentum evolves in time according to the simple equation (2.43) when there is a net external torque acting on the system, but is completely unaffected by the internal torques.

2.8 Exercises

2.1. Consider an isolated system of N point objects interacting via gravity. Let the mass and position vector of the ith object be m_i and $\mathbf{r}_i$, respectively. What is the vector equation of motion of the ith object? Write expressions for the total kinetic energy, K, and potential energy, U, of the system. Demonstrate from the equations of motion that $K+U$ is a conserved quantity.

2.2. Consider a function of many variables $f(x_1, x_2, \cdots, x_n)$. Such a function which satisfies

$$f(t\,x_1, t\,x_2, \cdots, t\,x_n) = t^a\, f(x_1, x_2, \cdots, x_n)$$

for all $t > 0$, and all values of the x_i, is termed a *homogenous function of degree* a. Prove the following theorem regarding homogeneous functions:

$$\sum_{i=1,n} x_i \frac{\partial f}{\partial x_i} = a\, f$$

2.3. Consider an isolated system of N point objects interacting via attractive central forces. Let the mass and position vector of the ith object be m_i and $\mathbf{r}_i$, respectively. Suppose that magnitude of the force exerted on object i by object j is $k_i\,k_j\,|\mathbf{r}_i - \mathbf{r}_j|^{-n}$. Here, the k_i measure some constant physical property of the particles (*e.g.*, their electric charges). Write an expression for the total potential energy U of the system. Is this a homogenous function? If so, what is

its degree? Write the equation of motion of the ith particle. Use the mathematical theorem from the previous exercise to demonstrate that

$$\frac{1}{2}\frac{d^2I}{dt^2} = 2K + (n-1)U,$$

where $I = \sum_{i=1,N} m_i r_i^2$, and K is the kinetic energy. This result is known as the *virial theorem*. Demonstrate that there are no bound steady-state equilibria for the system (*i.e.*, states in which the global system parameters do not evolve in time) when $n \geq 3$.

2.4. A star can be through of as a spherical system that consists of a very large number of particles interacting via gravity. Show that, for such a system, the virial theorem, introduced in the previous exercise, implies that

$$\frac{d^2I}{dt^2} = -2U + c,$$

where c is a constant, and the r_i are measured from the geometric center. Hence, deduce that the angular frequency of small amplitude radial pulsations of the star (in which the radial displacement is directly proportional to the radial distance from the center) takes the form

$$\omega = \left(\frac{|U_0|}{I_0}\right)^{1/2},$$

where U_0 and I_0 are the equilibrium values of U and I. Finally, show that if the mass density within the star varies as $r^{-\alpha}$, where r is the radial distance from the geometric center, and where $\alpha < 5/2$, then

$$\omega = \left(\frac{5-\alpha}{5-2\,\alpha}\,\frac{G\,M}{R^3}\right)^{1/2},$$

where M and R are the stellar mass and radius, respectively.

2.5. Consider a system of N point particles. Let the ith particle have mass m_i, electric charge q_i, and position vector $\mathbf{r}_i$. Suppose that the charge to mass ratio, q_i/m_i, is the same for all particles. The system is placed in a uniform magnetic field $\mathbf{B}$. Write the equation of motion of the ith particle. You may neglect any magnetic fields generated by the motion of the particles. Demonstrate that the total momentum $\mathbf{P}$ of the system precesses about $\mathbf{B}$ at the frequency $\Omega = q_i\,B/m_i$. Demonstrate that $L_\parallel + \Omega\,I_\parallel/2$ is a constant of the motion. Here, $L_\parallel$ is the total angular momentum of the system parallel to the magnetic field, and $I_\parallel$ is the moment of inertia of the system about an axis parallel to $\mathbf{B}$ which passes through the origin.

3 One-Dimensional Motion

3.1 Introduction

This chapter employs Newton's laws to investigate one-dimensional motion. Particular attention is given to the various mathematical techniques commonly used to analyze oscillatory motion.

3.2 Motion in a General One-Dimensional Potential

Consider a point particle of mass m moving in the x-direction, say, under the action of some x-directed force $f(x)$. Suppose that $f(x)$ is a conservative force: *e.g.*, gravity. In this case, according to Equation (2.19), we can write

$$f(x) = -\frac{dU(x)}{dx}, \tag{3.1}$$

where $U(x)$ is the potential energy of the particle at position x.

Let the curve $U(x)$ take the form shown in Figure 3.1. For instance, this curve might represent the gravitational potential energy of a cyclist freewheeling in a hilly region. Observe that we have set the potential energy at infinity to zero (which we are generally free to do, since potential energy is undefined to an arbitrary additive constant). This is a fairly common convention. What can we deduce about the motion of the particle in this potential?

Well, we know that the total energy, E—which is the sum of the kinetic energy, K, and the potential energy, U—is a *constant* of the motion—see Equation (2.22). Hence, we can write

$$K(x) = E - U(x). \tag{3.2}$$

However, we also know that a kinetic energy can never be negative [since $K = (1/2)\, m\, v^2$, and neither m nor v^2 can be negative]. Hence, the above expression tells us that the particle's motion is restricted to the region (or regions) in which the potential energy curve $U(x)$ falls below the value E. This idea is illustrated in Figure 3.1. Suppose that the total energy of the system is E_0. It is clear, from the figure, that the particle is trapped inside one or other of the two dips in the potential—these dips are generally referred to as *potential wells*. Suppose that we now raise the energy to E_1. In this case, the particle is free to enter or leave each of the potential wells, but its motion is still *bounded* to some extent, since it clearly cannot move off to infinity. Finally, let us raise the energy to E_2. Now the particle is *unbounded*: *i.e.*, it can move off to infinity. In conservative systems in which it makes sense to adopt the convention that the potential energy at infinity is zero, bounded systems are characterized by $E < 0$, whereas unbounded systems are characterized by $E > 0$.

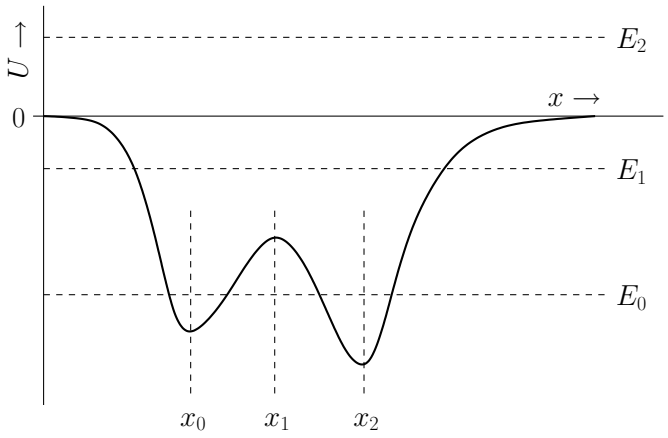


Figure 3.1: *A potential energy curve.*

The above discussion suggests that the motion of an particle moving in a potential generally becomes less bounded as the total energy E of the system increases. Conversely, we would expect the motion to become more bounded as E decreases. In fact, if the energy becomes sufficiently small then it appears likely that the system will settle down in some *equilibrium state* in which the particle is stationary. Let us try to identify any prospective equilibrium states in Figure 3.1. If the particle remains stationary then it must be subject to zero force (otherwise it would accelerate). Hence, according to Equation (3.1), an equilibrium state is characterized by

$$\frac{dU}{dx} = 0. \tag{3.3}$$

In other words, a equilibrium state corresponds to either a *maximum* or a *minimum* of the potential energy curve $U(x)$. It can be seen that the $U(x)$ curve shown in Figure 3.1 has three associated equilibrium states located at $x = x_0$, $x = x_1$, and $x = x_2$.

Let us now make a distinction between *stable* equilibrium points and *unstable* equilibrium points. When the particle is slightly displaced from a stable equilibrium point then the resultant force f acting on it must always be such as to return it to this point. In other words, if $x = x_0$ is an equilibrium point then we require

$$\left.\frac{df}{dx}\right|_{x_0} < 0 \tag{3.4}$$

for stability: *i.e.*, if the particle is displaced to the right, so that $x - x_0 > 0$, then the force must act to the left, so that $f < 0$, and *vice versa*. Likewise, if

$$\left.\frac{df}{dx}\right|_{x_0} > 0 \tag{3.5}$$

then the equilibrium point $x = x_0$ is unstable. It follows, from Equation (3.1), that stable

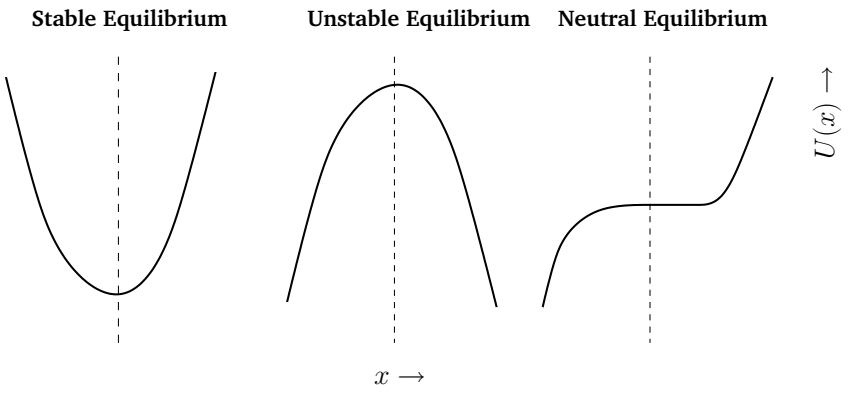


Figure 3.2: *Different types of equilibrium point.*

equilibrium points are characterized by

$$\frac{d^2U}{dx^2} > 0. \tag{3.6}$$

In other words, a stable equilibrium point corresponds to a *minimum* of the potential energy curve $U(x)$. Likewise, an unstable equilibrium point corresponds to a *maximum* of the $U(x)$ curve. Hence, we conclude that, in Figure 3.1, $x = x_0$ and $x = x_2$ are stable equilibrium points, whereas $x = x_1$ is an unstable equilibrium point. Of course, this makes perfect sense if we think of $U(x)$ as a gravitational potential energy curve, so that U is directly proportional to height. In this case, all we are saying is that it is easy to confine a low energy mass at the bottom of a valley, but very difficult to balance the same mass on the top of a hill (since any slight displacement of the mass will cause it to fall down the hill). Note, finally, that if

$$\frac{dU}{dx} = \frac{d^2U}{dx^2} = 0 \tag{3.7}$$

at any point (or in any region) then we have what is known as a *neutral equilibrium* point. We can move the particle slightly away from such a point and it will still remain in equilibrium (*i.e.*, it will neither attempt to return to its initial state, nor will it continue to move). A neutral equilibrium point corresponds to a *flat spot* in a $U(x)$ curve. See Figure 3.2.

The equation of motion of an particle moving in one dimension under the action of a conservative force is, in principle, integrable. Since $K = (1/2)\,m\,v^2$, the energy conservation equation (3.2) can be rearranged to give

$$v = \pm \left(\frac{2\,[E - U(x)]}{m} \right)^{1/2}, \tag{3.8}$$

where the $\pm$ signs correspond to motion to the left and to the right, respectively. However,

since $v = dx/dt$, this expression can be integrated to give

$$t = \pm \left(\frac{m}{2E}\right)^{1/2} \int_{x_0}^{x} \frac{dx'}{\sqrt{1 - U(x')/E}}, \tag{3.9}$$

where $x(t = 0) = x_0$. For sufficiently simple potential functions, $U(x)$, the above equation can be solved to give x as a function of t. For instance, if $U = (1/2)\,k\,x^2$, $x_0 = 0$, and the plus sign is chosen, then

$$t = \left(\frac{m}{k}\right)^{1/2} \int_0^{(k/2E)^{1/2}\,x} \frac{dy}{\sqrt{1 - y^2}} = \left(\frac{m}{k}\right)^{1/2} \sin^{-1}\left(\left[\frac{k}{2E}\right]^{1/2} x\right), \tag{3.10}$$

which can be inverted to give

$$x = a\,\sin(\omega\,t), \tag{3.11}$$

where $a = \sqrt{2E/k}$ and $\omega = \sqrt{k/m}$. Note that the particle reverses direction each time it reaches one of the so-called *turning points* ($x = \pm a$) at which $U = E$ (and, so $K = 0$).

3.3 Velocity Dependent Forces

Consider a particle of mass m moving in one dimension under the action of a force, f, which is a function of the particle's speed, v, but not of its displacement, x. Note that such a force is intrinsically non-conservative [since it clearly cannot be expressed as minus the gradient of some potential function, $U(x)$]. Now, the particle's equation of motion is written

$$m\,\frac{dv}{dt} = f(v). \tag{3.12}$$

Integrating this equation, we obtain

$$\int_{v_0}^{v} \frac{dv'}{f(v')} = \frac{t}{m}, \tag{3.13}$$

where $v(t = 0) = v_0$. In principle, the above equation can be solved to give $v(t)$. The equation of motion is also written

$$m\,v\,\frac{dv}{dx} = f(v), \tag{3.14}$$

since $v = dx/dt$. Integrating this equation, we obtain

$$\int_{v_0}^{v} \frac{v'\,dv'}{f(v')} = \frac{x - x_0}{m}, \tag{3.15}$$

where $x(t = 0) = x_0$. In principle, the above equation can be solved to give $v(x)$.

Let us now consider a specific example. Suppose that an object of mass m falls vertically under gravity. Let x be the height through which the object has fallen since $t = 0$, at which time the object is assumed to be at rest. It follows that $x_0 = v_0 = 0$. Suppose that, in addition to the force of gravity, $m\,g$, where g is the gravitational acceleration, our object is subject to a retarding air resistance force which is proportional to the square of its instantaneous velocity. The object's equation of motion is thus

$$m\frac{dv}{dt} = m\,g - c\,v^2, \tag{3.16}$$

where $c > 0$. This equation can be integrated to give

$$\int_0^v \frac{dv'}{1-(v'/v_t)^2} = g\,t, \tag{3.17}$$

where $v_t = (m\,g/c)^{1/2}$. Making a change of variable, we obtain

$$\int_0^{v/v_t} \frac{dy}{1-y^2} = \frac{g}{v_t}\,t. \tag{3.18}$$

The left-hand side of the above equation is now a standard integral, which can be solved to give

$$\tanh^{-1}\left(\frac{v}{v_t}\right) = \frac{g\,t}{v_t}, \tag{3.19}$$

or

$$v = v_t\,\tanh\left(\frac{g\,t}{v_t}\right). \tag{3.20}$$

Thus, when $t \ll v_t/g$, we obtain the standard result $v \simeq g\,t$, since $\tanh x \simeq x$ for $x \ll 1$. However, when $t \gg v_t/g$, we get $v \simeq v_t$, since $\tanh x \simeq 1$ for $x \gg 1$. It follows that air resistance prevents the downward velocity of our object from increasing indefinitely as it falls. Instead, at large times, the velocity asymptotically approaches the so-called *terminal velocity*, v_t (at which the gravitational and air resistance forces balance).

The equation of motion of our falling object is also written

$$m\,v\frac{dv}{dx} = m\,g - c\,v^2. \tag{3.21}$$

This equation can be integrated to give

$$\int_0^v \frac{v'\,dv'}{1-(v'/v_t)^2} = g\,x. \tag{3.22}$$

Making a change of variable, we obtain

$$\int_0^{(v/v_t)^2} \frac{dy}{1-y} = \frac{x}{x_t}, \tag{3.23}$$

where $x_t = m/(2\,c)$. The left-hand side of the above equation is now a standard integral, which can be solved to give

$$-\ln\left[1-\left(\frac{v}{v_t}\right)^2\right] = \frac{x}{x_t}, \tag{3.24}$$

or

$$v = v_t\left(1-e^{-x/x_t}\right)^{1/2}. \tag{3.25}$$

It follows that our object needs to fall a distance of order x_t before it achieves its terminal velocity.

Incidentally, it is quite easy to account for an air resistance force which scales as the square of projectile velocity. Let us imaging that our projectile is moving sufficiently rapidly that air does not have enough time to flow around it, and is instead simply knocked out of the way. If our projectile has cross-sectional area A, perpendicular to the direction of its motion, and is moving with speed v, then the mass of air that it knocks out of its way per second is $\rho_a\,A\,v$, where ρ_a is the mass density of air. Suppose that the air knocked out of the way is pushed in the direction of the projectile's motion with a speed of order v. It follows that the air gains momentum per unit time $\rho_a\,A\,v^2$ in the direction of the projectile's motion. Hence, by Newton's third law, the projectile loses the same momentum per unit time in the direction of its motion. In other words, the projectile is subject to a drag force of magnitude

$$f_{drag} = C\,\rho_a\,A\,v^2 \tag{3.26}$$

acting in the opposite direction to its motion. Here, C is an $\mathcal{O}(1)$ dimensionless constant, known as the *drag coefficient*, which depends on the exact shape of the projectile. Obviously, streamlined projectiles, such as arrows, have small drag coefficients, whereas non-streamlined projectiles, such as bricks, have large drag coefficients. From before, the terminal velocity of our projectile is $v_t = (m\,g/c)^{1/2}$, where m is its mass, and $c = C\,\rho_a\,A$. Writing $m = A\,d\,\rho$, where d is the typical linear dimension of the projectile, and ρ its mass density, we obtain

$$v_t = \left(\frac{\rho\,g\,d}{\rho_a\,C}\right)^{1/2}. \tag{3.27}$$

The above expression tells us that large, dense, streamlined projectiles (*e.g.*, medicine balls) tend to have large terminal velocities, and small, rarefied, non-streamlined projectiles (*e.g.*, feathers) tend to have small terminal velocities. Hence, the former type of projectile is relatively less affected by air resistance than the latter.

3.4 Simple Harmonic Motion

Consider the motion of a point particle of mass m which is slightly displaced from a stable equilibrium point at $x = 0$. Suppose that the particle is moving in the conservative force-field $f(x)$. According to the analysis of Section 3.2, in order for $x = 0$ to be a stable equilibrium point we require both

$$f(0) = 0, \tag{3.28}$$

and

$$\frac{df(0)}{dx} < 0. \tag{3.29}$$

Now, our particle obeys Newton's second law of motion,

$$m\,\frac{d^2x}{dt^2} = f(x). \tag{3.30}$$

Let us assume that it always stays fairly close to its equilibrium point. In this case, to a good approximation, we can represent $f(x)$ via a truncated Taylor expansion about this point. In other words,

$$f(x) \simeq f(0) + \frac{df(0)}{dx}\,x + \mathcal{O}(x^2). \tag{3.31}$$

However, according to (3.28) and (3.29), the above expression can be written

$$f(x) \simeq -m\,\omega_0^2\,x, \tag{3.32}$$

where $df(0)/dx = -m\,\omega_0^2$. Hence, we conclude that our particle satisfies the following approximate equation of motion,

$$\frac{d^2x}{dt^2} + \omega_0^2\,x \simeq 0, \tag{3.33}$$

provided that it does not stray too far from its equilibrium point: *i.e.*, provided $|x|$ does not become too large.

Equation (3.33) is called the *simple harmonic equation*, and governs the motion of *all* one-dimensional conservative systems which are slightly perturbed from some stable equilibrium state. The solution of Equation (3.33) is well-known:

$$x(t) = a\,\sin(\omega_0\,t - \phi_0). \tag{3.34}$$

The pattern of motion described by above expression, which is called *simple harmonic motion*, is *periodic* in time, with repetition period $T_0 = 2\pi/\omega_0$, and *oscillates* between $x = \pm a$. Here, a is called the *amplitude* of the motion. The parameter ϕ_0, known as the *phase angle*, simply shifts the pattern of motion backward and forward in time. Figure 3.3 shows some examples of simple harmonic motion. Here, $\phi_0 = 0$, $+\pi/4$, and $-\pi/4$ correspond to the solid, short-dashed, and long-dashed curves, respectively.

Note that the frequency, ω_0—and, hence, the period, T_0—of simple harmonic motion is determined by the parameters appearing in the simple harmonic equation, (3.33). However, the amplitude, a, and the phase angle, ϕ_0, are the two integration constants of this second-order ordinary differential equation, and are, thus, determined by the initial conditions: *i.e.*, by the particle's initial displacement and velocity.

Now, from Equations (3.1) and (3.32), the potential energy of our particle at position x is approximately

$$U(x) \simeq \frac{1}{2}\,m\,\omega_0^2\,x^2. \tag{3.35}$$

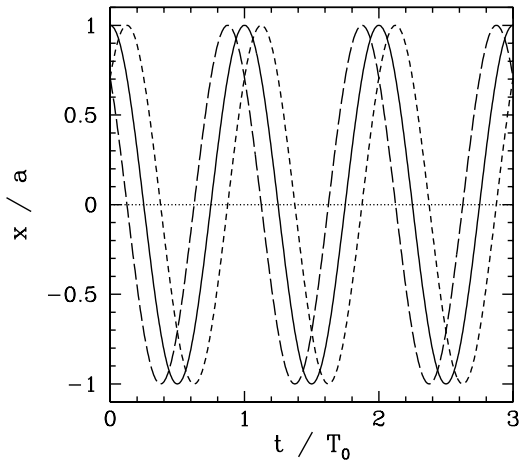


Figure 3.3: *Simple harmonic motion.*

Hence, the total energy is written

$$E = K + U = \frac{1}{2}\,m\left(\frac{dx}{dt}\right)^2 + \frac{1}{2}\,m\,\omega_0^2\,x^2, \tag{3.36}$$

giving

$$E = \frac{1}{2}\,m\,\omega_0^2\,a^2\,\cos^2(\omega_0\,t-\phi_0) + \frac{1}{2}\,m\,\omega_0^2\,a^2\,\sin^2(\omega_0\,t-\phi_0) = \frac{1}{2}\,m\,\omega_0^2\,a^2, \tag{3.37}$$

where use has been made of Equation (3.34), and the trigonometric identity $\cos^2\theta + \sin^2\theta \equiv 1$. Note that the total energy is *constant* in time, as is to be expected for a conservative system, and is proportional to the *amplitude squared* of the motion.

3.5 Damped Oscillatory Motion

According to Equation (3.34), a one-dimensional conservative system which is slightly perturbed from a stable equilibrium point (and then left alone) oscillates about this point with a fixed frequency and a constant amplitude. In other words, the oscillations never die away. This is not very realistic, since we know that, in practice, if we slightly perturb a dynamical system (such as a pendulum) from a stable equilibrium point then it will indeed oscillate about this point, but these oscillations will eventually die away due to frictional effects, which are present in virtually all real dynamical systems. In order to model this process, we need to include some sort of frictional drag force in our perturbed equation of motion, (3.33).

The most common model for a frictional drag force is one which is always directed in the *opposite direction* to the instantaneous velocity of the object upon which it acts, and

is *directly proportional* to the magnitude of this velocity. Let us adopt this model. So, our drag force can be written

$$f_{drag} = -2\,m\,\nu\,\frac{dx}{dt}, \tag{3.38}$$

where ν is a positive constant with the dimensions of frequency. Including such a force in our perturbed equation of motion, (3.33), we obtain

$$\frac{d^2x}{dt^2} + 2\,\nu\,\frac{dx}{dt} + \omega_0^2\,x = 0. \tag{3.39}$$

Thus, the positive constant ν parameterizes the strength of the frictional damping in our dynamical system.

Equation (3.39) is a linear second-order ordinary differential equation, which we suspect possesses oscillatory solutions. There is a standard trick for solving such an equation. We search for complex solutions of the form

$$x = a\,e^{-i\,\omega\,t}, \tag{3.40}$$

where the constants ω and a are both, in general, complex. Of course, the physical solution is the *real part* of the above expression: *i.e.*,

$$x = |a|\,\cos[\arg(a) - \mathrm{Re}(\omega)\,t]\,e^{\mathrm{Im}(\omega)\,t}. \tag{3.41}$$

Clearly, the modulus and argument of the complex amplitude, a, determine the amplitude (at $t = 0$) and phase of the oscillation, respectively, whereas the real and imaginary parts of the complex frequency, ω, determine its frequency and growth-rate, respectively. Note that this method of solution is only appropriate for *linear* differential equations. Incidentally, the method works because

$$\mathrm{Re}[\mathcal{L}(x)] \equiv \mathcal{L}(\mathrm{Re}[x]), \tag{3.42}$$

where x is a complex variable, and $\mathcal{L}$ some real linear differential operator which acts on this variable. [A linear operator satisfies $\mathcal{L}(a\,x) = a\,\mathcal{L}(x)$ for all a and x, where a is a constant. The differential operator appearing in Equation (3.39) is clearly of this type.]

Substituting Equation (3.40) into Equation (3.39), we obtain

$$a\left[-\omega^2 - i\,2\,\nu\,\omega + \omega_0^2\right] e^{-i\,\omega\,t} = 0, \tag{3.43}$$

which reduces to the following quadratic equation for ω [since $a\,\exp(-i\,\omega\,t) \neq 0$]:

$$\omega^2 + i\,2\,\nu\,\omega - \omega_0^2 = 0. \tag{3.44}$$

The solution to this equation is

$$\omega_\pm = -i\,\nu \pm \sqrt{\omega_0^2 - \nu^2}. \tag{3.45}$$

Thus, the most general physical solution to Equation (3.39) is

$$x(t) = \mathrm{Re}\left[a_+\,e^{-i\,\omega_+\,t} + a_-\,e^{-i\,\omega_-\,t}\right], \tag{3.46}$$

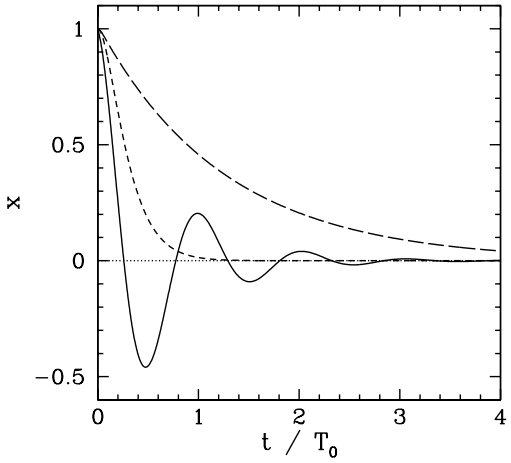


Figure 3.4: *Damped oscillatory motion.*

where $a_\pm$ are two arbitrary complex constants.

We can distinguish three different cases. In the first case, $\nu < \omega_0$, and the motion is said to be *underdamped*. The most general solution is written

$$x(t) = x_0\, e^{-\nu\, t}\, \cos(\omega_r\, t) + \left(\frac{\nu_0 + \nu\, x_0}{\omega_r}\right) e^{-\nu\, t}\, \sin(\omega_r\, t), \tag{3.47}$$

where $\omega_r = \sqrt{\omega_0^2 - \nu^2}$, $x_0 = x(0)$, and $\nu_0 = dx(0)/dt$. It can be seen that the solution oscillates at some real frequency, ω_r, which is somewhat less than the natural frequency of oscillation of the undamped system, ω_0, but also *decays* exponentially in time at a rate proportional to the damping coefficient, ν.

In the second case, $\nu = \omega_0$, and the motion is said to be *critically damped*. The most general solution is written

$$x(t) = [x_0\, (1 + \omega_0\, t) + \nu_0\, t]\, e^{-\omega_0\, t}. \tag{3.48}$$

It can be seen that the solution now decays without oscillating.

In the third case, $\nu > \omega_0$, and the motion is said to be *overdamped*. The most general solution is written

$$x(t) = -\left(\frac{\nu_0 + \nu_- x_0}{\nu_+ - \nu_-}\right) e^{-\nu_+\, t} + \left(\frac{\nu_0 + \nu_+ x_0}{\nu_+ - \nu_-}\right) e^{-\nu_-\, t}, \tag{3.49}$$

where $\nu_\pm = \nu \pm \sqrt{\nu^2 - \omega_0^2}$. It can be seen that the solution again decays without oscillating, except there are now *two* independent decay rates. The largest, ν_+, is always greater than the critically damped decay rate, ω_0, whereas the smaller, ν_-, is always less than this decay rate. This means that, in general, the critically damped solution is more rapidly damped than either the underdamped or overdamped solutions.

Figure 3.4 shows typical examples of underdamped (*i.e.*, $\nu = \omega_0/4$), critically damped (*i.e.*, $\nu = \omega_0$), and overdamped (*i.e.*, $\nu = 4\,\omega_0$) solutions, calculated with the initial conditions $x_0 = 1$ and $v_0 = 0$. Here, $T_0 = 2\pi/\omega_0$. The three solutions correspond to the solid, short-dashed, and long-dashed curves, respectively.

3.6 Quality Factor

The total energy of a damped oscillator is the sum of its kinetic and potential energies: *i.e.*,

$$E = \frac{1}{2}\,m\left(\frac{dx}{dt}\right)^2 + \frac{1}{2}\,m\,\omega_0^2\,x^2. \tag{3.50}$$

Differentiating the above expression with respect to time, we obtain

$$\frac{dE}{dt} = m\,\frac{dx}{dt}\,\frac{d^2x}{dt^2} + m\,\omega_0^2\,x\,\frac{dx}{dt} = m\,\frac{dx}{dt}\left(\frac{d^2x}{dt^2} + \omega_0^2\,x\right). \tag{3.51}$$

It follows from Equation (3.39) that

$$\frac{dE}{dt} = -2\,m\,\nu\left(\frac{dx}{dt}\right)^2. \tag{3.52}$$

We conclude that the presence of damping causes the oscillator energy to decrease monotonically in time, and, hence, causes the amplitude of the oscillation to eventually become negligibly small [see Equation (3.37)].

The energy loss rate of a weakly damped (*i.e.*, $\nu \ll \omega_0$) oscillator is conveniently characterized in terms of a parameter, Q, which is known as the *quality factor*. This parameter is defined to be 2π times the energy stored in the oscillator, divided by the energy lost in a single oscillation period. If the oscillator is weakly damped then the energy lost per period is relatively small, and Q is therefore much larger than unity. Roughly speaking, Q is the number of oscillations that the oscillator typically completes, after being set in motion, before its amplitude decays to a negligible value. Let us find an expression for Q.

Now, the most general solution for a weakly damped oscillator can be written in the form [*cf.*, Equation (3.47)]

$$x = x_0\,e^{-\nu\,t}\,\cos(\omega_r\,t - \phi_0), \tag{3.53}$$

where x_0 and ϕ_0 are constants, and $\omega_r = \sqrt{\omega_0^2 - \nu^2}$. It follows that

$$\frac{dx}{dt} = -x_0\,\nu\,e^{-\nu\,t}\,\cos(\omega_r\,t - \phi_0) - x_0\,\omega_r\,e^{-\nu\,t}\,\sin(\omega_r\,t - \phi_0). \tag{3.54}$$

Thus, making use of Equation (3.52), the energy lost during a single oscillation period is

$$\begin{aligned}\Delta E &= -\int_0^{T_r}\frac{dE}{dt}\,dt \qquad (3.55)\\ &= 2\,m\,\nu\,x_0^2\int_0^{T_r} e^{-2\,\nu\,t}\,[\nu\,\cos(\omega_r\,t - \phi_0) + \omega_r\,\sin(\omega_r\,t - \phi_0)]^2\,dt,\end{aligned}$$

where $T_r = 2\pi/\omega_r$. In the weakly damped limit, $\nu \ll \omega_r$, the exponential factor is approximately unity in the interval $t = 0$ to $t = T_r$, so that

$$\Delta E \simeq \frac{2\,m\,\nu\,x_0^2}{\omega_r}\int_0^{2\pi}\left(\nu^2\,\cos^2\theta + 2\,\nu\,\omega_r\,\cos\theta\,\sin\theta + \omega_r^2\,\sin^2\theta\right)d\theta. \tag{3.56}$$

Thus,

$$\Delta E \simeq \frac{2\pi\,m\,\nu\,x_0^2}{\omega_r}\,(\nu^2+\omega_r^2) = 2\pi\,m\,\omega_0^2\,x_0^2\left(\frac{\nu}{\omega_r}\right), \tag{3.57}$$

since $\cos^2\theta$ and $\sin^2\theta$ both have the average values $1/2$ in the interval 0 to 2π, whereas $\cos\theta\,\sin\theta$ has the average value 0. According to Equation (3.37), the energy stored in the oscillator (at $t = 0$) is

$$E = \frac{1}{2}\,m\,\omega_0^2\,x_0^2. \tag{3.58}$$

It follows that

$$Q = 2\pi\,\frac{E}{\Delta E} = \frac{\omega_r}{2\,\nu} \simeq \frac{\omega_0}{2\,\nu}. \tag{3.59}$$

3.7 Resonance

We have seen that when a one-dimensional dynamical system is slightly perturbed from a stable equilibrium point (and then left alone), it eventually returns to this point at a rate controlled by the amount of damping in the system. Let us now suppose that the same system is subject to *continuous* oscillatory constant amplitude external forcing at some fixed frequency, ω. In this case, we would expect the system to eventually settle down to some steady oscillatory pattern of motion with the same frequency as the external forcing. Let us investigate the properties of this type of driven oscillation.

Suppose that our system is subject to an external force of the form

$$f_{ext}(t) = m\,\omega_0^2\,X_1\,\cos(\omega\,t). \tag{3.60}$$

Here, X_1 is the amplitude of the oscillation at which the external force matches the restoring force, (3.32). Incorporating the above force into our perturbed equation of motion, (3.39), we obtain

$$\frac{d^2x}{dt^2} + 2\,\nu\,\frac{dx}{dt} + \omega_0^2\,x = \omega_0^2\,X_1\,\cos(\omega\,t). \tag{3.61}$$

Let us search for a solution of the form (3.40), and represent the right-hand side of the above equation as $\omega_0^2\,X_1\,\exp(-i\,\omega\,t)$. It is again understood that the physical solutions are the *real parts* of these expressions. Note that ω is now a real parameter. We obtain

$$a\left[-\omega^2 - i\,2\,\nu\,\omega + \omega_0^2\right]e^{-i\,\omega\,t} = \omega_0^2\,X_1\,e^{-i\,\omega\,t}. \tag{3.62}$$

Hence,

$$a = \frac{\omega_0^2\,X_1}{\omega_0^2 - \omega^2 - i\,2\,\nu\,\omega}. \tag{3.63}$$

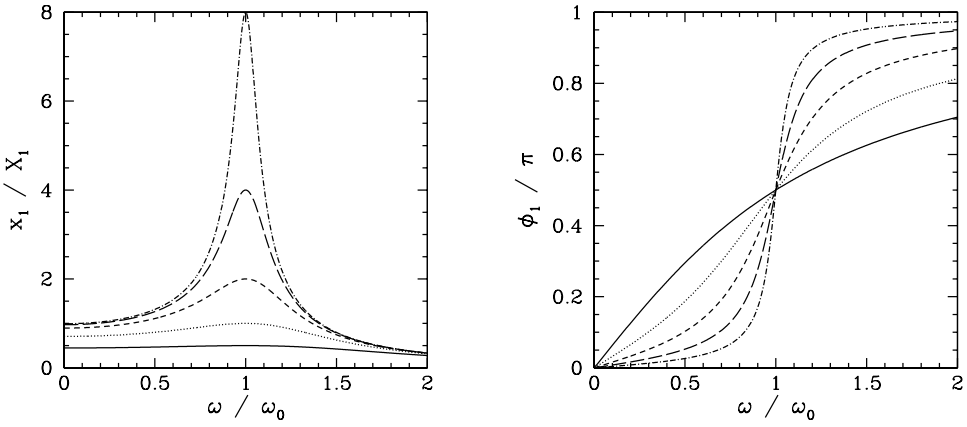


Figure 3.5: *Resonance.*

In general, a is a complex quantity. Thus, we can write

$$a = x_1\, e^{\,i\,\phi_1}, \tag{3.64}$$

where x_1 and ϕ_1 are both real. It follows from Equations (3.40), (3.63), and (3.64) that the physical solution takes the form

$$x(t) = x_1\, \cos(\omega\, t - \phi_1), \tag{3.65}$$

where

$$x_1 = \frac{\omega_0^2\, X_1}{\left[(\omega_0^2 - \omega^2)^2 + 4\,\nu^2\,\omega^2\right]^{1/2}}, \tag{3.66}$$

and

$$\phi_1 = \tan^{-1}\left(\frac{2\,\nu\,\omega}{\omega_0^2 - \omega^2}\right). \tag{3.67}$$

We conclude that, in response to the applied sinusoidal force, (3.60), the system executes a sinusoidal pattern of motion at the *same* frequency, with fixed amplitude x_1, and phase-lag ϕ_1 (with respect to the external force).

Let us investigate the variation of x_1 and ϕ_1 with the forcing frequency, ω. This is most easily done graphically. Figure 3.5 shows x_1 and ϕ_1 as functions of ω for various values of ν/ω_0. Here, $\nu/\omega_0 = 1$, $1/2$, $1/4$, $1/8$, and $1/16$ correspond to the solid, dotted, short-dashed, long-dashed, and dot-dashed curves, respectively. It can be seen that as the amount of damping in the system is decreased, the amplitude of the response becomes progressively more peaked at the natural frequency of oscillation of the system, ω_0. This effect is known as *resonance*, and ω_0 is termed the *resonant frequency*. Thus, a weakly damped system (*i.e.*, $\nu \ll \omega_0$) can be driven to large amplitude by the application of a relatively small external force which oscillates at a frequency close to the resonant frequency. Note that the response of the system is in phase (*i.e.*, $\phi_1 \simeq 0$) with the external driving force

for driving frequencies well below the resonant frequency, is in phase quadrature (*i.e.*, $\phi_1 = \pi/2$) at the resonant frequency, and is in anti-phase (*i.e.*, $\phi_1 \simeq \pi$) for frequencies well above the resonant frequency.

According to Equation (3.66),

$$\frac{x_1(\omega = \omega_0)}{x_1(\omega = 0)} = \frac{\omega_0}{2\,\nu} = Q. \tag{3.68}$$

In other words, the ratio of the driven amplitude at the resonant frequency to that at a typical non-resonant frequency (for the same drive amplitude) is of order the quality factor. Equation (3.66) also implies that, for a weakly damped oscillator ($\nu \ll \omega_0$),

$$\frac{x_1(\omega)}{x_1(\omega = \omega_0)} \simeq \frac{\nu}{[(\omega - \omega_0)^2 + \nu^2]^{1/2}}, \tag{3.69}$$

provided $|\omega - \omega_0| \ll \omega_0$. Hence, the width of the resonance peak (in frequency) is $\Delta\omega = 2\,\nu$, where the edges of peak are defined as the points at which the driven amplitude is reduced to $1/\sqrt{2}$ of its maximum value. It follows that the fractional width is

$$\frac{\Delta\omega}{\omega_0} = \frac{2\,\nu}{\omega_0} = \frac{1}{Q}. \tag{3.70}$$

We conclude that the height and width of the resonance peak of a weakly damped ($Q \gg 1$) oscillator scale as Q and Q^{-1}, respectively. Thus, the area under the resonance curve stays approximately constant as Q varies.

3.8 Periodic Driving Forces

In the last section, we investigated the response of a one-dimensional dynamical system, close to a stable equilibrium point, to an external force which varies as $\cos(\omega\,t)$. Let us now examine the response of the same system to a more complicated external force.

Consider a general external force which is *periodic* in time, with period T. By analogy with Equation (3.60), we can write such a force as

$$f_{ext}(t) = m\,\omega_0^2\,X(t), \tag{3.71}$$

where

$$X(t + T) = X(t) \tag{3.72}$$

for all t.

It is convenient to represent $X(t)$ as a *Fourier series* in time, so that

$$X(t) = \sum_{n=0}^{\infty} X_n\,\cos(n\,\omega\,t), \tag{3.73}$$

where $\omega = 2\pi/T$. By writing $X(t)$ in this form, we *automatically* satisfy the periodicity constraint (3.72). [Note that by choosing a cosine Fourier series we are limited to even functions in t: *i.e.*, $X(-t) = X(t)$. Odd functions in t can be represented by sine Fourier series, and mixed functions require a combination of cosine and sine Fourier series.] The constant coefficients X_n are known as *Fourier coefficients*. But, how do we determine these coefficients for a given functional form, $X(t)$?

Well, it follows from the periodicity of the cosine function that

$$\frac{1}{T}\int_0^T \cos(n\,\omega\,t)\,dt = \delta_{n\,0}, \tag{3.74}$$

where $\delta_{n\,n'}$ is unity if $n = n'$, and zero otherwise, and is known as the *Kronecker delta function*. Thus, integrating Equation (3.73) in t from $t = 0$ to $t = T$, and making use of Equation (3.74), we obtain

$$X_0 = \frac{1}{T}\int_0^T X(t)\,dt. \tag{3.75}$$

It is also easily demonstrated that

$$\frac{2}{T}\int_0^T \cos(n\,\omega\,t)\,\cos(n'\,\omega\,t)\,dt = \delta_{n\,n'}, \tag{3.76}$$

provided $n, n' > 0$. Thus, multiplying Equation (3.73) by $\cos(n\,\omega\,t)$, integrating in t from $t = 0$ to $t = T$, and making use of Equations (3.74) and (3.76), we obtain

$$X_n = \frac{2}{T}\int_0^T X(t)\,\cos(n\,\omega\,t)\,dt \tag{3.77}$$

for $n > 0$. Hence, we have now determined the Fourier coefficients of the general periodic function $X(t)$.

We can incorporate the periodic external force (3.71) into our perturbed equation of motion by writing

$$\frac{d^2x}{dt^2} + 2\,\nu\,\frac{dx}{dt} + \omega_0^2\,x = \omega_0^2\,\sum_{n=0}^{\infty} X_n\,e^{-i\,n\,\omega\,t}, \tag{3.78}$$

where we are again using the convention that the physical solution corresponds to the *real part* of the complex solution. Note that the above differential equation is *linear*. This means that if $x_a(t)$ and $x_b(t)$ represent two independent solutions to this equation then any linear combination of $x_a(t)$ and $x_b(t)$ is also a solution. We can exploit the linearity of the above equation to write the solution in the form

$$x(t) = \sum_{n=0}^{\infty} X_n\,a_n\,e^{-i\,n\,\omega\,t}, \tag{3.79}$$

where the a_n are the complex amplitudes of the solutions to

$$\frac{d^2x}{dt^2} + 2\,\nu\,\frac{dx}{dt} + \omega_0^2\,x = \omega_0^2\,e^{-i\,n\,\omega\,t}. \tag{3.80}$$

In other words, a_n is obtained by substituting $x = a_n \exp(-i\,n\,\omega\,t)$ into the above equation. Hence, it follows that

$$a_n = \frac{\omega_0^2}{\omega_0^2 - n^2\,\omega^2 - i\,2\,\nu\,n\,\omega}. \tag{3.81}$$

Thus, the physical solution takes the form

$$x(t) = \sum_{n=0}^{\infty} X_n\,x_n\,\cos(n\,\omega\,t - \phi_n), \tag{3.82}$$

where

$$a_n = x_n\,e^{\,i\,\phi_n}, \tag{3.83}$$

and x_n and ϕ_n are real parameters. It follows from Equation (3.81) that

$$x_n = \frac{\omega_0^2}{\left[(\omega_0^2 - n^2\,\omega^2)^2 + 4\,\nu^2\,n^2\,\omega^2\right]^{1/2}}, \tag{3.84}$$

and

$$\phi_n = \tan^{-1}\left(\frac{2\,\nu\,n\,\omega}{\omega_0^2 - n^2\,\omega^2}\right). \tag{3.85}$$

We have now fully determined the response of our dynamical system to a general periodic driving force.

As an example, suppose that the external force periodically delivers a brief kick to the system. For instance, let $X(t) = A$ for $0 \leq t \leq T/10$ and $9\,T/10 < t < T$, and $X(t) = 0$ otherwise (in the period $0 \leq t \leq T$). It follows from Equation (3.75) and (3.77) that, in this case,

$$X_0 = 0.2\,A, \tag{3.86}$$

and

$$X_n = \frac{2\,\sin(n\,\pi/5)\,A}{n\,\pi}, \tag{3.87}$$

for $n > 0$. Obviously, to obtain an exact solution, we would have to include every Fourier harmonic in Equation (3.82), which would be impractical. However, we can obtain a fairly accurate approximate solution by *truncating* the Fourier series (*i.e.*, by neglecting all the terms with $n > N$, where $N \gg 1$).

Figure 3.6 shows an example calculation in which the Fourier series is truncated after 100 terms. The parameters used in this calculation are $\omega = 1.2\,\omega_0$ and $\nu = 0.8\,\omega_0$. The left panel shows the Fourier reconstruction of the driving force, $X(t)$. The glitches at the rising and falling edges of the pulses are called *Gibbs phenomena*, and are an inevitable consequence of attempting to represent a discontinuous periodic function as a Fourier series. The right panel shows the Fourier reconstruction of the response, $x(t)$, of the dynamical system to the applied force.

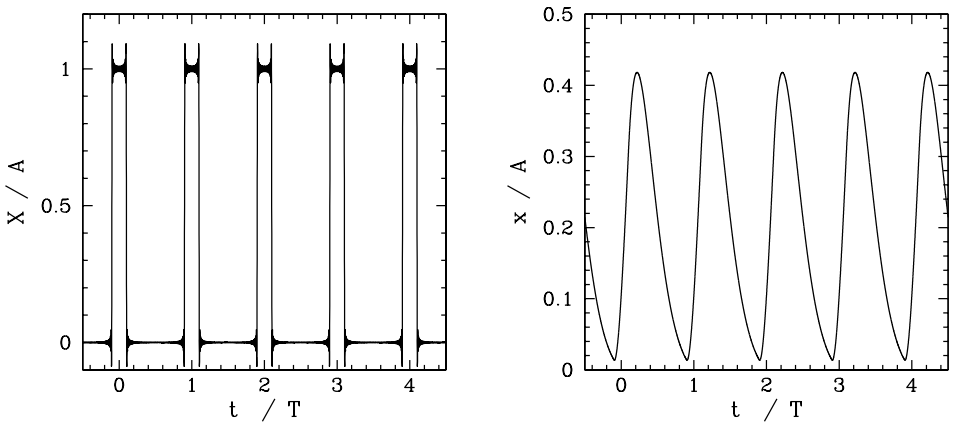


Figure 3.6: *Periodic forcing.*

3.9 Transients

We saw, in Section 3.7, that when a one-dimensional dynamical system, close to a stable equilibrium point, is subject to a sinusoidal external force of the form (3.60) then the equation of motion of the system is written

$$\frac{d^2x}{dt^2} + 2\,\nu\,\frac{dx}{dt} + \omega_0^2\,x = \omega_0^2\,X_1\,\cos(\omega\,t). \tag{3.88}$$

We also found that the solution to this equation which oscillates in sympathy with the applied force takes the form

$$x(t) = x_1\,\cos(\omega\,t - \phi_1), \tag{3.89}$$

where x_1 and ϕ_1 are specified in Equations (3.66) and (3.67), respectively. However, (3.89) is not the most general solution to Equation (3.88). It should be clear that we can take the above solution and add to it any solution of Equation (3.88) calculated with the right-hand side set to zero, and the result will also be a solution of Equation (3.88). Now, we investigated the solutions to (3.88) with the right-hand set to zero in Section 3.5. In the underdamped regime ($\nu < \omega_0$), we found that the most general such solution takes the form

$$x(t) = A\,e^{-\nu\,t}\,\cos(\omega_r\,t) + B\,e^{-\nu\,t}\,\sin(\omega_r\,t), \tag{3.90}$$

where A and B are two arbitrary constants [they are in fact the integration constants of the second-order ordinary differential equation (3.88)], and $\omega_r = \sqrt{\omega_0^2 - \nu^2}$. Thus, the most general solution to Equation (3.88) is written

$$x(t) = A\,e^{-\nu\,t}\,\cos(\omega_r\,t) + B\,e^{-\nu\,t}\,\sin(\omega_r\,t) + x_1\,\cos(\omega\,t - \phi_1). \tag{3.91}$$

The first two terms on the right-hand side of the above equation are called *transients*, since they decay in time. The transients are determined by the *initial conditions*. However, if we wait long enough after setting the system into motion then the transients will always

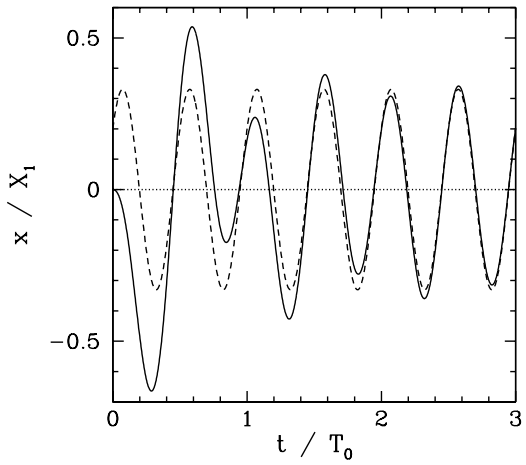


Figure 3.7: *Transients.*

decay away, leaving the time-asymptotic solution (3.89), which is independent of the initial conditions.

As an example, suppose that we set the system into motion at time $t = 0$ with the initial conditions $x(0) = dx(0)/dt = 0$. Setting $x(0) = 0$ in Equation (3.91), we obtain

$$A = -x_1 \cos\phi_1. \tag{3.92}$$

Moreover, setting $dx(0)/dt = 0$ in Equation (3.91), we get

$$B = -\frac{x_1\,(\nu\,\cos\phi_1 + \omega\,\sin\phi_1)}{\omega_r}. \tag{3.93}$$

Thus, we have now determined the constants A and B, and, hence, fully specified the solution for $t > 0$. Figure 3.7 shows this solution (solid curve) calculated for $\omega = 2\,\omega_0$ and $\nu = 0.2\,\omega_0$. Here, $T_0 = 2\pi/\omega_0$. The associated time-asymptotic solution (3.89) is also shown for the sake of comparison (dashed curve). It can be seen that the full solution quickly converges to the time-asymptotic solution.

3.10 Simple Pendulum

Consider a compact mass m suspended from a light inextensible string of length l, such that the mass is free to swing from side to side in a vertical plane, as shown in Figure 3.8. This setup is known as a *simple pendulum*. Let θ be the angle subtended between the string and the downward vertical. Obviously, the stable equilibrium state of the simple pendulum corresponds to the situation in which the mass is stationary, and hangs vertically down (*i.e.*, $\theta = 0$). From elementary mechanics, the angular equation of motion of the pendulum is

$$I\,\frac{d^2\theta}{dt^2} = \tau, \tag{3.94}$$

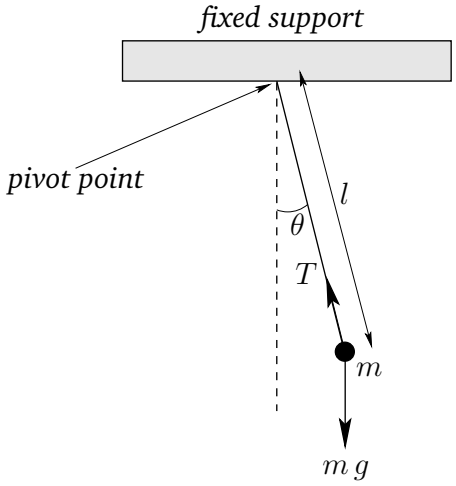


Figure 3.8: *A simple pendulum.*

where I is the moment of inertia of the mass (see Section 8.3), and τ is the torque acting about the pivot point (see Section A.7). For the case in hand, given that the mass is essentially a point particle, and is situated a distance l from the axis of rotation (*i.e.*, the pivot point), it is easily seen that $I = m\,l^2$.

The two forces acting on the mass are the downward gravitational force, $m\,g$, where g is the acceleration due to gravity, and the tension, T, in the string. Note, however, that the tension makes no contribution to the torque, since its line of action clearly passes through the pivot point. From simple trigonometry, the line of action of the gravitational force passes a distance $l\,\sin\theta$ from the pivot point. Hence, the magnitude of the gravitational torque is $m\,g\,l\,\sin\theta$. Moreover, the gravitational torque is a *restoring torque*: *i.e.*, if the mass is displaced slightly from its equilibrium state (*i.e.*, $\theta = 0$) then the gravitational torque clearly acts to push the mass back toward that state. Thus, we can write

$$\tau = -m\,g\,l\,\sin\theta. \tag{3.95}$$

Combining the previous two equations, we obtain the following angular equation of motion of the pendulum:

$$l\,\frac{d^2\theta}{dt^2} + g\,\sin\theta = 0. \tag{3.96}$$

Note that, unlike all of the other equations of motion which we have examined in this chapter, the above equation is *nonlinear* [since $\sin(a\,\theta) \neq a\,\sin\theta$, except when $\theta \ll 1$].

Let us assume, as usual, that the system does not stray very far from its equilibrium state ($\theta = 0$). If this is the case then we can make the small angle approximation $\sin\theta \simeq \theta$, and the above equation of motion simplifies to

$$\frac{d^2\theta}{dt^2} + \omega_0^2\,\theta \simeq 0, \tag{3.97}$$

where $\omega_0 = \sqrt{g/l}$. Of course, this is just the simple harmonic equation. Hence, we can immediately write the solution as

$$\theta(t) = \theta_0 \cos(\omega_0 t). \tag{3.98}$$

We conclude that the pendulum swings back and forth at a fixed frequency, ω_0, which depends on l and g, but is *independent* of the amplitude, θ_0, of the motion.

Suppose, now, that we desire a more accurate solution of Equation (3.96). One way in which we could achieve this would be to include more terms in the small angle expansion of $\sin\theta$, which is

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \cdots. \tag{3.99}$$

For instance, keeping the first two terms in this expansion, Equation (3.96) becomes

$$\frac{d^2\theta}{dt^2} + \omega_0^2\,(\theta - \theta^3/6) \simeq 0. \tag{3.100}$$

By analogy with (3.98), let us try a trial solution of the form

$$\theta(t) = \vartheta_0 \cos(\omega t). \tag{3.101}$$

Substituting this into Equation (3.100), and making use of the trigonometric identity

$$\cos^3 u \equiv (3/4)\cos u + (1/4)\cos(3u), \tag{3.102}$$

we obtain

$$\vartheta_0\left[\omega_0^2 - \omega^2 - (1/8)\,\omega_0^2\,\vartheta_0^2\right]\cos(\omega t) - (1/24)\,\omega_0^2\,\vartheta_0^3\cos(3\omega t) \simeq 0. \tag{3.103}$$

It is evident that the above equation cannot be satisfied for all values of t, except in the trivial case $\vartheta_0 = 0$. However, the form of this expression does suggest a better trial solution, namely

$$\theta(t) = \vartheta_0\cos(\omega t) + \alpha\,\vartheta_0^3\cos(3\omega t), \tag{3.104}$$

where α is $\mathcal{O}(1)$. Substitution of this expression into Equation (3.100) yields

$$\begin{aligned}\vartheta_0\left[\omega_0^2 - \omega^2 - (1/8)\,\omega_0^2\,\vartheta_0^2\right]\cos(\omega t) +\\ \vartheta_0^3\left[\alpha\,\omega_0^2 - 9\,\alpha\,\omega^2 - (1/24)\,\omega_0^2\right]\cos(3\omega t) + \mathcal{O}(\vartheta_0^5) \simeq 0.\end{aligned} \tag{3.105}$$

We can only satisfy the above equation at all values of t, and for non-zero ϑ_0, by setting the two expressions in square brackets to zero. This yields

$$\omega \simeq \omega_0\sqrt{1 - (1/8)\,\vartheta_0^2}, \tag{3.106}$$

and

$$\alpha \simeq -\frac{\omega_0^2}{192}. \tag{3.107}$$

Now, the amplitude of the motion is given by

$$\theta_0 = \vartheta_0 + \alpha\,\vartheta_0^3 = \vartheta_0 - \frac{\omega_0^2}{192}\,\vartheta_0^3. \tag{3.108}$$

Hence, Equation (3.106) simplifies to

$$\omega = \omega_0\left[1 - \frac{\theta_0^2}{16} + \mathcal{O}(\theta_0^4)\right]. \tag{3.109}$$

The above expression is only approximate, but it illustrates an important point: *i.e.*, that the frequency of oscillation of a simple pendulum is *not*, in fact, amplitude independent. Indeed, the frequency goes down slightly as the amplitude increases.

The above example illustrates how we might go about solving a nonlinear equation of motion by means of an expansion in a small parameter (in this case, the amplitude of the motion).

3.11 Exercises

3.1. If a train of mass M is subject to a retarding force $M\,(a + b\,v^2)$, show that if the engines are shut off when the speed is v_0 then the train will come to rest in a time

$$\frac{1}{\sqrt{a\,b}}\,\tan^{-1}\left(\sqrt{\frac{b}{a}}\,v_0\right),$$

after traveling a distance

$$\frac{1}{2\,b}\,\ln\left(1 + \frac{b\,v_0^2}{a}\right).$$

3.2. A particle is projected vertically upward from the Earth's surface with a velocity which would, if gravity were uniform, carry it to a height h. Show that if the variation of gravity with height is allowed for, but the resistance of air is neglected, then the height reached will be greater by $h^2/(R - h)$, where R is the Earth's radius.

3.3. A particle is projected vertically upward from the Earth's surface with a velocity just sufficient for it to reach infinity (neglecting air resistance). Prove that the time needed to reach a height h is

$$\frac{1}{3}\left(\frac{2\,R}{g}\right)^{1/2}\left[\left(1 + \frac{h}{R}\right)^{3/2} - 1\right].$$

where R is the Earth's radius, and g its surface gravitational acceleration.

3.4. A particle of mass m is constrained to move in one dimension such that its instantaneous displacement is x. The particle is released at rest from $x = b$, and is subject to a force of the form $f(x) = -k\,x^{-2}$. Show that the time required for the particle to reach the origin is

$$\pi\left(\frac{m\,b^3}{8\,k}\right)^{1/2}.$$

3.5. A block of mass m slides along a horizontal surface which is lubricated with heavy oil such that the block suffers a viscous retarding force of the form

$$F = -c\,v^n,$$

where $c > 0$ is a constant, and v is the block's instantaneous velocity. If the initial speed is v_0 at time $t = 0$, find v and the displacement x as functions of time t. Also find v as a function of x. Show that for $n = 1/2$ the block does not travel further than $2\,m\,v_0^{3/2}/(3\,c)$.

3.6. A particle is projected vertically upward in a constant gravitational field with an initial speed v_0. Show that if there is a retarding force proportional to the square of the speed then the speed of the particle when it returns to the initial position is

$$\frac{v_0\,v_t}{\sqrt{v_0^2 + v_t^2}},$$

where v_t is the terminal speed.

3.7. A particle of mass m moves (in one dimension) in a medium under the influence of a retarding force of the form $m\,k\,(v^3 + a^2\,v)$, where v is the particle speed, and k and a are positive constants. Show that for any value of the initial speed the particle will never move a distance greater than $\pi/(2\,k\,a)$, and will only come to rest as $t \to \infty$.

3.8. Two light springs have spring constants k_1 and k_2, respectively, and are used in a vertical orientation to support an object of mass m. Show that the angular frequency of oscillation is $[(k_1 + k_2)/m]^{1/2}$ if the springs are in parallel, and $[k_1\,k_2/(k_1 + k_2)\,m]^{1/2}$ if the springs are in series.

3.9. A body of uniform cross-sectional area A and mass density ρ floats in a liquid of density ρ_0 (where $\rho < \rho_0$), and at equilibrium displaces a volume V. Show that the period of small oscillations about the equilibrium position is

$$T = 2\pi\sqrt{\frac{V}{g\,A}}.$$

3.10. Show that the ratio of two successive maxima in the displacement of a damped harmonic oscillator is constant.

3.11. If the amplitude of a damped harmonic oscillator decreases to $1/e$ of its initial value after n periods show that the ratio of the period of oscillation to the period of the same oscillator with no damping is

$$\left(1 + \frac{1}{4\pi^2\,n^2}\right)^{1/2} \simeq 1 + \frac{1}{8\pi^2\,n^2}.$$

3.12. Consider a damped driven oscillator whose equation of motion is

$$\frac{d^2x}{dt^2} + 2\,\nu\,\frac{dx}{dt} + \omega_0^2\,x = F(t).$$

Let $x = 0$ and $dx/dt = v_0$ at $t = 0$.

(a) Find the solution for $t > 0$ when $F = \sin(\omega t)$.

(b) Find the solution for $t > 0$ when $F = \sin^2(\omega t)$.

3.13. Obtain the time asymptotic response of a damped linear oscillator of natural frequency ω_0 and damping coefficient ν to a square-wave periodic forcing function of amplitude $F_0 = m\,\omega_0^2\,X_0$ and frequency ω. Thus, $F(t) = F_0$ for $-\pi/2 < \omega t < \pi/2$, $3\pi/2 < \omega t < 5\pi/2$, *etc.*, and $F(t) = -F_0$ for $\pi/2 < \omega t < 3\pi/2$, $5\pi/2 < \omega t < 7\pi/2$, *etc.*

4 Multi-Dimensional Motion

4.1 Introduction

This chapter employs Newton's laws to investigate various aspects of multi-dimensional motion.

4.2 Motion in a Two-Dimensional Harmonic Potential

Consider a particle of mass m moving in the two-dimensional harmonic potential

$$U(x,y) = \frac{1}{2}\,k\,r^2, \tag{4.1}$$

where $r = \sqrt{x^2+y^2}$, and $k > 0$. It follows that the particle is subject to a force,

$$\mathbf{f} = -\nabla U = -k\,(x,\,y) = -k\,\mathbf{r}, \tag{4.2}$$

which always points *towards* the origin, and whose magnitude increases *linearly* with increasing distance from the origin. According to Newton's second law, the equation of motion of the particle is

$$m\,\frac{d^2\mathbf{r}}{dt^2} = \mathbf{f} = -k\,\mathbf{r}. \tag{4.3}$$

When written in component form, the above equation reduces to

$$\frac{d^2x}{dt^2} = -\omega_0^2\,x, \tag{4.4}$$

$$\frac{d^2y}{dt^2} = -\omega_0^2\,y, \tag{4.5}$$

where $\omega_0 = \sqrt{k/m}$.

Since Equations (4.4) and (4.5) are both *simple harmonic equations*, we can immediately write their general solutions:

$$x = A\,\cos(\omega_0\,t - \phi_1), \tag{4.6}$$

$$y = B\,\cos(\omega_0\,t - \phi_2). \tag{4.7}$$

Here, A, B, ϕ_1, and ϕ_2 are arbitrary constants of integration. We can simplify the above equations slightly by shifting the origin of time (which is, after all, arbitrary): *i.e.*,

$$t \rightarrow t' + \phi_1/\omega_0. \tag{4.8}$$

Hence, we obtain

$$x = A\cos(\omega_0 t'), \tag{4.9}$$

$$y = B\cos(\omega_0 t' - \Delta), \tag{4.10}$$

where $\Delta = \phi_2 - \phi_1$. Note that the motion is clearly *periodic* in time, with period $T = 2\pi/\omega_0$. Thus, the particle must trace out some *closed trajectory* in the x-y plane. The question, now, is what does this trajectory look like as a function of the relative phase-shift, Δ, between the oscillations in the x- and y-directions?

Using standard trigonometry, we can write Equation (4.10) in the form

$$y = B\left[\cos(\omega_0 t')\,\cos\Delta + \sin(\omega_0 t')\,\sin\Delta\right]. \tag{4.11}$$

Hence, using Equation (4.9), we obtain

$$\left(\frac{y}{B} - \frac{x}{A}\cos\Delta\right)^2 = \sin^2(\omega_0 t')\,\sin^2\Delta = \left(1 - \frac{x^2}{A^2}\right)\sin^2\Delta, \tag{4.12}$$

which simplifies to give

$$\frac{x^2}{A^2} - 2\,\frac{x\,y}{A\,B}\,\cos\Delta + \frac{y^2}{B^2} = \sin^2\Delta. \tag{4.13}$$

Unfortunately, the above equation is not immediately recognizable as being the equation of any particular geometric curve: *e.g.*, a circle, an ellipse, or a parabola, *etc*.

Perhaps our problem is that we are using the wrong coordinates. Suppose that we rotate our coordinate axes about the z-axis by an angle θ, as illustrated in Figure A.5. According to Equations (A.17) and (A.18), our old coordinates (x, y) are related to our new coordinates (x', y') via

$$x = x'\cos\theta - y'\sin\theta, \tag{4.14}$$

$$y = x'\sin\theta + y'\cos\theta. \tag{4.15}$$

Let us see whether Equation (4.13) takes a simpler form when expressed in terms of our new coordinates. Equations (4.13)–(4.15) yield

$$\begin{aligned} &x'^2\left[\frac{\cos^2\theta}{A^2} - \frac{2\,\cos\theta\,\sin\theta\,\cos\Delta}{A\,B} + \frac{\sin^2\theta}{B^2}\right] \\ &+y'^2\left[\frac{\sin^2\theta}{A^2} + \frac{2\,\cos\theta\,\sin\theta\,\cos\Delta}{A\,B} + \frac{\cos^2\theta}{B^2}\right] \\ &+x'\,y'\left[-\frac{2\,\sin\theta\,\cos\theta}{A^2} + \frac{2\,(\sin^2\theta - \cos^2\theta)\,\cos\Delta}{A\,B} + \frac{2\,\cos\theta\,\sin\theta}{B^2}\right] = \sin^2\Delta. \end{aligned} \tag{4.16}$$

We can simplify the above equation by setting the term involving $x'\,y'$ to zero. Hence,

$$-\frac{\sin(2\,\theta)}{A^2} - \frac{2\,\cos(2\,\theta)\,\cos\Delta}{A\,B} + \frac{\sin(2\,\theta)}{B^2} = 0, \tag{4.17}$$

where we have made use of some simple trigonometric identities. Thus, the $x'y'$ term disappears when θ takes the special value

$$\theta = \frac{1}{2}\,\tan^{-1}\left(\frac{2\,A\,B\,\cos\Delta}{A^2-B^2}\right). \tag{4.18}$$

In this case, Equation (4.16) reduces to

$$\frac{x'^2}{a^2}+\frac{y'^2}{b^2}=1, \tag{4.19}$$

where

$$\frac{1}{a^2} = \frac{1}{\sin^2\Delta}\left[\frac{\cos^2\theta}{A^2}-\frac{2\,\cos\theta\,\sin\theta\,\cos\Delta}{A\,B}+\frac{\sin^2\theta}{B^2}\right], \tag{4.20}$$

$$\frac{1}{b^2} = \frac{1}{\sin^2\Delta}\left[\frac{\sin^2\theta}{A^2}+\frac{2\,\cos\theta\,\sin\theta\,\cos\Delta}{A\,B}+\frac{\cos^2\theta}{B^2}\right]. \tag{4.21}$$

Of course, we immediately recognize Equation (4.19) as the equation of an *ellipse*, centered on the origin, whose major and minor axes are aligned along the x'- and y'-axes, and whose major and minor radii are a and b, respectively (assuming that $a > b$).

We conclude that, in general, a particle of mass m moving in the two-dimensional harmonic potential (4.1) executes a *closed elliptical orbit* (which is not necessarily aligned along the x- and y-axes), centered on the origin, with period $T = 2\pi/\omega_0$, where $\omega_0 = \sqrt{k/m}$.

Figure 4.1 shows some example trajectories calculated for $A = 2$, $B = 1$, and the following values of the phase difference, Δ: (a) $\Delta = 0$; (b) $\Delta = \pi/4$; (c) $\Delta = \pi/2$; (d) $\Delta = 3\pi/4$. Note that when $\Delta = 0$ the trajectory degenerates into a straight-line (which can be thought of as an ellipse whose minor radius is zero).

Perhaps, the main lesson to be learned from the above study of two-dimensional motion in a harmonic potential is that comparatively simple patterns of motion can be made to look complicated when expressed in terms of ill-chosen coordinates.

4.3 Projectile Motion with Air Resistance

Suppose that a projectile of mass m is launched, at $t = 0$, from ground level (in a flat plain), making an angle θ to the horizontal. Suppose, further, that, in addition to the force of gravity, the projectile is subject to an air resistance force which acts in the opposite direction to its instantaneous direction of motion, and whose magnitude is *directly proportional* to its instantaneous speed. This is not a particularly accurate model of the drag force due to air resistance (the magnitude of the drag force is typically proportion to the square of the speed—see Section 3.3), but it does lead to tractable equations of motion. Hence, by using this model we can, at least, get some idea of how air resistance modifies projectile trajectories.

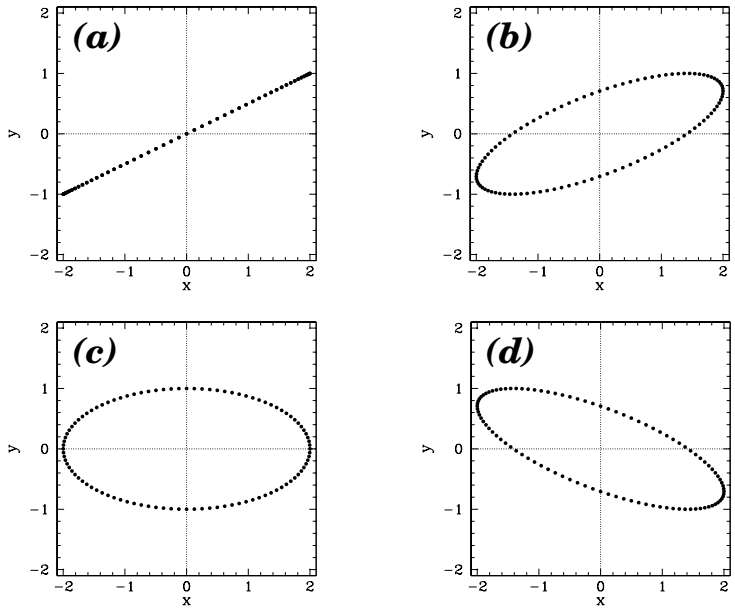


Figure 4.1: *Trajectories in a two-dimensional harmonic oscillator potential.*

Let us adopt a Cartesian coordinate system whose origin coincides with the launch point, and whose z-axis points vertically upward. Let the initial velocity of the projectile lie in the x-z plane. Note that, since neither gravity nor the drag force cause the projectile to move out of the x-z plane, we can effectively ignore the y coordinate in this problem.

The equation of motion of our projectile is written

$$m\,\frac{d\mathbf{v}}{dt} = m\,\mathbf{g} - c\,\mathbf{v}, \tag{4.22}$$

where $\mathbf{v} = (v_x, v_z)$ is the projectile velocity, $\mathbf{g} = (0, -g)$ the acceleration due to gravity, and c a positive constant. In component form, the above equation becomes

$$\frac{dv_x}{dt} = -g\,\frac{v_x}{v_t}, \tag{4.23}$$

$$\frac{dv_z}{dt} = -g\left(1 + \frac{v_z}{v_t}\right). \tag{4.24}$$

Here, $v_t = m\,g/c$ is the *terminal velocity*: *i.e.*, the velocity at which the drag force balances the gravitational force (for a projectile falling vertically downward).

Integrating Equation (4.23), we obtain

$$\int_{v_{x\,0}}^{v} \frac{dv_x}{v_x} = -\frac{g}{v_t}\,t, \tag{4.25}$$

where $v_{x\,0} = v_0\,\cos\theta$ is the x-component of the launch velocity. Hence,

$$\ln\left(\frac{v_x}{v_{x\,0}}\right) = -\frac{g}{v_t}\,t, \tag{4.26}$$

or

$$v_x = v_0 \cos\theta \, e^{-g\,t/v_t}. \tag{4.27}$$

It is clear, from the above equation, that air drag causes the projectile's horizontal velocity, which would otherwise be constant, to *decay* exponentially on a time-scale of order v_t/g.

Integrating Equation (4.24), we get

$$\int_{v_{z\,0}}^{v_z} \frac{dv_z}{v_t + v_z} = -\frac{g}{v_t}\, t, \tag{4.28}$$

where $v_{z\,0} = v_0 \sin\theta$ is the z-component of the launch velocity. Hence,

$$\ln\left(\frac{v_t + v_z}{v_t + v_{z\,0}}\right) = -\frac{g}{v_t}\, t, \tag{4.29}$$

or

$$v_z = v_0 \sin\theta \, e^{-g\,t/v_t} - v_t \left(1 - e^{-g\,t/v_t}\right). \tag{4.30}$$

It thus follows, from Equations (4.27) and (4.30), that if the projectile stays in the air much longer than a time of order v_t/g then it ends up falling vertically downward at the terminal velocity, v_t, irrespective of its initial launch angle.

Integration of (4.27) yields

$$x = \frac{v_0\, v_t \cos\theta}{g}\left(1 - e^{-g\,t/v_t}\right). \tag{4.31}$$

In the limit $t \ll v_t/g$, the above equation reduces to

$$x = v_0 \cos\theta \, t, \tag{4.32}$$

which is the standard result in the absence of air drag. In the opposite limit, $t \gg v_t/g$, we get

$$x = \frac{v_0\, v_t \cos\theta}{g}. \tag{4.33}$$

The above expression clearly sets an effective upper limit on how far the projectile can travel in the horizontal direction.

Integration of (4.30) gives

$$z = \frac{v_t}{g}\,(v_0 \sin\theta + v_t)\left(1 - e^{-g\,t/v_t}\right) - v_t\, t. \tag{4.34}$$

In the limit $t \ll v_t/g$, this equation reduces to

$$z = v_0 \sin\theta \, t - \frac{g}{2}\, t^2, \tag{4.35}$$

which is the standard result in the absence of air drag. In the opposite limit, $t \gg v_t/g$, we get

$$z = \frac{v_t}{g}\,(v_0 \sin\theta + v_t) - v_t\, t. \tag{4.36}$$

Incidentally, the above analysis implies that air resistance only starts to have an appreciable effect on the trajectory after the projectile has been in the air a time of order v_t/g.

It is clear, from the previous two equations, that the time of flight of the projectile (*i.e.*, the time at which $z = 0$, excluding the trivial result $t = 0$) is

$$t_f = \frac{2\,v_0\,\sin\theta}{g} \tag{4.37}$$

when $t \ll v_t/g$, which implies that $v_0\,\sin\theta \ll v_t$, and

$$t_f = \frac{v_0\,\sin\theta}{g} \tag{4.38}$$

when $t \gg v_t/g$, which implies that $v_0\,\sin\theta \gg v_t$ (*i.e.*, the vertical component of the launch velocity is much greater than the terminal velocity). It thus follows, from Equations (4.32) and (4.33), that the horizontal range [*i.e.*, $x(t_f)$] of the projectile is

$$R = \frac{v_0^2\,\sin(2\,\theta)}{g} \tag{4.39}$$

when $v_0\,\sin\theta \ll v_t$, and

$$R = \frac{v_0\,v_t\,\cos\theta}{g} \tag{4.40}$$

when $v_0\,\sin\theta \gg v_t$. Equation (4.39) is, of course, the standard result without air resistance. This result implies that, in the absence of air resistance, the maximum horizontal range, v_0^2/g, is achieved when the launch angle θ takes the value 45°. On the other hand, Equation (4.40) implies that, in the presence of air resistance, the maximum horizontal range, $v_0\,v_t/g$, is achieved when θ is made as small as possible. However, θ cannot be made too small, since expression (4.40) is only valid when $v_0\,\sin\theta \gg v_t$. In fact, assuming that $v_0 \gg v_t$, the maximum horizontal range, $v_0\,v_t/g$, is achieved when $\theta \sim v_t/v_0 \ll 1$. We thus conclude that if air resistance is significant then it causes the horizontal range of the projectile to scale *linearly*, rather than quadratically, with the launch velocity, v_0. Moreover, the maximum horizontal range is achieved with a launch angle which is *much shallower* than the standard result, 45°.

Figure 4.2 shows some example trajectories calculated, from the above model, with the same launch angle, 45°, but with different values of the ratio v_0/v_t. Here, $X = x/(v_0^2/g)$ and $Z = z/(v_0^2/g)$. The solid, short-dashed, long-dashed, and dot-dashed curves correspond to $v_0/v_t = 0$, 1, 2, and 4, respectively. It can be seen that as the air resistance strength increases (*i.e.*, as v_0/v_t increases), the range of the projectile decreases. Furthermore, there is always an initial time interval during which the trajectory is identical to that calculated in the absence of air resistance (*i.e.*, $v_0/v_t = 0$). Finally, in the presence of air resistance, the projectile tends to fall more steeply than it rises. Indeed, in the presence of strong air resistance (*i.e.*, $v_0/v_t = 4$), the projectile falls almost vertically.

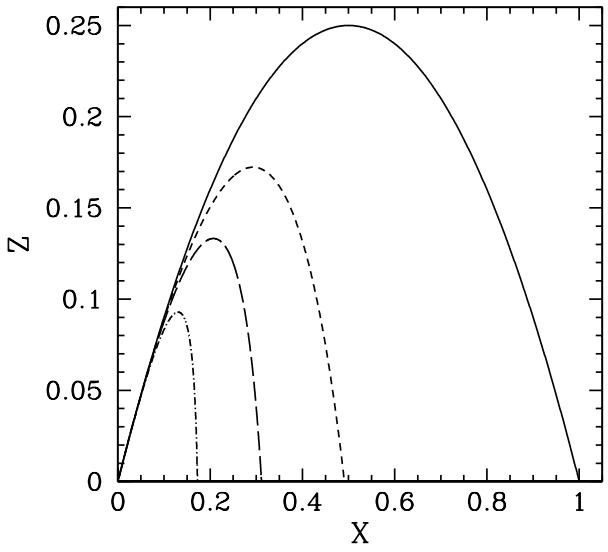


Figure 4.2: *Projectile trajectories in the presence of air resistance.*

4.4 Charged Particle Motion in Electric and Magnetic Fields

Consider a particle of mass m and electric charge q moving in the *uniform* electric and magnetic fields, $\mathbf{E}$ and $\mathbf{B}$. Suppose that the fields are "crossed" (*i.e.*, perpendicular to one another), so that $\mathbf{E}\cdot\mathbf{B}=0$.

The force acting on the particle is given by the familiar Lorentz law:

$$\mathbf{f} = q\,(\mathbf{E}+\mathbf{v}\times\mathbf{B}), \tag{4.41}$$

where $\mathbf{v}$ is the particle's instantaneous velocity. Hence, from Newton's second law, the particle's equation of motion can be written

$$m\frac{d\mathbf{v}}{dt} = q\,(\mathbf{E}+\mathbf{v}\times\mathbf{B}). \tag{4.42}$$

It turns out that we can eliminate the electric field from the above equation by transforming to a different inertial frame. Thus, writing

$$\mathbf{v} = \frac{\mathbf{E}\times\mathbf{B}}{B^2}+\mathbf{v}', \tag{4.43}$$

Equation (4.42) reduces to

$$m\frac{d\mathbf{v}'}{dt} = q\,\mathbf{v}'\times\mathbf{B}, \tag{4.44}$$

where we have made use of a standard vector identity (see Section A.10), as well as the fact that $\mathbf{E}\cdot\mathbf{B}=0$. Hence, we conclude that the addition of an electric field perpendicular to

a given magnetic field simply causes the particle to drift perpendicular to both the electric and magnetic field with the fixed velocity

$$\mathbf{v}_{EB} = \frac{\mathbf{E}\times\mathbf{B}}{B^2}, \tag{4.45}$$

irrespective of its charge or mass. It follows that the electric field has no effect on the particle's motion in a frame of reference which is co-moving with the so-called *E-cross-B velocity* given above.

Let us suppose that the magnetic field is directed along the z-axis. As we have just seen, in the $\mathbf{E}\times\mathbf{B}$ frame, the particle's equation of motion reduces to Equation (4.44), which can be written:

$$\frac{dv'_x}{dt} = \Omega\, v'_y, \tag{4.46}$$

$$\frac{dv'_y}{dt} = -\Omega\, v'_x, \tag{4.47}$$

$$\frac{dv'_z}{dt} = 0. \tag{4.48}$$

Here,

$$\Omega = \frac{q\,B}{m} \tag{4.49}$$

is the so-called *cyclotron frequency*. Equations (4.46)–(4.48) can be integrated to give

$$v'_x = v_\perp \sin(\Omega\, t), \tag{4.50}$$

$$v'_y = v_\perp \cos(\Omega\, t) \tag{4.51}$$

$$v'_z = v_\parallel, \tag{4.52}$$

where we have judiciously chosen the origin of time so as to eliminate any phase offset in the arguments of the above trigonometrical functions. According to Equations (4.50)–(4.52), in the $\mathbf{E}\times\mathbf{B}$ frame, our charged particle gyrates at the cyclotron frequency in the plane perpendicular to the magnetic field with some fixed speed $v_\perp$, and drifts parallel to the magnetic field with some fixed speed $v_\parallel$. The fact that the cyclotron frequency is positive for positively charged particles, and negative for negatively charged particles, just means that oppositely charged particles gyrate in opposite directions in the plane perpendicular to the magnetic field.

Equations (4.50)–(4.52) can be integrated to give

$$x' = -\rho\, \cos(\Omega\, t), \tag{4.53}$$

$$y' = \rho\, \sin(\Omega\, t) \tag{4.54}$$

$$z' = v_\parallel\, t, \tag{4.55}$$

where we have judiciously chosen the origin of our coordinate system so as to eliminate any constant offsets in the above equations. Here,

$$\rho = \frac{v_\perp}{\Omega} \tag{4.56}$$

is called the *Larmor radius*. Equations (4.53)–(4.55) are the equations of a *spiral* of radius ρ, aligned along the direction of the magnetic field (*i.e.*, the z-direction).

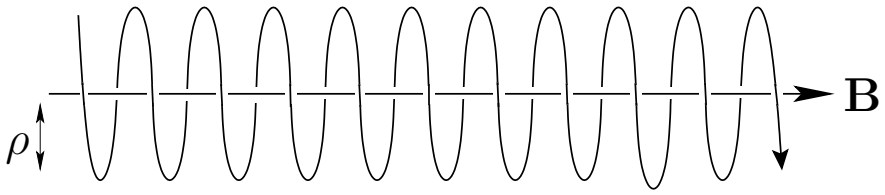


Figure 4.3: *The spiral trajectory of a negatively charged particle in a magnetic field.*

We conclude that the general motion of a charged particle in crossed electric and magnetic field is a combination of $\mathbf{E} \times \mathbf{B}$ drift [see Equation (4.45)] and spiral motion aligned along the direction of the magnetic field—see Figure 4.3. Particles drift parallel to the magnetic field with constant speeds, and gyrate at the cyclotron frequency in the plane perpendicular to the magnetic field with constant speeds. Oppositely charged particles gyrate in opposite directions.

4.5 Exercises

4.1. An electron of mass m and charge $-e$ moves in a uniform y-directed electric field of magnitude E, and a uniform z-directed magnetic field of magnitude B. The electron is situated at the origin at $t = 0$ with an initial x-directed velocity of magnitude v_0. Show that the electron traces out a cycloid of the general form

$$\begin{aligned} x &= a \sin(\omega t) + b t, \\ y &= c\,[1 - \cos(\omega t)], \\ z &= 0. \end{aligned}$$

Find the values of a, b, c, and ω, and sketch the electron's trajectory in the x-y plane when $v_0 < E/B$, $E/B < v_0 < 2\,E/B$, and $v_0 > 2\,E/B$.

4.2. A particle of mass m and charge q moves in the x-y plane under the influence of a constant amplitude rotating electric field which is such that $E_x = E_0 \cos(\omega t)$ and $E_y = E_0 \sin(\omega t)$. The particle starts at rest from the origin. Determine its subsequent motion. What shape is the particle's trajectory?

4.3. A particle of mass m slides on a frictionless surface whose height is a function of x only: *i.e.*, $z = z(x)$. The function $z(x)$ is specified by the parametric equations

$$\begin{aligned} x &= A\,[2\,\phi + \sin(2\,\phi)], \\ z &= A\,[1 - \cos(2\,\phi)], \end{aligned}$$

where ϕ is the parameter. Show that the total energy of the particle can be written

$$E = \frac{m}{2}\left(\frac{ds}{dt}\right)^2 + \frac{1}{2}\,\frac{m\,g}{4\,A}\,s^2,$$

where $s = 4\,A\,\sin\phi$. Deduce that the particle undergoes periodic motion whose frequency is amplitude independent (even when the amplitude is large). Demonstrate that the frequency of the motion is given by $4\pi\,(A/g)^{1/2}$.

5 Planetary Motion

5.1 Introduction

Newtonian dynamics was initially developed in order to account for the motion of the Planets around the Sun. Let us now investigate this problem.

5.2 Kepler's Laws

As is well-known, Johannes Kepler was the first astronomer to correctly describe planetary motion in the Solar System (in works published between 1609 and 1619). The motion of the Planets is summed up in three simple laws:

1. The planetary orbits are all ellipses which are confocal with the Sun (*i.e.*, the Sun lies at one of the focii of each ellipse).

2. The radius vectors connecting each planet to the Sun sweep out equal areas in equal time intervals.

3. The squares of the orbital periods of the planets are proportional to the cubes of their orbital major radii.

Let us now see if we can derive Kepler's laws from Newton's laws of motion.

5.3 Newtonian Gravity

The force which maintains the Planets in orbit around the Sun is called *gravity*, and was first correctly described by Isaac Newton (in 1687). According to Newton, any two point mass objects (or spherically symmetric objects of finite extent) exert a force of attraction on one another. This force points along the line of centers joining the objects, is directly proportional to the product of the objects' masses, and inversely proportional to the square of the distance between them. Suppose that the first object is the Sun, which is of mass M, and is located at the origin of our coordinate system. Let the second object be some planet, of mass m, which is located at position vector $\mathbf{r}$. The gravitational force exerted on the planet by the Sun is thus written

$$\mathbf{f} = -\frac{G\,M\,m}{r^3}\,\mathbf{r}. \tag{5.1}$$

The constant of proportionality, G, is called the *gravitational constant*, and takes the value

$$G = 6.67300 \times 10^{-11}\,m^3\,kg^{-1}\,s^{-2}. \tag{5.2}$$

An equal and opposite force to (5.1) acts on the Sun. However, we shall assume that the Sun is so much more massive than the planet in question that this force does not cause the Sun's position to shift appreciably. Hence, the Sun will always remain at the origin of our coordinate system. Likewise, we shall neglect the gravitational forces exerted on our planet by the other planets in the Solar System, compared to the much larger gravitational force exerted by the Sun.

Incidentally, there is something rather curious about Equation (5.1). According to this law, the gravitational force acting on an object is directly proportional to its inertial mass. But why should inertia be related to the force of gravity? After all, inertia measures the reluctance of a given body to deviate from its preferred state of uniform motion in a straight-line, in response to some external force. What has this got to do with gravitational attraction? This question perplexed physicists for many years, and was only answered when Albert Einstein published his general theory of relativity in 1916. According to Einstein, inertial mass acts as a sort of gravitational charge since it impossible to distinguish an acceleration produced by a gravitational field from an apparent acceleration generated by observing in a non-inertial reference frame. The assumption that these two types of acceleration are indistinguishable leads directly to all of the strange predictions of general relativity: *e.g.*, clocks in different gravitational potentials run at different rates, mass bends space, *etc.*

According to Equation (5.1), and Newton's second law, the equation of motion of our planet takes the form

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{G\,M}{r^3}\,\mathbf{r}. \tag{5.3}$$

Note that the planetary mass, m, has canceled out on both sides of the above equation.

5.4 Conservation Laws

Now gravity is a *conservative force*. Hence, the gravitational force (5.1) can be written (see Section 2.5)

$$\mathbf{f} = -\nabla U, \tag{5.4}$$

where the potential energy, $U(\mathbf{r})$, of our planet in the Sun's gravitational field takes the form

$$U(\mathbf{r}) = -\frac{G\,M\,m}{r}. \tag{5.5}$$

It follows that the *total energy* of our planet is a conserved quantity—see Section 2.5. In other words,

$$\mathcal{E} = \frac{v^2}{2} - \frac{G\,M}{r} \tag{5.6}$$

is constant in time. Here, $\mathcal{E}$ is actually the planet's total energy *per unit mass*, and $\mathbf{v} = d\mathbf{r}/dt$.

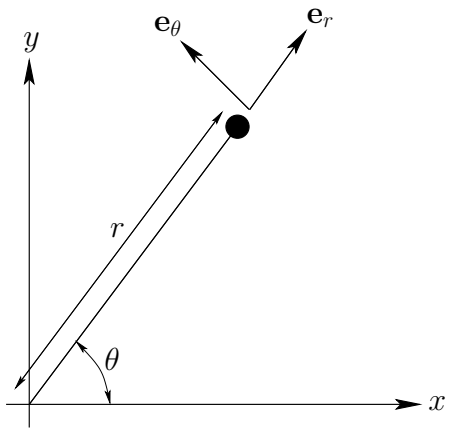


Figure 5.1: *Polar coordinates.*

Gravity is also a *central force*. Hence, the *angular momentum* of our planet is a conserved quantity—see Section 2.6. In other words,

$$\mathbf{h} = \mathbf{r} \times \mathbf{v}, \tag{5.7}$$

which is actually the planet's angular momentum *per unit mass*, is constant in time. Taking the scalar product of the above equation with $\mathbf{r}$, we obtain

$$\mathbf{h} \cdot \mathbf{r} = 0. \tag{5.8}$$

This is the equation of a *plane* which passes through the origin, and whose normal is parallel to $\mathbf{h}$. Since $\mathbf{h}$ is a constant vector, it always points in the *same* direction. We, therefore, conclude that the motion of our planet is *two-dimensional* in nature: *i.e.*, it is confined to some fixed plane which passes through the origin. Without loss of generality, we can let this plane coincide with the x-y plane.

5.5 Polar Coordinates

We can determine the instantaneous position of our planet in the x-y plane in terms of standard Cartesian coordinates, (x, y), or polar coordinates, (r, θ), as illustrated in Figure 5.1. Here, $r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}(y/x)$. It is helpful to define two unit vectors, $\mathbf{e}_r \equiv \mathbf{r}/r$ and $\mathbf{e}_\theta \equiv \mathbf{e}_z \times \mathbf{e}_r$, at the instantaneous position of the planet. The first always points radially away from the origin, whereas the second is normal to the first, in the direction of increasing θ. As is easily demonstrated, the Cartesian components of $\mathbf{e}_r$ and $\mathbf{e}_\theta$ are

$$\mathbf{e}_r = (\cos\theta, \sin\theta), \tag{5.9}$$

$$\mathbf{e}_\theta = (-\sin\theta, \cos\theta), \tag{5.10}$$

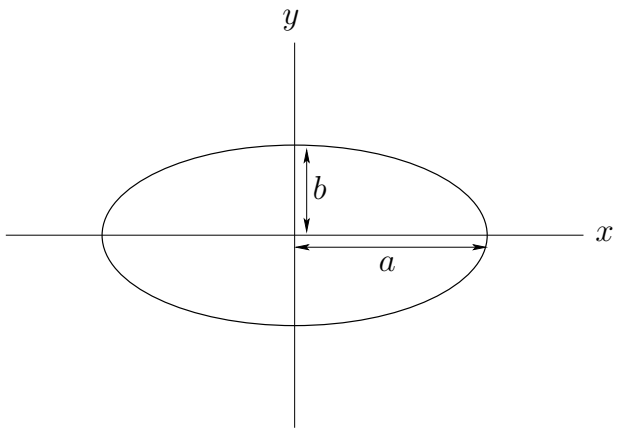


Figure 5.2: *An ellipse.*

respectively.

We can write the position vector of our planet as

$$\mathbf{r} = r\,\mathbf{e}_r. \tag{5.11}$$

Thus, the planet's velocity becomes

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \dot{r}\,\mathbf{e}_r + r\,\dot{\mathbf{e}}_r, \tag{5.12}$$

where $\dot{}$ is shorthand for d/dt. Note that $\mathbf{e}_r$ has a non-zero time-derivative (unlike a Cartesian unit vector) because its direction *changes* as the planet moves around. As is easily demonstrated, from differentiating Equation (5.9) with respect to time,

$$\dot{\mathbf{e}}_r = \dot{\theta}\,(-\sin\theta,\,\cos\theta) = \dot{\theta}\;\mathbf{e}_\theta. \tag{5.13}$$

Thus,

$$\mathbf{v} = \dot{r}\;\mathbf{e}_r + r\,\dot{\theta}\;\mathbf{e}_\theta. \tag{5.14}$$

Now, the planet's acceleration is written

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2} = \ddot{r}\,\mathbf{e}_r + \dot{r}\,\dot{\mathbf{e}}_r + (\dot{r}\,\dot{\theta} + r\,\ddot{\theta})\,\mathbf{e}_\theta + r\,\dot{\theta}\;\dot{\mathbf{e}}_\theta. \tag{5.15}$$

Again, $\mathbf{e}_\theta$ has a non-zero time-derivative because its direction *changes* as the planet moves around. Differentiation of Equation (5.10) with respect to time yields

$$\dot{\mathbf{e}}_\theta = \dot{\theta}\,(-\cos\theta,\,-\sin\theta) = -\dot{\theta}\,\mathbf{e}_r. \tag{5.16}$$

Hence,

$$\mathbf{a} = (\ddot{r} - r\,\dot{\theta}^{\,2})\,\mathbf{e}_r + (r\,\ddot{\theta} + 2\,\dot{r}\,\dot{\theta})\,\mathbf{e}_\theta. \tag{5.17}$$

It follows that the equation of motion of our planet, (5.3), can be written

$$\mathbf{a} = (\ddot{r} - r\,\dot{\theta}^2)\,\mathbf{e}_r + (r\,\ddot{\theta} + 2\,\dot{r}\,\dot{\theta})\,\mathbf{e}_\theta = -\frac{G\,M}{r^2}\,\mathbf{e}_r. \tag{5.18}$$

Since $\mathbf{e}_r$ and $\mathbf{e}_\theta$ are mutually orthogonal, we can separately equate the coefficients of both, in the above equation, to give a *radial equation of motion*,

$$\ddot{r} - r\,\dot{\theta}^2 = -\frac{G\,M}{r^2}, \tag{5.19}$$

and a *tangential equation of motion*,

$$r\,\ddot{\theta} + 2\,\dot{r}\,\dot{\theta} = 0. \tag{5.20}$$

5.6 Conic Sections

The ellipse, the parabola, and the hyperbola are collectively known as *conic sections*, since these three types of curve can be obtained by taking various different plane sections of a right cone. It turns out that the possible solutions of Equations (5.19) and (5.20) are all conic sections. It is, therefore, appropriate for us to briefly review these curves.

An *ellipse*, centered on the origin, of major radius a and minor radius b, which are aligned along the x- and y-axes, respectively (see Figure 5.2), satisfies the following well-known equation:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{5.21}$$

Likewise, a parabola which is aligned along the $+x$-axis, and passes through the origin (see Figure 5.3), satisfies:

$$y^2 - b\,x = 0, \tag{5.22}$$

where $b > 0$.

Finally, a hyperbola which is aligned along the $+x$-axis, and whose asymptotes intersect at the origin (see Figure 5.4), satisfies:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{5.23}$$

Here, a is the distance of closest approach to the origin. The asymptotes subtend an angle $\phi = \tan^{-1}(b/a)$ with the x-axis.

It is not clear, at this stage, what the ellipse, the parabola, and the hyperbola have in common (other than being conic sections). Well, it turns out that what these three curves have in common is that they can all be represented as the locus of a movable point whose distance from a fixed point is in a constant ratio to its perpendicular distance to some fixed straight-line. Let the fixed point (which is termed the *focus* of the ellipse/parabola/hyperbola) lie at the origin, and let the fixed line correspond to $x = -d$

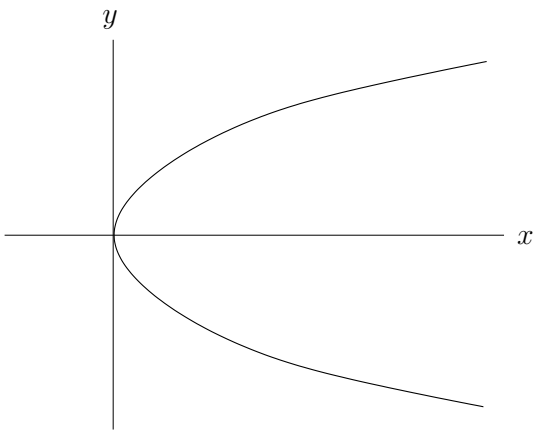


Figure 5.3: *A parabola.*

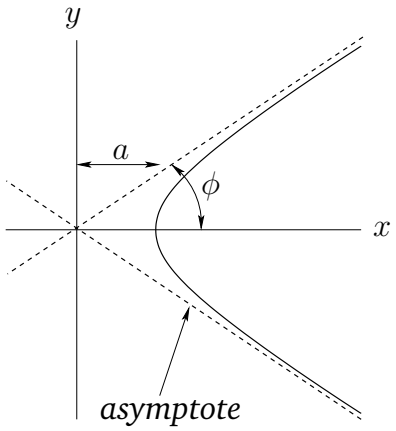


Figure 5.4: *A hyperbola.*

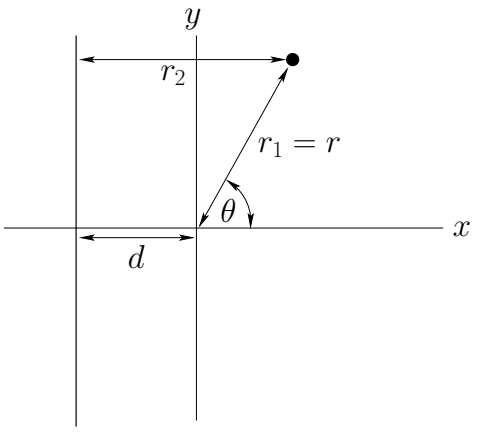


Figure 5.5: *Conic sections in polar coordinates.*

(with $d > 0$). Thus, the distance of a general point (x, y) (which lies to the right of the line $x = -d$) from the origin is $r_1 = \sqrt{x^2 + y^2}$, whereas the perpendicular distance of the point from the line $x = -d$ is $r_2 = x + d$—see Figure 5.5. In polar coordinates, $r_1 = r$ and $r_2 = r\cos\theta + d$. Hence, the locus of a point for which r_1 and r_2 are in a fixed ratio satisfies the following equation:

$$\frac{r_1}{r_2} = \frac{\sqrt{x^2 + y^2}}{x + d} = \frac{r}{r\cos\theta + d} = e, \tag{5.24}$$

where $e \geq 0$ is a constant. When expressed in terms of polar coordinates, the above equation can be rearranged to give

$$r = \frac{r_c}{1 - e\cos\theta}, \tag{5.25}$$

where $r_c = e\,d$.

When written in terms of Cartesian coordinates, (5.24) can be rearranged to give

$$\frac{(x - x_c)^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{5.26}$$

for $e < 1$. Here,

$$a = \frac{r_c}{1 - e^2}, \tag{5.27}$$

$$b = \frac{r_c}{\sqrt{1 - e^2}} = \sqrt{1 - e^2}\,a, \tag{5.28}$$

$$x_c = \frac{e\,r_c}{1 - e^2} = e\,a. \tag{5.29}$$

Equation (5.26) can be recognized as the equation of an *ellipse* whose center lies at $(x_c, 0)$, and whose major and minor radii, a and b, are aligned along the x- and y-axes, respectively [*cf.*, Equation (5.21)].

When again written in terms of Cartesian coordinates, Equation (5.24) can be rearranged to give

$$y^2 - 2\,r_c\,(x - x_c) = 0, \tag{5.30}$$

for $e = 1$. Here, $x_c = -r_c/2$. This is the equation of a *parabola* which passes through the point $(x_c,\ 0)$, and which is aligned along the $+x$-direction [*cf.*, Equation (5.22)].

Finally, when written in terms of Cartesian coordinates, Equation (5.24) can be rearranged to give

$$\frac{(x - x_c)^2}{a^2} - \frac{y^2}{b^2} = 1, \tag{5.31}$$

for $e > 1$. Here,

$$a = \frac{r_c}{e^2 - 1}, \tag{5.32}$$

$$b = \frac{r_c}{\sqrt{e^2 - 1}} = \sqrt{e^2 - 1}\,a, \tag{5.33}$$

$$x_c = -\frac{e\,r_c}{e^2 - 1} = -e\,a. \tag{5.34}$$

Equation (5.31) can be recognized as the equation of a *hyperbola* whose asymptotes intersect at $(x_c,\ 0)$, and which is aligned along the $+x$-direction. The asymptotes subtend an angle

$$\phi = \tan^{-1}\left(\frac{b}{a}\right) = \tan^{-1}(\sqrt{e^2 - 1}) \tag{5.35}$$

with the x-axis [*cf.*, Equation (5.23)].

In conclusion, Equation (5.25) is the polar equation of a *general conic section* which is *confocal with the origin*. For $e < 1$, the conic section is an *ellipse*. For $e = 1$, the conic section is a *parabola*. Finally, for $e > 1$, the conic section is a *hyperbola*.

5.7 Kepler's Second Law

Multiplying our planet's tangential equation of motion, (5.20), by r, we obtain

$$r^2\,\ddot{\theta} + 2\,r\,\dot{r}\,\dot{\theta} = 0. \tag{5.36}$$

However, the above equation can be also written

$$\frac{d(r^2\,\dot{\theta})}{dt} = 0, \tag{5.37}$$

which implies that

$$h = r^2\,\dot{\theta} \tag{5.38}$$

is constant in time. It is easily demonstrated that h is the magnitude of the vector $\mathbf{h}$ defined in Equation (5.7). Thus, the fact that h is constant in time is equivalent to the statement

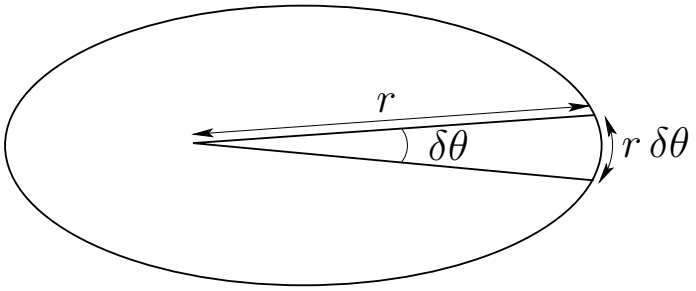


Figure 5.6: *Kepler's second law.*

that the angular momentum of our planet is a constant of its motion. As we have already mentioned, this is the case because gravity is a central force.

Suppose that the radius vector connecting our planet to the origin (*i.e.*, the Sun) sweeps out an angle $\delta\theta$ between times t and $t+\delta t$—see Figure 5.6. The approximately triangular region swept out by the radius vector has the area

$$\delta A \simeq \frac{1}{2}\, r^2\, \delta\theta, \tag{5.39}$$

since the area of a triangle is half its base ($r\,\delta\theta$) times its height (r). Hence, the rate at which the radius vector sweeps out area is

$$\frac{dA}{dt} = \lim_{\delta t\rightarrow 0} \frac{r^2\,\delta\theta}{2\,\delta t} = \frac{r^2}{2}\,\frac{d\theta}{dt} = \frac{h}{2}. \tag{5.40}$$

Thus, the radius vector sweeps out area at a constant rate (since h is constant in time)—this is Kepler's second law. We conclude that Kepler's second law of planetary motion is a direct consequence of *angular momentum conservation*.

5.8 Kepler's First Law

Our planet's radial equation of motion, (5.19), can be combined with Equation (5.38) to give

$$\ddot{r} - \frac{h^2}{r^3} = -\frac{G\,M}{r^2}. \tag{5.41}$$

Suppose that $r = u^{-1}$. It follows that

$$\dot{r} = -\frac{\dot{u}}{u^2} = -r^2\,\frac{du}{d\theta}\,\frac{d\theta}{dt} = -h\,\frac{du}{d\theta}. \tag{5.42}$$

Likewise,

$$\ddot{r} = -h\,\frac{d^2u}{d\theta^2}\,\dot{\theta} = -u^2\,h^2\,\frac{d^2u}{d\theta^2}. \tag{5.43}$$

Hence, Equation (5.41) can be written in the *linear* form

$$\frac{d^2u}{d\theta^2} + u = \frac{G\,M}{h^2}. \tag{5.44}$$

The general solution to the above equation is

$$u(\theta) = \frac{G\,M}{h^2}\,[1 - e\,\cos(\theta - \theta_0)]\,, \tag{5.45}$$

where e and θ_0 are arbitrary constants. Without loss of generality, we can set $\theta_0 = 0$ by rotating our coordinate system about the z-axis. Thus, we obtain

$$r(\theta) = \frac{r_c}{1 - e\,\cos\theta}, \tag{5.46}$$

where

$$r_c = \frac{h^2}{G\,M}. \tag{5.47}$$

We immediately recognize Equation (5.46) as the equation of a conic section which is confocal with the origin (*i.e.*, with the Sun). Specifically, for $e < 1$, Equation (5.46) is the equation of an *ellipse* which is *confocal with the Sun*. Thus, the orbit of our planet around the Sun in a confocal ellipse—this is Kepler's first law of planetary motion. Of course, a planet cannot have a parabolic or a hyperbolic orbit, since such orbits are only appropriate to objects which are ultimately able to escape from the Sun's gravitational field.

5.9 Kepler's Third Law

We have seen that the radius vector connecting our planet to the origin sweeps out area at the constant rate $dA/dt = h/2$ [see Equation (5.40)]. We have also seen that the planetary orbit is an ellipse. Suppose that the major and minor radii of the ellipse are a and b, respectively. It follows that the area of the ellipse is $A = \pi\,a\,b$. Now, we expect the radius vector to sweep out the whole area of the ellipse in a single orbital period, T. Hence,

$$T = \frac{A}{(dA/dt)} = \frac{2\pi\,a\,b}{h}. \tag{5.48}$$

It follows from Equations (5.27), (5.28), and (5.47) that

$$T^2 = \frac{4\pi^2\,a^3}{G\,M}. \tag{5.49}$$

In other words, the *square* of the orbital *period* of our planet is proportional to the *cube* of its orbital *major radius*—this is Kepler's third law.

Note that for an elliptical orbit the closest distance to the Sun—the so-called *perihelion* distance—is [see Equations (5.27) and (5.46)]

$$r_p = \frac{r_c}{1 + e} = a\,(1 - e). \tag{5.50}$$

Likewise, the furthest distance from the Sun—the so-called *aphelion* distance—is

$$r_a = \frac{r_c}{1-e} = a\,(1+e). \tag{5.51}$$

It follows that the major radius, a, is simply the mean of the perihelion and aphelion distances,

$$a = \frac{r_p + r_a}{2}. \tag{5.52}$$

The parameter

$$e = \frac{r_a - r_p}{r_a + r_p} \tag{5.53}$$

is called the *eccentricity*, and measures the deviation of the orbit from circularity. Thus, $e = 0$ corresponds to a circular orbit, whereas $e \rightarrow 1$ corresponds to an infinitely elongated elliptical orbit.

As is easily demonstrated from the above analysis, Kepler laws of planetary motion can be written in the convenient form

$$r = \frac{a\,(1-e^2)}{1 - e\,\cos\theta}, \tag{5.54}$$

$$r^2\,\dot{\theta} = (1-e^2)^{1/2}\,n\,a^2, \tag{5.55}$$

$$G\,M = n^2\,a^3, \tag{5.56}$$

where a is the mean orbital radius (*i.e.*, the major radius), e the orbital eccentricity, and $n = 2\pi/T$ the mean orbital angular velocity.

5.10 Orbital Energies

According to Equations (5.6) and (5.14), the total energy per unit mass of an object in orbit around the Sun is given by

$$\mathcal{E} = \frac{\dot{r}^2 + r^2\,\dot{\theta}^2}{2} - \frac{G\,M}{r}. \tag{5.57}$$

It follows from Equations (5.38), (5.42), and (5.47) that

$$\mathcal{E} = \frac{h^2}{2}\left[\left(\frac{du}{d\theta}\right)^2 + u^2 - 2\,u\,u_c\right], \tag{5.58}$$

where $u = r^{-1}$, and $u_c = r_c^{-1}$. However, according to Equation (5.46),

$$u(\theta) = u_c\,(1 - e\,\cos\theta). \tag{5.59}$$

The previous two equations can be combined with Equations (5.47) and (5.50) to give

$$\mathcal{E} = \frac{u_c^2\, h^2}{2}\,(e^2-1) = \frac{G\,M}{2\,r_p}\,(e-1). \tag{5.60}$$

We conclude that *elliptical* orbits ($e < 1$) have *negative* total energies, whereas *parabolic* orbits ($e = 1$) have *zero* total energies, and *hyperbolic* orbits ($e > 1$) have *positive* total energies. This makes sense, since in a conservative system in which the potential energy at infinity is set to zero [see Equation (5.5)] we expect *bounded* orbits to have *negative* total energies, and *unbounded* orbits to have *positive* total energies—see Section 3.2. Thus, elliptical orbits, which are clearly bounded, should indeed have negative total energies, whereas hyperbolic orbits, which are clearly unbounded, should indeed have positive total energies. Parabolic orbits are marginally bounded (*i.e.*, an object executing a parabolic orbit only just escapes from the Sun's gravitational field), and thus have zero total energy. For the special case of an elliptical orbit, whose major radius a is *finite*, we can write

$$\mathcal{E} = -\frac{G\,M}{2\,a}. \tag{5.61}$$

It follows that the energy of such an orbit is completely determined by the orbital *major radius*.

Consider an artificial satellite in an elliptical orbit around the Sun (the same considerations also apply to satellites in orbit around the Earth). At perihelion, $\dot{r} = 0$, and Equations (5.57) and (5.60) can be combined to give

$$\frac{v_t}{v_c} = \sqrt{1+e}. \tag{5.62}$$

Here, $v_t = r\,\dot{\theta}$ is the satellite's tangential velocity, and $v_c = \sqrt{G\,M/r_p}$ is the tangential velocity that it would need in order to maintain a circular orbit at the perihelion distance. Likewise, at aphelion,

$$\frac{v_t}{v_c} = \sqrt{1-e}, \tag{5.63}$$

where $v_c = \sqrt{G\,M/r_a}$ is now the tangential velocity that the satellite would need in order to maintain a circular orbit at the aphelion distance.

Suppose that our satellite is initially in a circular orbit of radius r_1, and that we wish to transfer it into a circular orbit of radius r_2, where $r_2 > r_1$. We can achieve this by temporarily placing the satellite in an elliptical orbit whose perihelion distance is r_1, and whose aphelion distance is r_2. It follows, from Equation (5.53), that the required eccentricity of the elliptical orbit is

$$e = \frac{r_2 - r_1}{r_2 + r_1}. \tag{5.64}$$

According to Equation (5.62), we can transfer our satellite from its initial circular orbit into the temporary elliptical orbit by increasing its tangential velocity (by briefly switching on the satellite's rocket motor) by a factor

$$\alpha_1 = \sqrt{1+e}. \tag{5.65}$$

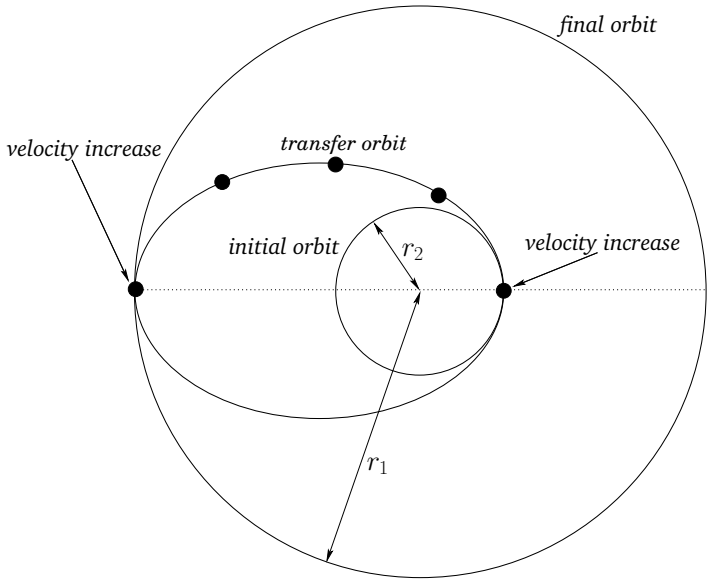


Figure 5.7: *A transfer orbit between two circular orbits.*

We must next allow the satellite to execute half an orbit, so that it attains its aphelion distance, and then boost the tangential velocity by a factor [see Equation (5.63)]

$$\alpha_2 = \frac{1}{\sqrt{1-e}}. \tag{5.66}$$

The satellite will now be in a circular orbit at the aphelion distance, r_2. This process is illustrated in Figure 5.7. Obviously, we can transfer our satellite from a larger to a smaller circular orbit by performing the above process in reverse. Note, finally, from Equation (5.62), that if we increase the tangential velocity of a satellite in a circular orbit about the Sun by a factor greater than $\sqrt{2}$ then we will transfer it into a hyperbolic orbit ($e > 1$), and it will eventually escape from the Sun's gravitational field.

5.11 Kepler Problem

In a nutshell, the so-called *Kepler problem* consists of determining the radial and angular coordinates, r and θ, respectively, of an object in a Keplerian orbit about the Sun as a function of time.

Consider an object in a general Keplerian orbit about the Sun which passes through its perihelion point, $r = r_p$ and $\theta = 0$, at $t = 0$. It follows from the previous analysis that

$$r = \frac{r_p\,(1+e)}{1+e\,\cos\theta}, \tag{5.67}$$

and

$$\mathcal{E} = \frac{\dot{r}^2}{2} + \frac{h^2}{2\,r^2} - \frac{G\,M}{r}, \tag{5.68}$$

where e, $h = \sqrt{G\,M\,r_p\,(1+e)}$, and $\mathcal{E} = G\,M\,(e-1)/(2\,r_p)$ are the orbital eccentricity, angular momentum per unit mass, and energy per unit mass, respectively. The above equation can be rearranged to give

$$\dot{r}^2 = (e-1)\,\frac{G\,M}{r_p} - (e+1)\,\frac{r_p\,G\,M}{r^2} + \frac{2\,G\,M}{r}. \tag{5.69}$$

Taking the square-root, and integrating, we obtain

$$\int_{r_p}^{r} \frac{r\,dr}{[2\,r + (e-1)\,r^2/r_p - (e+1)\,r_p]^{1/2}} = \sqrt{G\,M}\;t. \tag{5.70}$$

Consider an elliptical orbit characterized by $0 < e < 1$. Let us write

$$r = \frac{r_p}{1-e}\,(1 - e\,\cos E), \tag{5.71}$$

where E is termed the *elliptic anomaly*. In fact, E is an angle which varies between $-\pi$ and $+\pi$. Moreover, the perihelion point corresponds to $E = 0$, and the aphelion point to $E = \pi$. Now,

$$dr = \frac{r_p}{1-e}\,e\,\sin E\,dE, \tag{5.72}$$

whereas

$$\begin{aligned} 2\,r + (e-1)\,\frac{r^2}{r_p} - (e+1)\,r_p &= \frac{r_p}{1-e}\,e^2\,(1-\cos^2 E) \\ &= \frac{r_p}{1-e}\,e^2\,\sin^2 E. \end{aligned} \tag{5.73}$$

Thus, Equation (5.70) reduces to

$$\int_0^E (1 - e\,\cos E)\,dE = \left(\frac{G\,M}{a^3}\right)^{1/2} t, \tag{5.74}$$

where $a = r_p/(1-e)$. This equation can immediately be integrated to give

$$E - e\,\sin E = 2\pi\left(\frac{t}{T}\right), \tag{5.75}$$

where $T = 2\pi\,(a^3/GM)^{1/2}$ is the orbital period. Equation (5.75), which is known as *Kepler's equation*, is a transcendental equation which does not possess a simple analytic solution. Fortunately, it is fairly straightforward to solve numerically. For instance, using an iterative approach, if E_n is the nth guess then

$$E_{n+1} = 2\pi\left(\frac{t}{T}\right) + e\,\sin E_n. \tag{5.76}$$

The above iteration scheme converges very rapidly (except in the limit as $e \rightarrow 1$).

Equations (5.67) and (5.71) can be combined to give

$$\cos\theta = \frac{\cos E - e}{1 - e\,\cos E}. \tag{5.77}$$

Thus,

$$1 + \cos\theta = 2\,\cos^2(\theta/2) = \frac{2\,(1-e)\,\cos^2(E/2)}{1 - e\,\cos E}, \tag{5.78}$$

and

$$1 - \cos\theta = 2\,\sin^2(\theta/2) = \frac{2\,(1+e)\,\sin^2(E/2)}{1 - e\,\cos E}. \tag{5.79}$$

The previous two equations imply that

$$\tan(\theta/2) = \left(\frac{1+e}{1-e}\right)^{1/2} \tan(E/2). \tag{5.80}$$

We conclude that, in the case of an elliptical orbit, the solution of the Kepler problem reduces to the solution of the following three equations:

$$E - e\,\sin E = 2\pi\left(\frac{t}{T}\right), \tag{5.81}$$

$$r = a\,(1 - e\,\cos E), \tag{5.82}$$

$$\tan(\theta/2) = \left(\frac{1+e}{1-e}\right)^{1/2} \tan(E/2). \tag{5.83}$$

Here, $T = 2\pi\,(a^3/GM)^{1/2}$ and $a = r_p/(1-e)$. Incidentally, it is clear that if $t \rightarrow t + T$ then $E \rightarrow E + 2\pi$, and $\theta \rightarrow \theta + 2\pi$. In other words, the motion is periodic with period T.

For the case of a parabolic orbit, characterized by $e = 1$, similar analysis to the above yields:

$$P + \frac{P^3}{3} = \left(\frac{G\,M}{2\,r_p^3}\right)^{1/2} t, \tag{5.84}$$

$$r = r_p\,(1 + P^2), \tag{5.85}$$

$$\tan(\theta/2) = P. \tag{5.86}$$

Here, P is termed the *parabolic anomaly*, and varies between $-\infty$ and $+\infty$, with the perihelion point corresponding to $P = 0$. Note that Equation (5.84) is a cubic equation, possessing a single real root, which can, in principle, be solved analytically. However, a numerical solution is generally more convenient.

Finally, for the case of a hyperbolic orbit, characterized by $e > 1$, we obtain:

$$e \sinh H - H = \left(\frac{G M}{a^3}\right)^{1/2} t, \tag{5.87}$$

$$r = a\,(e \cosh H - 1), \tag{5.88}$$

$$\tan(\theta/2) = \left(\frac{e+1}{e-1}\right)^{1/2} \tanh(H/2). \tag{5.89}$$

Here, H is termed the *hyperbolic anomaly*, and varies between $-\infty$ and $+\infty$, with the perihelion point corresponding to $H = 0$. Moreover, $a = r_p/(e-1)$. As in the elliptical case, Equation (5.87) is a transcendental equation which is most easily solved numerically.

5.12 Motion in a General Central Force-Field

Consider the motion of an object in a general (attractive) central force-field characterized by the potential energy *per unit mass* function $V(r)$. Since the force-field is central, it still remains true that

$$h = r^2\,\dot{\theta} \tag{5.90}$$

is a constant of the motion. As is easily demonstrated, Equation (5.44) generalizes to

$$\frac{d^2u}{d\theta^2} + u = -\frac{1}{h^2}\frac{dV}{du}, \tag{5.91}$$

where $u = r^{-1}$.

Suppose, for instance, that we wish to find the potential $V(r)$ which causes an object to execute the spiral orbit

$$r = r_0\,\theta^2. \tag{5.92}$$

Substitution of $u = (r_0\,\theta^2)^{-1}$ into Equation (5.91) yields

$$\frac{dV}{du} = -h^2\left(6\,r_0\,u^2 + u\right). \tag{5.93}$$

Integrating, we obtain

$$V(u) = -h^2\left(2\,r_0\,u^3 + \frac{u^2}{2}\right), \tag{5.94}$$

or

$$V(r) = -h^2\left(\frac{2\,r_0}{r^3} + \frac{1}{2\,r^2}\right). \tag{5.95}$$

In other words, the spiral pattern (5.92) is obtained from a mixture of an inverse-square and inverse-cube potential.

5.13 Motion in a Nearly Circular Orbit

In principle, a circular orbit is a possible orbit for any attractive central force. However, not all forces result in *stable* circular orbits. Let us now consider the stability of circular orbits in a general central force-field. Equation (5.41) generalizes to

$$\ddot{r} - \frac{h^2}{r^3} = f(r), \tag{5.96}$$

where $f(r)$ is the radial force *per unit mass*. For a circular orbit, $\ddot{r} = 0$, and the above equation reduces to

$$-\frac{h^2}{r_c^3} = f(r_c), \tag{5.97}$$

where r_c is the radius of the orbit.

Let us now consider *small* departures from circularity. Let

$$x = r - r_c. \tag{5.98}$$

Equation (5.96) can be written

$$\ddot{x} - \frac{h^2}{(r_c + x)^3} = f(r_c + x). \tag{5.99}$$

Expanding the two terms involving $r_c + x$ as power series in x/r_c, and keeping all terms up to first order, we obtain

$$\ddot{x} - \frac{h^2}{r_c^3}\left(1 - 3\,\frac{x}{r_c}\right) = f(r_c) + f'(r_c)\,x, \tag{5.100}$$

where $'$ denotes a derivative. Making use of Equation (5.97), the above equation reduces to

$$\ddot{x} + \left[-\frac{3\,f(r_c)}{r_c} - f'(r_c)\right] x = 0. \tag{5.101}$$

If the term in square brackets is *positive* then we obtain a simple harmonic equation, which we already know has bounded solutions—*i.e.*, the orbit is *stable* to small perturbations. On the other hand, if the term is square brackets is *negative* then we obtain an equation whose solutions grow exponentially in time—*i.e.*, the orbit is *unstable* to small oscillations. Thus, the stability criterion for a circular orbit of radius r_c in a central force-field characterized by a radial force (per unit mass) function $f(r)$ is

$$f(r_c) + \frac{r_c}{3}\,f'(r_c) < 0. \tag{5.102}$$

For example, consider an attractive power-law force function of the form

$$f(c) = -c\,r^n, \tag{5.103}$$

where $c > 0$. Substituting into the above stability criterion, we obtain

$$-c\, r_c^{\,n} - \frac{c\, n}{3}\, r_c^{\,n} < 0, \tag{5.104}$$

or

$$n > -3. \tag{5.105}$$

We conclude that circular orbits in attractive central force-fields which decay faster than r^{-3} are unstable. The case $n = -3$ is special, since the first-order terms in the expansion of Equation (5.99) cancel out exactly, and it is necessary to retain the second-order terms. Doing this, it is easily demonstrated that circular orbits are also unstable for inverse-cube ($n = -3$) forces.

An *apsis* is a point in an orbit at which the radial distance, r, assumes either a *maximum* or a *minimum* value. Thus, the perihelion and aphelion points are the apsides of planetary orbits. The angle through which the radius vector rotates in going between two consecutive apsides is called the *apsidal angle*. Thus, the apsidal angle for elliptical orbits in an inverse-square force-field is π.

For the case of stable nearly circular orbits, we have seen that r oscillates sinusoidally about its mean value, r_c. Indeed, it is clear from Equation (5.101) that the period of the oscillation is

$$T = \frac{2\pi}{[-3\, f(r_c)/r_c - f'(r_c)]^{1/2}}. \tag{5.106}$$

The apsidal angle is the amount by which θ increases in going between a maximum and a minimum of r. The time taken to achieve this is clearly $T/2$. Now, $\dot{\theta} = h/r^2$, where h is a constant of the motion, and r is almost constant. Thus, $\dot{\theta}$ is approximately constant. In fact,

$$\dot{\theta} \simeq \frac{h}{r_c^2} = \left[-\frac{f(r_c)}{r_c}\right]^{1/2}, \tag{5.107}$$

where use has been made of Equation (5.97). Thus, the apsidal angle, ψ, is given by

$$\psi = \frac{T}{2}\,\dot{\theta} = \pi \left[3 + r_c\, \frac{f'(r_c)}{f(r_c)}\right]^{-1/2} \tag{5.108}$$

For the case of attractive power-law central forces of the form $f(r) = -c\, r^n$, where $c > 0$, the apsidal angle becomes

$$\psi = \frac{\pi}{(3+n)^{1/2}}. \tag{5.109}$$

Now, it should be clear that if an orbit is going to close on itself then the apsidal angle needs to be a *rational* fraction of 2π. There are, in fact, only two small integer values of the power-law index, n, for which this is the case. As we have seen, for an inverse-square force law (*i.e.*, $n = -2$), the apsidal angle is π. For a linear force law (*i.e.*, $n = 1$), the apsidal angle is $\pi/2$—see Section 4.2. However, for quadratic (*i.e.*, $n = 2$) or cubic (*i.e.*,

$n = 3$) force laws, the apsidal angle is an *irrational* fraction of 2π, which means that non-circular orbits in such force-fields never close on themselves.

Let us, finally, calculate the apsidal angle for a nearly circular orbit of radius r_c in a slightly modified (attractive) inverse-square force law of the form

$$f(r) = -\frac{k}{r^2} - \frac{\epsilon}{r^4}, \tag{5.110}$$

where $\epsilon/(k\, r_c^2)$ is small. Substitution into Equation (5.108) yields

$$\psi = \pi \left[3 + r_c\, \frac{2\, k\, r_c^{-3} + 4\, \epsilon\, r_c^{-5}}{-k\, r_c^{-2} - \epsilon\, r_c^{-4}}\right]^{-1/2} = \pi \left[3 - 2\, \frac{1 + 2\, \epsilon/(k\, r_c^2)}{1 + \epsilon/(k\, r_c^2)}\right]^{-1/2}. \tag{5.111}$$

Expanding to first-order in $\epsilon/(k\, r_c^2)$, we obtain

$$\psi \simeq \pi \left(3 - 2\,[1 + \epsilon/(k\, r_c^2)]\right)^{-1/2} = \pi \left[1 - 2\, \epsilon/(k\, r_c^2)\right]^{-1/2} \simeq \pi\,[1 + \epsilon/(k\, r_c^2)]. \tag{5.112}$$

We conclude that if $\epsilon > 0$ then the perihelion (or aphelion) of the orbit *advances* by an angle $2\, \epsilon/(k\, r_c^2)$ every rotation period. It turns out that the general relativistic corrections to Newtonian gravity give rise to a small $1/r^4$ modification (with $\epsilon > 0$) to the Sun's gravitational field. Hence, these corrections generate a small precession in the perihelion of each planet orbiting the Sun. This effect is particularly large for Mercury—see Section 12.13.

5.14 Exercises

5.1. Halley's comet has an orbital eccentricity of $e = 0.967$ and a perihelion distance of $55,000,000$ miles. Find the orbital period, and the comet's speed at perihelion and aphelion.

5.2. A comet is first seen at a distance of d astronomical units (1 astronomical unit is the mean Earth-Sun distance) from the Sun, and is traveling with a speed of q times the Earth's mean speed. Show that the orbit of the comet is hyperbolic, parabolic, or elliptical, depending on whether the quantity $q^2\, d$ is greater than, equal to, or less than 2, respectively.

5.3. Consider a planet in a Keplerian orbit of major radius a and eccentricity e about the Sun. Suppose that the eccentricity of the orbit is small (*i.e.*, $0 < e \ll 1$), as is indeed the case for all of the planets except Mercury and Pluto. Demonstrate that, to first-order in e, the orbit can be approximated as a *circle* whose center is shifted a distance $e\, a$ from the Sun, and that the planet's angular motion appears *uniform* when viewed from a point (called the Equant) which is shifted a distance $2\, e\, a$ from the Sun, in the same direction as the center of the circle. This theorem is the basis of the Ptolomaic model of planetary motion.

5.4. How long (in days) does it take the Sun-Earth radius vector to rotate through 90°, starting at the perihelion point? How long does it take starting at the aphelion point? The period and eccentricity of the Earth's orbit are $T = 365.24$ days, and $e = 0.01673$, respectively.

5.5. Solve the Kepler problem for a parabolic orbit to obtain Equations (5.84)–(5.86).

5.6. Solve the Kepler problem for a hyperbolic orbit to obtain Equations (5.87)–(5.89).

5.7. A comet is in a parabolic orbit lying in the plane of the Earth's orbit. Regarding the Earth's orbit as a circle of radius a, show that the points where the comet intersects the Earth's orbit are given by

$$\cos\theta = -1 + \frac{2p}{a},$$

where p is the perihelion distance of the comet, defined at $\theta = 0$. Show that the time interval that the comet remains inside the Earth's orbit is the faction

$$\frac{2^{1/2}}{3\pi}\left(\frac{2p}{a}+1\right)\left(1-\frac{p}{a}\right)^{1/2}$$

of a year, and that the maximum value of this time interval is $2/3\pi$ year, or about 11 weeks.

5.8. Prove that in the case of a central force varying inversely as the cube of the distance

$$r^2 = A\,t^2 + B\,t + C,$$

where A, B, C are constants.

5.9. The orbit of a particle moving in a central field is a circle passing through the origin, namely $r = r_0\cos\theta$. Show that the force law is inverse-fifth power.

5.10. A particle moving in a central field describes a spiral orbit $r = r_0\exp(k\,\theta)$. Show that the force law is inverse-cube, and that θ varies logarithmically with t. Show that there are two other possible types of orbit in this force-field, and give their equations.

5.11. A particle moves in a spiral orbit given by $r = a\,\theta$. Suppose that θ increases linearly with t. Is the force acting on the particle central in nature? If not, determine how θ would have to vary with t in order to make the force central.

5.12. A particle moves in a circular orbit of radius r_0 in an attractive central force-field of the form $f(r) = -c\,\exp(-r/a)/r^2$, where $c > 0$ and $a > 0$. Demonstrate that the orbit is only stable provided that $r_0 < a$.

5.13. A particle moves in a circular orbit in an attractive central force-field of the form $f(r) = -c\,r^{-3}$, where $c > 0$. Demonstrate that the orbit is unstable to small perturbations.

5.14. If the Solar System were embedded in a uniform dust cloud, what would the apsidal angle of a planet be for motion in a nearly circular orbit? Express your answer in terms of the ratio of the mass of dust contained in a sphere, centered on the Sun, whose radius is that of the orbit, to the mass of the Sun. This model was once suggested as a possible explanation for the advance of the perihelion of Mercury.

5.15. The potential energy per unit mass of a particle in the gravitational field of an oblate spheroid, like the Earth, is

$$V(r) = -\frac{G\,M}{r}\left(1+\frac{\epsilon}{r^2}\right),$$

where r refers to distances in the equatorial plane, M is the Earth's mass, and $\epsilon = (2/5)\,R\,\Delta R$. Here, $R = 4000$ mi is the Earth's equatorial radius, and $\Delta R = 13$ mi the difference between the equatorial and polar radii. Find the apsidal angle for a satellite moving in a nearly circular orbit in the equatorial plane of the Earth.

6 Two-Body Dynamics

6.1 Introduction

This chapter examines the motion of dynamical systems consisting of two freely moving and mutually interacting point objects.

6.2 Reduced Mass

Suppose that our first object is of mass m_1, and is located at position vector $\mathbf{r}_1$. Likewise, our second object is of mass m_2, and is located at position vector $\mathbf{r}_2$. Let the first object exert a force $\mathbf{f}_{21}$ on the second. By Newton's third law, the second object exerts an equal and opposite force, $\mathbf{f}_{12} = -\mathbf{f}_{21}$, on the first. Suppose that there are no other forces in the problem. The equations of motion of our two objects are thus

$$m_1 \frac{d^2\mathbf{r}_1}{dt^2} = -\mathbf{f}, \tag{6.1}$$

$$m_2 \frac{d^2\mathbf{r}_2}{dt^2} = \mathbf{f}, \tag{6.2}$$

where $\mathbf{f} = \mathbf{f}_{21}$.

Now, the center of mass of our system is located at

$$\mathbf{r}_{cm} = \frac{m_1\,\mathbf{r}_1 + m_2\,\mathbf{r}_2}{m_1 + m_2}. \tag{6.3}$$

Hence, we can write

$$\mathbf{r}_1 = \mathbf{r}_{cm} - \frac{m_2}{m_1 + m_2}\,\mathbf{r}, \tag{6.4}$$

$$\mathbf{r}_2 = \mathbf{r}_{cm} + \frac{m_1}{m_1 + m_2}\,\mathbf{r}, \tag{6.5}$$

where $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. Substituting the above two equations into Equations (6.1) and (6.2), and making use of the fact that the center of mass of an isolated system *does not accelerate* (see Section 2.6), we find that both equations yield

$$\mu \frac{d^2\mathbf{r}}{dt^2} = \mathbf{f}, \tag{6.6}$$

where

$$\mu = \frac{m_1\,m_2}{m_1 + m_2} \tag{6.7}$$

is called the *reduced mass*. Hence, we have effectively converted our original two-body problem into an equivalent one-body problem. In the equivalent problem, the force $\mathbf{f}$ is the *same* as that acting on both objects in the original problem (modulo a minus sign). However, the mass, μ, is *different*, and is less than either of m_1 or m_2 (which is why it is called the "reduced" mass).

6.3 Binary Star Systems

Approximately half of the stars in our galaxy are members of so-called *binary star systems*. Such systems consist of two stars orbiting about their common center of mass. The distance separating the stars is always much less than the distance to the nearest neighbour star. Hence, a binary star system can be treated as a two-body dynamical system to a very good approximation.

In a binary star system, the gravitational force which the first star exerts on the second is

$$\mathbf{f} = -\frac{G\, m_1\, m_2}{r^3}\,\mathbf{r}, \tag{6.8}$$

where $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. As we have seen, a two-body system can be reduced to an equivalent one-body system whose equation of motion is of the form (6.6), where $\mu = m_1\, m_2/(m_1 + m_2)$. Hence, in this particular case, we can write

$$\frac{m_1\, m_2}{m_1 + m_2}\,\frac{d^2\mathbf{r}}{dt^2} = -\frac{G\, m_1\, m_2}{r^3}\,\mathbf{r}, \tag{6.9}$$

which gives

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{G\, M}{r^3}\,\mathbf{r}, \tag{6.10}$$

where

$$M = m_1 + m_2. \tag{6.11}$$

Equation (6.10) is identical to Equation (5.3), which we have already solved. Hence, we can immediately write down the solution:

$$\mathbf{r} = (r\,\cos\theta,\, r\,\sin\theta,\, 0), \tag{6.12}$$

where

$$r = \frac{a\,(1-e^2)}{1 - e\,\cos\theta}, \tag{6.13}$$

and

$$\frac{d\theta}{dt} = \frac{h}{r^2}, \tag{6.14}$$

with

$$a = \frac{h^2}{(1-e^2)\,G\,M}. \tag{6.15}$$

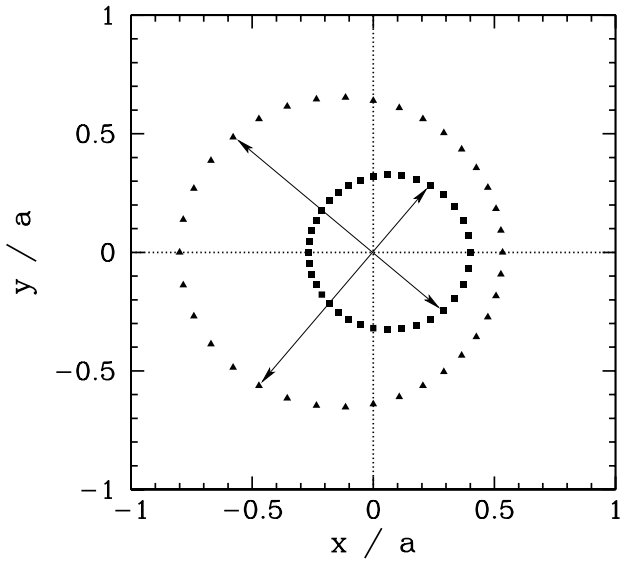


Figure 6.1: *An example binary star orbit.*

Here, h is a constant, and we have aligned our Cartesian axes so that the plane of the orbit coincides with the x-y plane. According to the above solution, the second star executes a Keplerian elliptical orbit, with major radius a and eccentricity e, relative to the first star, and *vice versa*. From Equation (5.49), the period of revolution, T, is given by

$$T = \sqrt{\frac{4\pi^2\, a^3}{G\, M}}. \tag{6.16}$$

In the *inertial* frame of reference whose origin always coincides with the center of mass—the so-called *center of mass frame*—the position vectors of the two stars are

$$\mathbf{r}_1 = -\frac{m_2}{m_1+m_2}\,\mathbf{r}, \tag{6.17}$$

$$\mathbf{r}_2 = \frac{m_1}{m_1+m_2}\,\mathbf{r}, \tag{6.18}$$

where $\mathbf{r}$ is specified above. Figure 6.1 shows an example binary star orbit, in the center of mass frame, calculated with $m_1/m_2 = 0.5$ and $e = 0.2$. Here, the triangles and squares denote the positions of the first and second star, respectively (which are always diagrammatically opposite one another, as indicated by the arrows). It can be seen that both stars execute elliptical orbits about their common center of mass.

Binary star systems have been very useful to astronomers, since it is possible to determine the masses of both stars in such a system by careful observation. The *sum* of the masses of the two stars, $M = m_1 + m_2$, can be found from Equation (6.16) after a measurement of the major radius, a (which is the mean of the greatest and smallest distance apart of the two stars during their orbit), and the orbital period, T. The *ratio* of the masses of

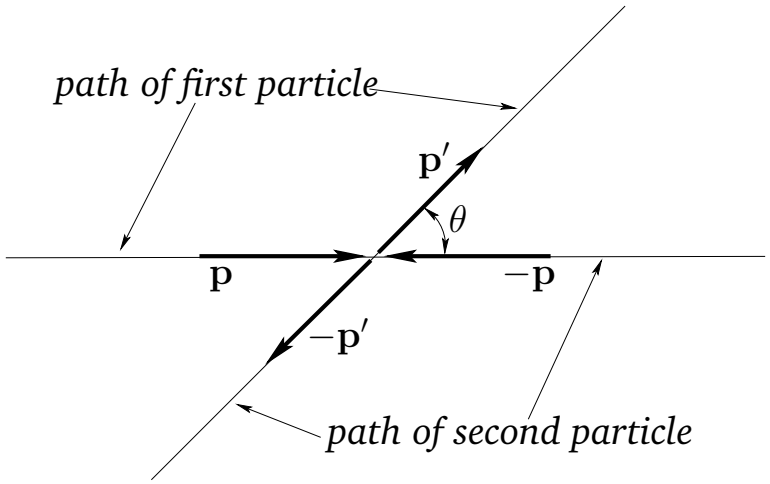


Figure 6.2: *A collision viewed in the center of mass frame.*

the two stars, m_1/m_2, can be determined from Equations (6.17) and (6.18) by observing the fixed ratio of the relative distances of the two stars from the common center of mass about which they both appear to rotate. Obviously, given the sum of the masses, and the ratio of the masses, the individual masses themselves can then be calculated.

6.4 Scattering in the Center of Mass Frame

Let us now consider scattering due to the collision of two particles. We shall restrict our discussion to particles which interact via *conservative central forces*. It turns out that scattering looks particularly simple when viewed in the *center of mass frame*. Let us, therefore, start our investigation by considering two-particle scattering in the center of mass frame.

As before, the first particle is of mass m_1, and is located at position vector $\mathbf{r}_1$, whereas the second particle is of mass m_2, and is located at $\mathbf{r}_2$. By definition, there is *zero net linear momentum* in the center of mass frame at all times. Hence, if the first particle approaches the collision point with momentum $\mathbf{p}$ then the second must approach with momentum $-\mathbf{p}$. Likewise, after the collision, if the first particle recedes from the collision point with momentum $\mathbf{p}'$ then the second must recede with momentum $-\mathbf{p}'$—see Figure 6.2. Furthermore, since the interaction force is *conservative*, the total kinetic energy before and after the collision must be the *same*. It follows that the *magnitude* of the final momentum vector, $\mathbf{p}'$, is equal to the magnitude of the initial momentum vector, $\mathbf{p}$. Because of this, the collision event is completely specified once the angle θ through which the first particle is scattered is given. Of course, in the center of mass frame, the second particle is scattered through the same angle—see Figure 6.2.

Suppose that the two particles interact via the potential $U(r)$, where r is the distance separating the particles. As we have seen, the two-body problem sketched in Figure 6.2 can be converted into the equivalent one-body problem sketched in Figure 6.3. In this

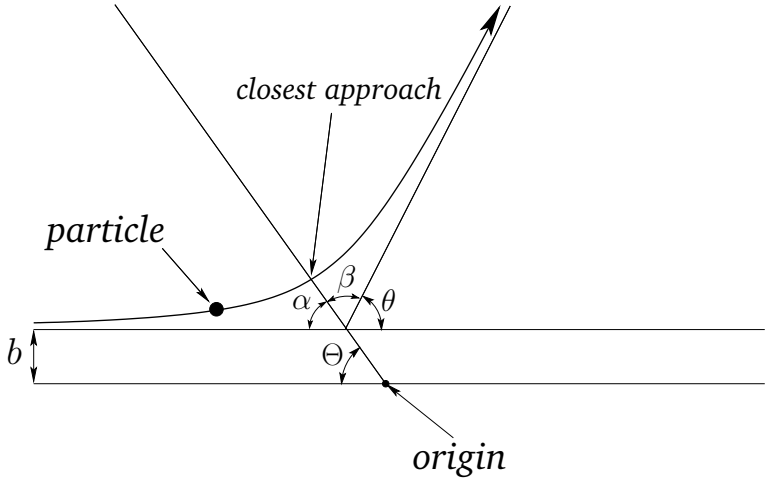


Figure 6.3: *The one-body equivalent to the previous figure.*

equivalent problem, a particle of mass $\mu = m_1\,m_2/(m_1+m_2)$ is scattered in the fixed potential $U(r)$, where r is now the distance from the origin. The vector position $\mathbf{r}$ of the particle in the equivalent problem corresponds to the relative position vector $\mathbf{r}_2-\mathbf{r}_1$ in the original problem. It follows that the angle θ through which the particle is scattered in the equivalent problem is the same as the scattering angle θ in the original problem.

The scattering angle, θ, is largely determined by the so-called *impact parameter*, b, which is the distance of closest approach of the two particles in the *absence* of an interaction potential. In the equivalent problem, b is the distance of closest approach to the origin in the absence of an interaction potential—see Figure 6.3. If $b=0$ then we have a head-on collision. In this case, we expect the two particles to reverse direction after colliding: *i.e.*, we expect $\theta=\pi$. Likewise, if b is large then we expect the two particles to miss one another entirely, in which case $\theta=0$. It follows that the scattering angle, θ, is a *decreasing* function of the impact parameter, b.

Suppose that the polar coordinates of the particle in the equivalent problem are (r,ϑ). Let the particle approach the origin from the direction $\vartheta=0$, and attain its closest distance to the origin when $\vartheta=\Theta$. From symmetry, the angle α in Figure 6.3 is equal to the angle β. However, from simple geometry, $\alpha=\Theta$. Hence,

$$\theta=\pi-2\,\Theta. \tag{6.19}$$

Now, by analogy with Equation (5.58), the conserved total energy E in the equivalent problem, which can easily be shown to be the same as the total energy in the original problem, is given by

$$E=\frac{\mu\,h^2}{2}\left[\left(\frac{du}{d\vartheta}\right)^2+u^2\right]+U(u), \tag{6.20}$$

where $u=r^{-1}$, and h is the angular momentum per unit mass in the equivalent problem.

It is easily seen that

$$h = b\, v_\infty = b \left(\frac{2\,E}{\mu}\right)^{1/2}, \tag{6.21}$$

where v_∞ is the approach velocity in the equivalent problem at large r. It follows that

$$E = E\, b^2 \left[\left(\frac{du}{d\vartheta}\right)^2 + u^2\right] + U(u). \tag{6.22}$$

The above equation can be rearranged to give

$$\frac{d\vartheta}{du} = \frac{b}{\sqrt{1 - b^2\, u^2 - U(u)/E}}. \tag{6.23}$$

Integration yields

$$\Theta = \int_0^{u_{max}} \frac{b\, du}{\sqrt{1 - b^2\, u^2 - U(u)/E}}. \tag{6.24}$$

Here, $u_{max} = 1/r_{min}$, where r_{min} is the distance of closest approach. Since, by symmetry, $(du/d\vartheta)_{u_{max}} = 0$, it follows from Equation (6.23) that

$$1 - b^2\, u_{max}^2 - U(u_{max})/E = 0. \tag{6.25}$$

Equations (6.19) and (6.24) enable us to calculate the function $b(\theta)$ for a given interaction potential, $U(r)$, and a given total energy, E, of the two particles in the center of mass frame. The function $b(\theta)$ tells us which impact parameter corresponds to which scattering angle, and *vice versa*.

Instead of two particles, suppose that we now have two counter-propagating *beams* of identical particles (with the same properties as the two particles described above) which scatter one another via binary collisions. What is the angular distribution of the scattered particles? Consider pairs of particles whose impact parameters lie in the range b to $b + db$. These particles are scattered in such a manner that their scattering angles lie in the range θ to $\theta + d\theta$, where θ is determined from inverting the function $b(\theta)$, and

$$d\theta = \frac{db}{|db(\theta)/d\theta|}. \tag{6.26}$$

Incidentally, we must take the modulus of $db(\theta)/d\theta$ because $b(\theta)$ is a decreasing function of θ. Assuming, as seems reasonable, that the scattering is azimuthally symmetric, the range of *solid angle* into which the particles are scattered is

$$d\Omega = 2\pi \sin\theta\, d\theta = \frac{2\pi \sin\theta\, db}{|db/d\theta|} \tag{6.27}$$

Finally, the *cross-sectional area* of the *annulus* through which incoming particles must pass if they are to have impact parameters in the range b to $b + db$ is

$$d\sigma = 2\pi\, b\, db. \tag{6.28}$$

The previous two equations allow us to define the *differential scattering cross-section*:

$$\frac{d\sigma}{d\Omega} = \frac{b}{\sin\theta}\left|\frac{db}{d\theta}\right| \tag{6.29}$$

The differential scattering cross-section has units of area per steradian, and specifies the effective target area for scattering into a given range of solid angle. For two *uniform* beams scattering off one another, the differential scattering cross-section thus effectively specifies the *probability* of scattering into a given range of solid angle. The *total scattering cross-section* is the integral of the differential cross-section over all solid angles,

$$\sigma = \int \frac{d\sigma}{d\Omega}\, d\Omega, \tag{6.30}$$

and measures the effective target area for scattering in *any* direction. Thus, if the flux of particles per unit area per unit time, otherwise known as the *intensity*, of the two beams is I, then the number of particles of a given type scattered per unit time is simply $I\,\sigma$.

Let us now calculate the scattering cross-section for the following very simple interaction potential:

$$U(r) = \begin{cases} 0 & r > a \\ \infty & r \leq a \end{cases}. \tag{6.31}$$

This is the interaction potential of *impenetrable spheres* which only exert a force on one another when they are in physical contact (*e.g.*, billiard balls). If the particles in the first beam have radius R_1, and the particles in the second beam have radius R_2, then $a = R_1 + R_2$. In other words, the centers of two particles, one from either beam, can never be less than a distance a apart, where a is the sum of their radii (since the particles are impenetrable spheres).

Equations (6.19), (6.24), and (6.31) yield

$$\theta = \pi - 2\int_0^{1/a} \frac{b\,du}{\sqrt{1 - b^2\,u^2}} = \pi - 2\,\sin^{-1}(b/a). \tag{6.32}$$

The above formula can be rearranged to give

$$b(\theta) = a\,\cos(\theta/2). \tag{6.33}$$

Note that

$$b\left|\frac{db}{d\theta}\right| = \frac{1}{2}\left|\frac{db^2}{d\theta}\right| = \frac{a^2}{2}\,\sin(\theta/2)\,\cos(\theta/2) = \frac{a^2}{4}\,\sin\theta. \tag{6.34}$$

Hence, Equations (6.29) and (6.34) yield

$$\frac{d\sigma}{d\Omega} = \frac{a^2}{4}. \tag{6.35}$$

We thus conclude that when two beams of impenetrable spheres collide, in the center of mass frame, the particles in the two beams have an equal probability of being scattered in any direction. The total scattering cross-section is

$$\sigma = \int \frac{d\sigma}{d\Omega}\, d\Omega = \pi\, a^2. \tag{6.36}$$

Obviously, this result makes a lot of sense—the total scattering cross-section for two impenetrable spheres is simply the area of a circle whose radius is the sum of the radii of the two spheres.

Let us now consider scattering by an inverse-square interaction force whose potential takes the form

$$U(r) = \frac{k}{r}. \tag{6.37}$$

It follows from Equations (6.24) and (6.25) that

$$\Theta = \int_0^{u_{max}} \frac{b\, du}{\sqrt{1 - b^2\, u^2 - k\, u/E}} = \int_0^{x_{max}} \frac{dx}{\sqrt{1 - x^2 - \alpha\, x}}, \tag{6.38}$$

where $\alpha = k/(E\, b)$, and

$$1 - x_{max}^2 - \alpha\, x_{max} = 0. \tag{6.39}$$

Integration yields

$$\Theta = \frac{\pi}{2} - \sin^{-1}\left(\frac{\alpha}{\sqrt{4 + \alpha^2}}\right). \tag{6.40}$$

Hence, from Equation (6.19), we obtain

$$\theta = 2\, \sin^{-1}\left(\frac{\alpha}{\sqrt{4 + \alpha^2}}\right). \tag{6.41}$$

The above equation can be rearranged to give

$$b^2 = \frac{k^2}{4\, E^2}\, \cot^2(\theta/2). \tag{6.42}$$

Thus,

$$2\, b \left|\frac{db}{d\theta}\right| = \frac{k^2}{8\, E^2} \frac{\sin\theta}{\sin^4(\theta/2)}. \tag{6.43}$$

Finally, using Equation (6.29), we get

$$\frac{d\sigma}{d\Omega} = \frac{k^2}{16\, E^2} \frac{1}{\sin^4(\theta/2)}. \tag{6.44}$$

There are a number of things to note about the above formula. First, the scattering cross-section is proportional to k^2. This means that *repulsive* ($k > 0$) and *attractive* ($k < 0$)

inverse-square interaction forces of the same strength give rise to *identical* angular distributions of scattered particles. Second, the scattering cross-section is proportional to E^{-2}. This means that inverse-square interaction forces are much more effective at scattering low energy, rather than high energy, particles. Finally, the differential scattering cross-section is proportional to $\sin^{-4}(\theta/2)$. This means that, with an inverse-square interaction force, the overwhelming majority of "collisions" consist of *small angle* scattering events (*i.e.*, $\theta \ll 1$).

Let us now consider a specific case. Suppose that we have particles of electric charge q scattering off particles of the same charge. The interaction potential due to the Coulomb force between the particles is simply

$$U(r) = \frac{q^2}{4\pi\,\epsilon_0\,r}. \tag{6.45}$$

Thus, it follows from Equation (6.44) [with $k = q^2/(4\pi\,\epsilon_0)$] that the differential scattering cross-section takes the form

$$\frac{d\sigma}{d\Omega} = \frac{q^4}{16\,(4\,\pi\,\epsilon_0)^2\,E^2}\frac{1}{\sin^4(\theta/2)}. \tag{6.46}$$

This very famous formula is known as the *Rutherford scattering cross-section*, since it was first derived by Earnst Rutherford for use in his celebrated α-particle scattering experiment.

Note, finally, that if we try to integrate the Rutherford formula to obtain the *total* scattering cross-section then we find that the integral is *divergent*, due to the very strong increase in $d\sigma/d\Omega$ as $\theta \rightarrow 0$. This implies that the Coulomb potential (or any other inverse-square-law potential) has an effectively *infinite* range. In practice, however, an electric charge is generally surrounded by charges of the opposite sign which *shield* the Coulomb potential of the charge beyond a certain distance. This shielding effect allows the charge to have a *finite* total scattering cross-section (for the scattering of other electric charges). However, the total scattering cross-section of the charge depends (albeit, logarithmically) on the shielding distance, and, hence, on the nature and distribution of the charges surrounding it.

6.5 Scattering in the Laboratory Frame

We have seen that two-particle scattering looks fairly simple when viewed in the center of mass frame. Unfortunately, we are not usually in a position to do this. In the laboratory, the most common scattering scenario is one in which the second particle is initially at rest. Let us now investigate this situation.

Suppose that, in the center of mass frame, the first particle has velocity $\mathbf{v}_1$ before the collision, and velocity $\mathbf{v}_1'$ after the collision. Likewise, the second particle has velocity $\mathbf{v}_2$ before the collision, and $\mathbf{v}_2'$ after the collision. We know that

$$m_1\,\mathbf{v}_1 + m_2\,\mathbf{v}_2 = m_1\,\mathbf{v}_1' + m_2\,\mathbf{v}_2' = \mathbf{0} \tag{6.47}$$

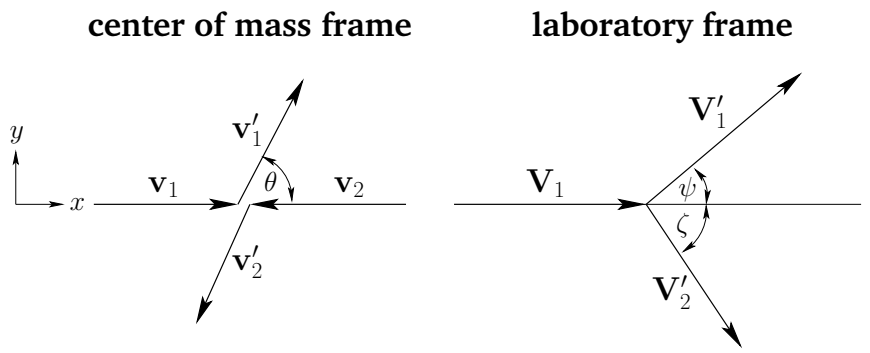


Figure 6.4: *Scattering in the center of mass and laboratory frames.*

in the center of mass frame. Moreover, since the collision is assumed to be elastic (*i.e.*, energy conserving),

$$v_1' = v_1, \tag{6.48}$$

$$v_2' = v_2. \tag{6.49}$$

Let us transform to a new inertial frame of reference—which we shall call the *laboratory frame*—which is moving with the uniform velocity $-\mathbf{v}_2$ with respect to the center of mass frame. In the new reference frame, the first particle has initial velocity $\mathbf{V}_1 = \mathbf{v}_1 - \mathbf{v}_2$, and final velocity $\mathbf{V}_1' = \mathbf{v}_1' - \mathbf{v}_2$. Furthermore, the second particle is initially at *rest*, and has the final velocity $\mathbf{V}_2' = \mathbf{v}_2' - \mathbf{v}_2$. The relationship between scattering in the center of mass frame and scattering in the laboratory frame is illustrated in Figure 6.4.

In the center of mass frame, both particles are scattered through the same angle θ. However, in the laboratory frame, the first and second particles are scattered by the (generally different) angles ψ and ζ, respectively.

Defining x- and y-axes, as indicated in Figure 6.4, it is easily seen that the Cartesian components of the various velocity vectors in the two frames of reference are:

$$\mathbf{v}_1 = v_1\,(1,\,0), \tag{6.50}$$

$$\mathbf{v}_2 = (m_1/m_2)\,v_1\,(-1,\,0), \tag{6.51}$$

$$\mathbf{v}_1' = v_1\,(\cos\theta,\,\sin\theta), \tag{6.52}$$

$$\mathbf{v}_2' = (m_1/m_2)\,v_1(-\cos\theta,\,-\sin\theta), \tag{6.53}$$

$$\mathbf{V}_1 = (1 + m_1/m_2)\,v_1\,(1,\,0), \tag{6.54}$$

$$\mathbf{V}_1' = v_1\,(\cos\theta + m_1/m_2,\,\sin\theta), \tag{6.55}$$

$$\mathbf{V}_2' = (m_1/m_2)\,v_1\,(1-\cos\theta,\,-\sin\theta). \tag{6.56}$$

In the center of mass frame, let E be the total energy, let $E_1 = (1/2)\,m_1\,v_1^2$ and $E_2 = (1/2)\,m_2\,v_2^2$ be the kinetic energies of the first and second particles, respectively, before the collision, and let $E_1' = (1/2)\,m_1\,v_1'^2$ and $E_2' = (1/2)\,m_2\,v_2'^2$ be the kinetic energies of the first

and second particles, respectively, after the collision. Of course, $E = E_1 + E_2 = E_1' + E_2'$. In the laboratory frame, let $\mathcal{E}$ be the total energy. This is, of course, equal to the kinetic energy of the first particle before the collision. Likewise, let $\mathcal{E}_1' = (1/2)\, m_1\, V_1'^2$ and $\mathcal{E}_2' = (1/2)\, m_2\, V_2'^2$ be the kinetic energies of the first and second particles, respectively, after the collision. Of course, $\mathcal{E} = \mathcal{E}_1' + \mathcal{E}_2'$.

The following results can easily be obtained from the above definitions and Equations (6.50)–(6.56). First,

$$\mathcal{E} = \left(\frac{m_1 + m_2}{m_2}\right) E. \tag{6.57}$$

Hence, the total energy in the laboratory frame is always *greater* than that in the center of mass frame. In fact, it can be demonstrated that the total energy in the center of mass frame is less than the total energy in *any* other inertial frame. Second,

$$E_1 = E_1' = \left(\frac{m_2}{m_1 + m_2}\right) E, \tag{6.58}$$

$$E_2 = E_2' = \left(\frac{m_1}{m_1 + m_2}\right) E. \tag{6.59}$$

These equations specify how the total energy in the center of mass frame is distributed between the two particles. Note that this distribution is *unchanged* by the collision. Finally,

$$\mathcal{E}_1' = \left[\frac{m_1^2 + 2\, m_1\, m_2\, \cos\theta + m_2^2}{(m_1 + m_2)^2}\right] \mathcal{E}, \tag{6.60}$$

$$\mathcal{E}_2' = \left[\frac{2\, m_1\, m_2\, (1 - \cos\theta)}{(m_1 + m_2)^2}\right] \mathcal{E}. \tag{6.61}$$

These equations specify how the total energy in the laboratory frame is distributed between the two particles after the collision. Note that the energy distribution in the laboratory frame is *different* before and after the collision.

Equations (6.50)–(6.56), and some simple trigonometry, yield

$$\tan\psi = \frac{\sin\theta}{\cos\theta + m_1/m_2}, \tag{6.62}$$

and

$$\tan\zeta = \frac{\sin\theta}{1 - \cos\theta} = \tan\left(\frac{\pi}{2} - \frac{\theta}{2}\right). \tag{6.63}$$

The last equation implies that

$$\zeta = \frac{\pi}{2} - \frac{\theta}{2}. \tag{6.64}$$

Differentiating Equation (6.62) with respect to θ, we obtain

$$\frac{d\tan\psi}{d\theta} = \frac{1 + (m_1/m_2)\, \cos\theta}{(\cos\theta + m_1/m_2)^2}. \tag{6.65}$$

Thus, $\tan\psi$ attains an extreme value, which can be shown to correspond to a *maximum* possible value of ψ, when the numerator of the above expression is zero: *i.e.*, when

$$\cos\theta = -\frac{m_2}{m_1}. \tag{6.66}$$

Note that it is only possible to solve the above equation when $m_1 > m_2$. If this is the case then Equation (6.62) yields

$$\tan\psi_{max} = \frac{m_2/m_1}{\sqrt{1-(m_2/m_1)^2}}, \tag{6.67}$$

which reduces to

$$\psi_{max} = \sin^{-1}\left(\frac{m_2}{m_1}\right). \tag{6.68}$$

Hence, we conclude that when $m_1 > m_2$ there is a *maximum* possible value of the scattering angle, ψ, in the laboratory frame. This maximum value is always less than $\pi/2$, which implies that there is *no backward scattering* (*i.e.*, $\psi > \pi/2$) at all when $m_1 > m_2$. For the special case when $m_1 = m_2$, the maximum scattering angle is $\pi/2$. However, for $m_1 < m_2$ there is no maximum value, and the scattering angle in the laboratory frame can thus range all the way to π.

Equations (6.57)–(6.64) enable us to relate the particle energies and scattering angles in the laboratory frame to those in the center of mass frame. In general, these relationships are fairly complicated. However, there are two special cases in which the relationships become much simpler.

The first special case is when $m_2 \gg m_1$. In this limit, it is easily seen from Equations (6.57)–(6.64) that the second mass is *stationary* both before and after the collision, and that the center of mass frame *coincides* with the laboratory frame (since the energies and scattering angles in the two frames are the same). Hence, the simple analysis outlined in Section 6.4 is applicable in this case.

The second special case is when $m_1 = m_2$. In this case, Equation (6.62) yields

$$\tan\psi = \frac{\sin\theta}{\cos\theta + 1} = \tan(\theta/2). \tag{6.69}$$

Hence,

$$\psi = \frac{\theta}{2}. \tag{6.70}$$

In other words, the scattering angle of the first particle in the laboratory frame is *half* of the scattering angle in the center of mass frame. The above equation can be combined with Equation (6.64) to give

$$\psi + \zeta = \frac{\pi}{2}. \tag{6.71}$$

Thus, in the laboratory frame, the two particles move off at *right-angles* to one another after the collision. Equation (6.57) yields

$$\mathcal{E} = 2\,\mathrm{E}. \tag{6.72}$$

In other words, the total energy in the laboratory frame is *twice* that in the center of mass frame. According to Equations (6.58) and (6.59),

$$E_1 = E_1' = E_2 = E_2' = \frac{E}{2}. \tag{6.73}$$

Hence, the total energy in the center of mass frame is divided *equally* between the two particles. Finally, Equations (6.60) and (6.61) give

$$\mathcal{E}_1' = \left(\frac{1+\cos\theta}{2}\right)\mathcal{E} = \cos^2(\theta/2)\ \mathcal{E} = \cos^2\psi\ \mathcal{E}, \tag{6.74}$$

$$\mathcal{E}_2' = \left(\frac{1-\cos\theta}{2}\right)\mathcal{E} = \sin^2(\theta/2)\ \mathcal{E} = \sin^2\psi\ \mathcal{E}. \tag{6.75}$$

Thus, in the laboratory frame, the unequal energy distribution between the two particles after the collision is simply related to the scattering angle ψ.

What is the angular distribution of scattered particles when a beam of particles of the first type scatter off stationary particles of the second type? Well, we can define a differential scattering cross-section, $d\sigma(\psi)/d\Omega'$, in the laboratory frame, where $d\Omega' = 2\pi\sin\psi\,d\psi$ is an element of solid angle in this frame. Thus, $(d\sigma(\psi)/d\Omega')\,d\Omega'$ is the effective cross-sectional area in the laboratory frame for scattering into the range of scattering angles ψ to $\psi + d\psi$. Likewise, $(d\sigma(\theta)/d\Omega)\,d\Omega$ is the effective cross-sectional area in the center of mass frame for scattering into the range of scattering angles θ to $\theta + d\theta$. Note that $d\Omega = 2\pi\sin\theta\,d\theta$. However, a cross-sectional area is not changed when we transform between different inertial frames. Hence, we can write

$$\frac{d\sigma(\psi)}{d\Omega'}\,d\Omega' = \frac{d\sigma(\theta)}{d\Omega}\,d\Omega, \tag{6.76}$$

provided that ψ and θ are related via Equation (6.62). This equation can be rearranged to give

$$\frac{d\sigma(\psi)}{d\Omega'} = \frac{d\Omega}{d\Omega'}\,\frac{d\sigma(\theta)}{d\Omega}, \tag{6.77}$$

or

$$\frac{d\sigma(\psi)}{d\Omega'} = \frac{\sin\theta}{\sin\psi}\,\frac{d\theta}{d\psi}\,\frac{d\sigma(\theta)}{d\Omega}. \tag{6.78}$$

The above equation allows us to relate the differential scattering cross-section in the laboratory frame to that in the center of mass frame. In general, this relationship is extremely complicated. However, for the special case where the masses of the two types of particles are *equal*, we have seen that $\psi = \theta/2$ [see Equation (6.70)]. Hence, it follows from Equation (6.78) that

$$\frac{d\sigma(\psi)}{d\Omega'} = 4\,\cos\psi\,\frac{d\sigma(\theta = 2\,\psi)}{d\Omega}. \tag{6.79}$$

Let us now consider some specific examples. We saw earlier that, in the center of mass frame, the differential scattering cross-section for *impenetrable spheres* is [see Equation (6.35)]

$$\frac{d\sigma(\theta)}{d\Omega} = \frac{a^2}{4}, \tag{6.80}$$

where a is the sum of the radii. According to Equation (6.79), the differential scattering cross-section (for equal mass spheres) in the laboratory frame is

$$\frac{d\sigma(\psi)}{d\Omega'} = a^2 \cos\psi. \tag{6.81}$$

Note that this cross-section is *negative* for $\psi > \pi/2$. This just tells us that there is *no scattering* with scattering angles greater than $\pi/2$ (*i.e.*, there is no backward scattering). Comparing Equations (6.80) and (6.81), we can see that the scattering is *isotropic* in the center of mass frame, but appears concentrated in the forward direction in the laboratory frame. We can integrate Equation (6.81) over all solid angles to obtain the total scattering cross-section in the laboratory frame. Note that we only integrate over angular regions where the differential scattering cross-section is *positive*. Doing this, we get

$$\sigma = \pi a^2, \tag{6.82}$$

which is the same as the total scattering cross-section in the center of mass frame [see Equation (6.36)]. This is a general result. The total scattering cross-section is *frame independent*, since a cross-sectional area is not modified by switching between different frames of reference.

As we have seen, the Rutherford scattering cross-section takes the form [see Equation (6.46)]

$$\frac{d\sigma}{d\Omega} = \frac{q^4}{16\,(4\pi\epsilon_0)^2\,E^2}\frac{1}{\sin^4(\theta/2)} \tag{6.83}$$

in the center of mass frame. It follows, from Equation (6.79), that the Rutherford scattering cross-section (for equal mass particles) in the laboratory frame is written

$$\frac{d\sigma}{d\Omega'} = \frac{q^4}{(4\pi\epsilon_0)^2\,\mathcal{E}^2}\frac{\cos\psi}{\sin^4\psi}. \tag{6.84}$$

Here, we have made use of the fact that $\mathcal{E} = 2E$ for equal mass particles [see Equation (6.72)]. Note, again, that this cross-section is negative for $\psi > \pi/2$, indicating the absence of backward scattering.

6.6 Exercises

6.1. A particle subject to a repulsive force varying as $1/r^3$ is projected from infinity with a velocity that would carry it to a distance a from the center of force, if it were directed toward the latter. Actually, it is projected along a line whose closest distance from the center of force would be b if there were no repulsion. Prove that the particle's least distance from the center is $\sqrt{a^2+b^2}$, and that the angle between the two asymptotes of its path is $\pi\,b/\sqrt{a^2+b^2}$.

6.2. A particle subject to a repulsive force varying as $1/r^5$ is projected from infinity with a velocity V that would carry it to a distance a from the center of force, if it were directed toward the latter. Actually, it is projected along a line whose closest distance from the center of force would be b if there were no repulsion. Show that the least velocity of the particle is

$$V\,\frac{b^2}{a^2}\left[\left(\frac{a^4}{b^4}+\frac{1}{4}\right)^{1/2}-\frac{1}{2}\right]^{1/2}.$$

6.3. Using the notation of Section 6.2, show that the angular momentum of a two-body system takes the form

$$\mathbf{L} = M\,\mathbf{r}_{cm}\times\dot{\mathbf{r}}_{cm}+\mu\,\mathbf{r}\times\dot{\mathbf{r}},$$

where $M = m_1 + m_2$.

6.4. Consider the case of Rutherford scattering in the event that $m_1 \gg m_2$. Demonstrate that the differential scattering cross-section in the laboratory frame is approximately

$$\frac{d\sigma}{d\Omega'}\simeq\frac{q_1^2\,q_2^2}{4\,(4\pi\epsilon_0)^2\,\mathcal{E}^2}\,\frac{\psi_{max}^{-4}}{[1-\sqrt{1-(\psi/\psi_{max})^2}]^2\;[1-(\psi/\psi_{max})^2]^{1/2}},$$

where $\psi_{max} = m_2/m_1$.

6.5. Show that the energy distribution of particles recoiling from an elastic collision is always directly proportional to the differential scattering cross-section in the center of mass frame.

6.6. It is found experimentally that in the elastic scattering of neutrons by protons ($m_n \simeq m_p$) at relatively low energies the energy distribution of the recoiling protons in the laboratory frame is constant up to a maximum energy, which is the energy of the incident neutrons. What is the angular distribution of the scattering in the center of mass frame?

6.7. The most energetic α-particles available to Earnst Rutherford and his colleagues for the famous Rutherford scattering experiment were 7.7 MeV. For the scattering of 7.7 MeV α-particles from Uranium 238 nuclei (initially at rest) at a scattering angle in the laboratory frame of 90°, find the following (in the laboratory frame, unless otherwise specified):

(a) The recoil scattering angle of the Uranium nucleus.

(b) The scattering angles of the α-particle and Uranium nucleus in the center of mass frame.

(c) The kinetic energies of the scattered α-particle and Uranium nucleus (in MeV).

(d) The impact parameter, b.

(e) The distance of closest approach.

(f) The differential scattering cross-section at 90°.

6.8. Consider scattering by the repulsive potential $U = k/r^2$ (where $k > 0$) viewed in the center of mass frame. Demonstrate that the differential scattering cross-section is

$$\frac{d\sigma}{d\Omega}=\frac{k}{E\,\pi}\,\frac{1}{\sin\theta}\,\frac{(1-\theta/\pi)}{(\theta/\pi)^2\,(2-\theta/\pi)^2}.$$

7 Rotating Reference Frames

7.1 Introduction

As we saw in Chapter 2, Newton's second law of motion is only valid in *inertial* frames of reference. Unfortunately, we are sometimes forced to observe motion in *non-inertial* reference frames. For instance, it is most convenient for us to observe the motions of the objects in our immediate vicinity in a reference frame which is *fixed* relative to the *surface of the Earth*. Such a frame is *non-inertial* in nature, since it *accelerates* with respect to a standard inertial frame due to the Earth's daily rotation about its axis. (Note that the accelerations of this frame due to the Earth's orbital motion about the Sun, or the Sun's orbital motion about the galactic center, *etc*., are negligible compared to the acceleration due to the Earth's axial rotation.) Let us now investigate motion in a *rotating* reference frame.

7.2 Rotating Reference Frames

Suppose that a given object has position vector $\mathbf{r}$ in some *non-rotating inertial* reference frame. Let us observe the motion of this object in a *non-inertial* reference frame which *rotates* with *constant* angular velocity $\boldsymbol{\Omega}$ about an axis passing through the origin of the inertial frame. Suppose, first of all, that our object appears *stationary* in the rotating reference frame. Hence, in the non-rotating frame, the object's position vector $\mathbf{r}$ will appear to *precess* about the origin with angular velocity $\boldsymbol{\Omega}$. It follows, from Equation (A.49), that in the non-rotating reference frame

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \boldsymbol{\Omega} \times \mathbf{r}. \tag{7.1}$$

Suppose, now, that our object appears to move in the rotating reference frame with instantaneous velocity $\mathbf{v}'$. It is fairly obvious that the appropriate generalization of the above equation is simply

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \mathbf{v}' + \boldsymbol{\Omega} \times \mathbf{r}. \tag{7.2}$$

Let and $\mathrm{d}/\mathrm{d}t$ and $\mathrm{d}/\mathrm{d}t'$ denote apparent time derivatives in the non-rotating and rotating frames of reference, respectively. Since an object which is stationary in the rotating reference frame appears to move in the non-rotating frame, it is clear that $\mathrm{d}/\mathrm{d}t \neq \mathrm{d}/\mathrm{d}t'$. Writing the apparent velocity, $\mathbf{v}'$, of our object in the rotating reference frame as $\mathrm{d}\mathbf{r}/\mathrm{d}t'$, the above equation takes the form

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t'} + \boldsymbol{\Omega} \times \mathbf{r}, \tag{7.3}$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t'} + \boldsymbol{\Omega}\times, \tag{7.4}$$

since $\mathbf{r}$ is a general position vector. Equation (7.4) expresses the relationship between apparent time derivatives in the non-rotating and rotating reference frames.

Operating on the general position vector $\mathbf{r}$ with the time derivative (7.4), we get

$$\mathbf{v} = \mathbf{v}' + \boldsymbol{\Omega} \times \mathbf{r}. \tag{7.5}$$

This equation relates the apparent velocity, $\mathbf{v} = \mathrm{d}\mathbf{r}/\mathrm{d}t$, of an object with position vector $\mathbf{r}$ in the non-rotating reference frame to its apparent velocity, $\mathbf{v}' = \mathrm{d}\mathbf{r}/\mathrm{d}t'$, in the rotating reference frame.

Operating twice on the position vector $\mathbf{r}$ with the time derivative (7.4), we obtain

$$\mathbf{a} = \left(\frac{\mathrm{d}}{\mathrm{d}t'} + \boldsymbol{\Omega}\times\right)(\mathbf{v}' + \boldsymbol{\Omega} \times \mathbf{r}), \tag{7.6}$$

or

$$\mathbf{a} = \mathbf{a}' + \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) + 2\,\boldsymbol{\Omega} \times \mathbf{v}'. \tag{7.7}$$

This equation relates the apparent acceleration, $\mathbf{a} = \mathrm{d}^2\mathbf{r}/\mathrm{d}t^2$, of an object with position vector $\mathbf{r}$ in the non-rotating reference frame to its apparent acceleration, $\mathbf{a}' = \mathrm{d}^2\mathbf{r}/\mathrm{d}t'^2$, in the rotating reference frame.

Applying Newton's second law of motion in the inertial (*i.e.*, non-rotating) reference frame, we obtain

$$m\,\mathbf{a} = \mathbf{f}. \tag{7.8}$$

Here, m is the mass of our object, and $\mathbf{f}$ is the (non-fictitious) force acting on it. Note that these quantities are the *same* in both reference frames. Making use of Equation (7.7), the apparent equation of motion of our object in the rotating reference frame takes the form

$$m\,\mathbf{a}' = \mathbf{f} - m\,\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) - 2\,m\,\boldsymbol{\Omega} \times \mathbf{v}'. \tag{7.9}$$

The last two terms in the above equation are so-called "fictitious forces". Such forces are always needed to account for motion observed in non-inertial reference frames. Note that fictitious forces can always be distinguished from non-fictitious forces in Newtonian dynamics because the former have no associated reactions. Let us now investigate the two fictitious forces appearing in Equation (7.9).

7.3 Centrifugal Acceleration

Let our non-rotating inertial frame be one whose origin lies at the center of the Earth, and let our rotating frame be one whose origin is fixed with respect to some point, of latitude λ, on the Earth's surface—see Figure 7.1. The latter reference frame thus rotates with respect to the former (about an axis passing through the Earth's center) with an angular

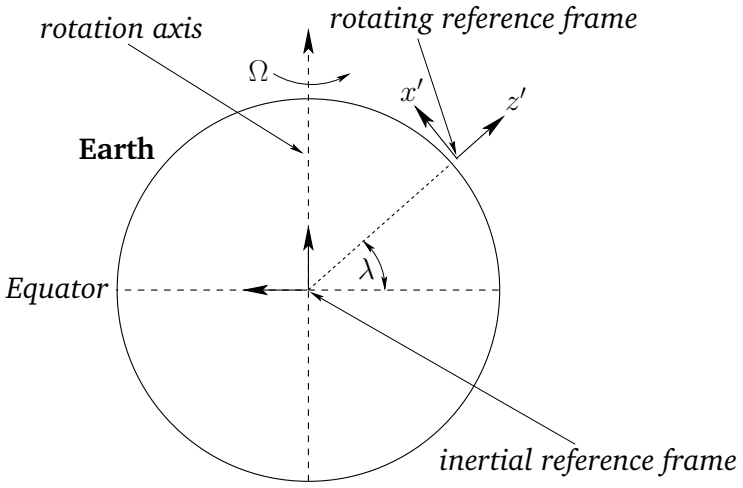


Figure 7.1: *Inertial and non-inertial reference frames.*

velocity vector, $\boldsymbol{\Omega}$, which points from the center of the Earth toward its north pole, and is of magnitude

$$\Omega = \frac{2\pi}{24 \text{ hrs}} = 7.27 \times 10^{-5} \text{ rad./s}. \tag{7.10}$$

Consider an object which appears stationary in our rotating reference frame: *i.e.*, an object which is stationary with respect to the Earth's surface. According to Equation (7.9), the object's apparent equation of motion in the rotating frame takes the form

$$m\,\mathbf{a}' = \mathbf{f} - m\,\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}). \tag{7.11}$$

Let the non-fictitious force acting on our object be the force of gravity, $\mathbf{f} = m\,\mathbf{g}$. Here, the local gravitational acceleration, $\mathbf{g}$, points directly toward the center of the Earth. It follows, from the above, that the apparent gravitational acceleration in the rotating frame is written

$$\mathbf{g}' = \mathbf{g} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R}), \tag{7.12}$$

where $\mathbf{R}$ is the displacement vector of the origin of the rotating frame (which lies on the Earth's surface) with respect to the center of the Earth. Here, we are assuming that our object is situated relatively close to the Earth's surface (*i.e.*, $\mathbf{r} \simeq \mathbf{R}$).

It can be seen, from Equation (7.12), that the apparent gravitational acceleration of a stationary object close to the Earth's surface has two components. First, the true gravitational acceleration, $\mathbf{g}$, of magnitude $g \simeq 9.8\,\text{m/s}^2$, which always points directly toward the center of the Earth. Second, the so-called *centrifugal acceleration*, $-\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R})$. This acceleration is normal to the Earth's axis of rotation, and always points directly away from this axis. The magnitude of the centrifugal acceleration is $\Omega^2\,\rho = \Omega^2\,R\,\cos\lambda$, where ρ is the perpendicular distance to the Earth's rotation axis, and $R = 6.37 \times 10^6\,\text{m}$ is the Earth's radius—see Figure 7.2.

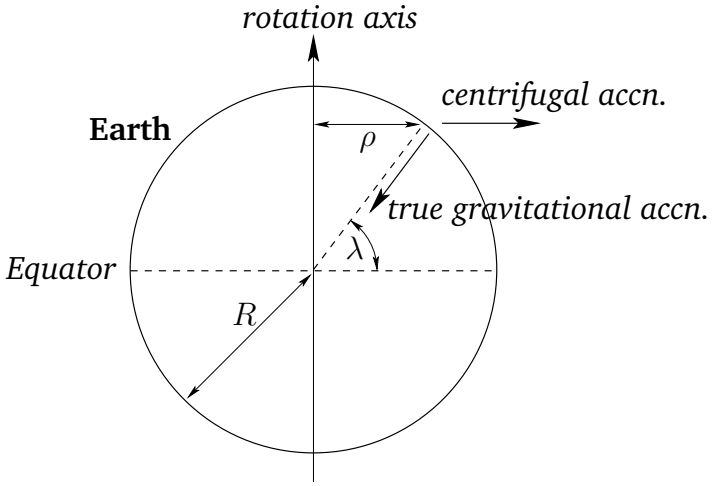


Figure 7.2: *Centrifugal acceleration.*

It is convenient to define Cartesian axes in the rotating reference frame such that the z'-axis points vertically upward, and x'- and y'-axes are horizontal, with the x'-axis pointing directly northward, and the y'-axis pointing directly westward—see Figure 7.1. The Cartesian components of the Earth's angular velocity are thus

$$\mathbf{\Omega} = \Omega\,(\cos\lambda,\, 0,\, \sin\lambda), \tag{7.13}$$

whilst the vectors $\mathbf{R}$ and $\mathbf{g}$ are written

$$\mathbf{R} = (0,\, 0,\, R), \tag{7.14}$$

$$\mathbf{g} = (0,\, 0,\, -g), \tag{7.15}$$

respectively. It follows that the Cartesian coordinates of the apparent gravitational acceleration, (7.12), are

$$\mathbf{g}' = \left(-\Omega^2\, R\, \cos\lambda\, \sin\lambda,\, 0,\, -g + \Omega^2\, R\, \cos^2\lambda\right). \tag{7.16}$$

The magnitude of this acceleration is approximately

$$g' \simeq g - \Omega^2\, R\, \cos^2\lambda \simeq 9.8 - 0.034\, \cos^2\lambda\ \ \mathrm{m/s^2}. \tag{7.17}$$

According to the above equation, the centrifugal acceleration causes the magnitude of the apparent gravitational acceleration on the Earth's surface to *vary* by about 0.3%, being largest at the poles, and smallest at the equator. This variation in apparent gravitational acceleration, due (ultimately) to the Earth's rotation, causes the Earth itself to *bulge* slightly at the equator (see Section 12.6), which has the effect of further intensifying the variation, since a point on the surface of the Earth at the equator is slightly further away from the Earth's center than a similar point at one of the poles (and, hence, the true gravitational acceleration is slightly weaker in the former case).

Another consequence of centrifugal acceleration is that the apparent gravitational acceleration on the Earth's surface has a *horizontal* component aligned in the north/south direction. This horizontal component ensures that the apparent gravitational acceleration *does not* point directly toward the center of the Earth. In other words, a plumb-line on the surface of the Earth does not point vertically downward, but is deflected slightly away from a true vertical in the north/south direction. The angular deviation from true vertical can easily be calculated from Equation (7.16):

$$\theta_{dev} \simeq -\frac{\Omega^2\, R}{2\, g}\, \sin(2\,\lambda) \simeq -0.1^\circ\, \sin(2\,\lambda). \tag{7.18}$$

Here, a positive angle denotes a northward deflection, and *vice versa*. Thus, the deflection is *southward* in the *northern hemisphere* (*i.e.*, $\lambda > 0$) and *northward* in the *southern hemisphere* (*i.e.*, $\lambda < 0$). The deflection is zero at the poles and at the equator, and reaches its maximum magnitude (which is very small) at middle latitudes.

7.4 Coriolis Force

We have now accounted for the first fictitious force, $-m\,\boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\mathbf{r})$, in Equation (7.9). Let us now investigate the second, which takes the form $-2\,m\,\boldsymbol{\Omega}\times\mathbf{v}'$, and is called the *Coriolis force*. Obviously, this force only affects objects which are *moving* in the rotating reference frame.

Consider a particle of mass m free-falling under gravity in our rotating reference frame. As before, we define Cartesian axes in the rotating frame such that the z'-axis points vertically upward, and the x'- and y'-axes are horizontal, with the x'-axis pointing directly northward, and the y'-axis pointing directly westward. It follows, from Equation (7.9), that the Cartesian equations of motion of the particle in the rotating reference frame take the form:

$$\ddot{x}' = 2\,\Omega\,\sin\lambda\,\dot{y}', \tag{7.19}$$

$$\ddot{y}' = -2\,\Omega\,\sin\lambda\,\dot{x}' + 2\,\Omega\,\cos\lambda\,\dot{z}', \tag{7.20}$$

$$\ddot{z}' = -g - 2\,\Omega\,\cos\lambda\,\dot{y}'. \tag{7.21}$$

Here, $\dot{} \equiv d/dt$, and g is the local acceleration due to gravity. In the above, we have neglected the centrifugal acceleration, for the sake of simplicity. This is reasonable, since the only effect of the centrifugal acceleration is to slightly modify the magnitude and direction of the local gravitational acceleration. We have also neglected air resistance, which is less reasonable.

Consider a particle which is dropped (at $t = 0$) from rest a height h above the Earth's surface. The following solution method exploits the fact that the Coriolis force is much smaller in magnitude that the force of gravity: hence, Ω can be treated as a *small* parameter. To lowest order (*i.e.*, neglecting Ω), the particle's vertical motion satisfies $\ddot{z}' = -g$,

which can be solved, subject to the initial conditions, to give

$$z' = h - \frac{g\,t^2}{2}. \tag{7.22}$$

Substituting this expression into Equations (7.19) and (7.20), neglecting terms involving Ω^2, and solving subject to the initial conditions, we obtain $x' \simeq 0$, and

$$y' \simeq -g\,\Omega\,\cos\lambda\,\frac{t^3}{3}. \tag{7.23}$$

In other words, the particle is deflected *eastward* (*i.e.*, in the negative y'-direction). Now, the particle hits the ground when $t \simeq \sqrt{2\,h/g}$. Hence, the net eastward deflection of the particle as strikes the ground is

$$d_{east} = \frac{\Omega}{3}\,\cos\lambda\,\left(\frac{8\,h^3}{g}\right)^{1/2}. \tag{7.24}$$

Note that this deflection is in the *same* direction as the Earth's rotation (*i.e.*, west to east), and is greatest at the equator, and zero at the poles. A particle dropped from a height of 100 m at the equator is deflected by about 2.2 cm.

Consider a particle launched *horizontally* with some fairly large velocity

$$\mathbf{V} = V_0\,(\cos\theta, -\sin\theta, 0). \tag{7.25}$$

Here, θ is the *compass bearing* of the velocity vector (so north is 0°, east is 90°, *etc.*). Neglecting any vertical motion, Equations (7.19) and (7.20) yield

$$\dot{v}_{x'} \simeq -2\,\Omega\,V_0\,\sin\lambda\,\sin\theta, \tag{7.26}$$

$$\dot{v}_{y}' \simeq -2\,\Omega\,V_0\,\sin\lambda\,\cos\theta, \tag{7.27}$$

which can be integrated to give

$$v_{x'} \simeq V_0\,\cos\theta - 2\,\Omega\,V_0\,\sin\lambda\,\sin\theta\,t, \tag{7.28}$$

$$v_{y'} \simeq -V_0\,\sin\theta - 2\,\Omega\,V_0\,\sin\lambda\,\cos\theta\,t. \tag{7.29}$$

To lowest order in Ω, the above equations are equivalent to

$$v_{x'} \simeq V_0\,\cos(\theta + 2\,\Omega\,\sin\lambda\,t), \tag{7.30}$$

$$v_{y'} \simeq -V_0\,\sin(\theta + 2\,\Omega\,\sin\lambda\,t). \tag{7.31}$$

If follows that the Coriolis force causes the compass bearing of the particle's velocity vector to *rotate* steadily as time progresses. The rotation rate is

$$\frac{d\theta}{dt} \simeq 2\,\Omega\,\sin\lambda. \tag{7.32}$$

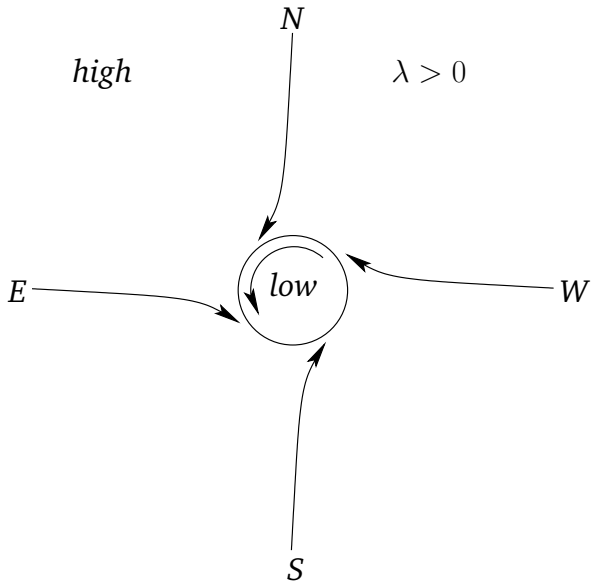


Figure 7.3: *A cyclone in the northern hemisphere.*

Hence, the rotation is *clockwise* (looking from above) in the *northern hemisphere*, and *counter-clockwise* in the *southern hemisphere*. The rotation rate is zero at the equator, and greatest at the poles.

The Coriolis force has a significant effect on terrestrial weather patterns. Near equatorial regions, the intense heating of the Earth's surface due to the Sun results in hot air rising. In the northern hemisphere, this causes cooler air to move in a southerly direction toward the equator. The Coriolis force deflects this moving air in a clockwise sense (looking from above), resulting in the *trade winds*, which blow toward the *southwest*. In the southern hemisphere, the cooler air moves northward, and is deflected by the Coriolis force in a counter-clockwise sense, resulting in trade winds which blow toward the *northwest*.

Furthermore, as air flows from high to low pressure regions, the Coriolis force deflects the air in a clockwise/counter-clockwise manner in the northern/southern hemisphere, producing *cyclonic* rotation—see Figure 7.3. It follows that cyclonic rotation is *counter-clockwise* in the *northern hemisphere*, and *clockwise* in the *southern hemisphere*. Thus, this is the direction of rotation of tropical storms (*e.g.*, hurricanes, typhoons) in each hemisphere.

7.5 Foucault Pendulum

Consider a pendulum consisting of a compact mass m suspended from a light cable of length l in such a manner that the pendulum is free to oscillate in any plane whose normal is parallel to the Earth's surface. The mass is subject to three forces: first, the force of gravity $\mathrm{m}\,\mathbf{g}$, which is directed vertically downward (we are again ignoring centrifugal acceleration); second, the tension $\mathbf{T}$ in the cable, which is directed upward along the

cable; and, third, the Coriolis force. It follows that the apparent equation of motion of the mass, in a frame of reference which co-rotates with the Earth, is [see Equation (7.9)]

$$m\,\ddot{\mathbf{r}}' = m\,\mathbf{g} + \mathbf{T} - 2\,m\,\boldsymbol{\Omega}\times\dot{\mathbf{r}}'. \tag{7.33}$$

Let us define our usual Cartesian coordinates (x', y', z'), and let the origin of our coordinate system correspond to the equilibrium position of the mass. If the pendulum cable is deflected from the downward vertical by a *small* angle θ then it is easily seen that $x' \sim l\,\theta$, $y' \sim l\,\theta$, and $z' \sim l\,\theta^2$. In other words, the change in height of the mass, z', is negligible compared to its horizontal displacement. Hence, we can write $z' \simeq 0$, provided that $\theta \ll 1$. The tension $\mathbf{T}$ has the vertical component $T\cos\theta \simeq T$, and the horizontal component $\mathbf{T}_{hz} = -T\sin\theta\,\mathbf{r}'/r' \simeq -T\,\mathbf{r}'/l$, since $\sin\theta \simeq r'/l$—see Figure 7.4. Hence, the Cartesian equations of motion of the mass are written [*cf.*, Equations (7.19)–(7.21)]

$$\ddot{x}' = -\frac{T}{l\,m}\,x' + 2\,\Omega\,\sin\lambda\,\dot{y}', \tag{7.34}$$

$$\ddot{y}' = -\frac{T}{l\,m}\,y' - 2\,\Omega\,\sin\lambda\,\dot{x}', \tag{7.35}$$

$$0 = \frac{T}{m} - g - 2\,\Omega\,\cos\lambda\,\dot{y}'. \tag{7.36}$$

To lowest order in Ω (*i.e.*, neglecting Ω), the final equation, which is just vertical force balance, yields $T \simeq m\,g$. Hence, Equations (7.34) and (7.35) reduce to

$$\ddot{x}' \simeq -\frac{g}{l}\,x' + 2\,\Omega\,\sin\lambda\,\dot{y}', \tag{7.37}$$

$$\ddot{y}' \simeq -\frac{g}{l}\,y' - 2\,\Omega\,\sin\lambda\,\dot{x}'. \tag{7.38}$$

Let

$$s = x' + i\,y'. \tag{7.39}$$

Equations (7.37) and (7.38) can be combined to give a single complex equation for s:

$$\ddot{s} = -\frac{g}{l}\,s - i\,2\,\Omega\,\sin\lambda\,\dot{s}. \tag{7.40}$$

Let us look for a sinusoidally oscillating solution of the form

$$s = s_0\,e^{-i\,\omega\,t}. \tag{7.41}$$

Here, ω is the (real) angular frequency of oscillation, and s_0 is an arbitrary complex constant. Equations (7.40) and (7.41) yield the following quadratic equation for ω:

$$\omega^2 - 2\,\Omega\,\sin\lambda\,\omega - \frac{g}{l} = 0. \tag{7.42}$$

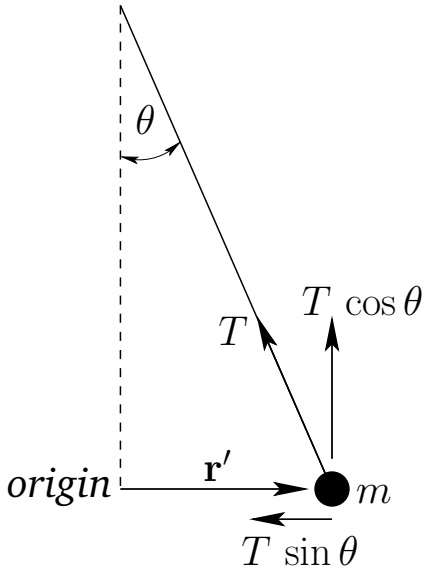


Figure 7.4: *The Foucault pendulum.*

The solutions are approximately

$$\omega_\pm \simeq \Omega\, \sin\lambda \pm \sqrt{\frac{g}{l}}, \tag{7.43}$$

where we have neglected terms involving Ω^2. Hence, the general solution of (7.41) takes the form

$$s = s_+\, e^{-i\,\omega_+\, t} + s_-\, e^{-i\,\omega_-\, t}, \tag{7.44}$$

where s_+ and s_- are two arbitrary complex constants.

Making the specific choice $s_+ = s_- = a/2$, where a is real, the above solution reduces to

$$s = a\, e^{-i\,\Omega\, \sin\lambda\, t}\, \cos\left(\sqrt{\frac{g}{l}}\, t\right). \tag{7.45}$$

Now, it is clear from Equation (7.39) that x' and y' are the real and imaginary parts of s, respectively. Thus, it follows from the above that

$$x' = a\, \cos(\Omega\, \sin\lambda\, t)\, \cos\left(\sqrt{\frac{g}{l}}\, t\right), \tag{7.46}$$

$$y' = -a\, \sin(\Omega\, \sin\lambda\, t)\, \cos\left(\sqrt{\frac{g}{l}}\, t\right). \tag{7.47}$$

These equations describe sinusoidal oscillations, in a plane whose normal is parallel to the Earth's surface, at the standard pendulum frequency $\sqrt{g/l}$. The Coriolis force, however, causes the plane of oscillation to slowly *precess* at the angular frequency $\Omega\, \sin\lambda$. The period of the precession is

$$T = \frac{2\pi}{\Omega\, \sin\lambda} = \frac{24}{\sin\lambda}\ \text{hrs}. \tag{7.48}$$

For example, according to the above equations, the pendulum oscillates in the x'-direction (*i.e.*, north/south) at $t \simeq 0$, in the y'-direction (*i.e.*, east/west) at $t \simeq T/4$, in the x'-direction again at $t \simeq T/2$, *etc.* The precession is *clockwise* (looking from above) in the *northern hemisphere*, and *counter-clockwise* in the *southern hemisphere*.

The precession of the plane of oscillation of a pendulum, due to the Coriolis force, is used in many museums and observatories to demonstrate that the Earth is rotating. This method of making the Earth's rotation manifest was first devised by Foucault in 1851.

7.6 Exercises

7.1. A pebble is dropped down an elevator shaft in the Empire State Building ($h = 1250$ ft, latitude $= 41^\circ$ N). Find the pebble's horizontal deflection (magnitude and direction) due to the Coriolis force at the bottom of the shaft. Neglect air resistance.

7.2. If a bullet is fired due east, at an elevation angle α, from a point on the Earth whose latitude is $+\lambda$ show that it will strike the Earth with a lateral deflection $4\,\Omega\,v_0^3\,\sin\lambda\,\sin^2\alpha\,\cos\alpha/g^2$. Is the deflection northward or southward? Here, Ω is the Earth's angular velocity, v_0 is the bullet's initial speed, and g is the acceleration due to gravity. Neglect air resistance.

7.3. A particle is thrown vertically with initial speed v_0, reaches a maximum height, and falls back to the ground. Show that the horizontal Coriolis deflection of the particle when it returns to the ground is opposite in direction, and four times greater in magnitude, than the Coriolis deflection when it is dropped at rest from the same maximum height. Neglect air resistance.

7.4. The surface of the Diskworld is a disk which rotates (counter-clockwise looking down) with angular frequency Ω about a perpendicular axis passing through its center. Diskworld gravitational acceleration is of magnitude g, and is everywhere directed normal to the disk. A projectile is launched from the surface of the disk at a point whose radial distance from the axis of rotation is R. The initial velocity of the projectile (in a co-rotating frame) is of magnitude v_0, is directly radially outwards, and is inclined at an angle α to the horizontal. What are the radial and tangential displacements of the impact point from that calculated by neglecting the centrifugal and Coriolis forces? Neglect air resistance. You may assume that the displacements are small compared to both R and the horizontal range of the projectile.

7.5. Demonstrate that the Coriolis force causes conical pendulums to rotate clockwise and counter-clockwise with slightly different angular frequencies. What is the frequency difference as a function of terrestrial latitude?

7.6. A satellite is in a circular orbit of radius a about the Earth. Let us define a set of co-moving Cartesian coordinates, centered on the satellite, such that the x-axis always points toward the center of the Earth, the y-axis in the direction of the satellite's orbital motion, and the z-axis in the direction of the satellite's orbital angular velocity, $\boldsymbol{\omega}$. Demonstrate that the equation of motion of a small mass in orbit about the satellite are

$$\begin{aligned} \ddot{x} &= 3\,\omega^2\,x + 2\,\omega\,\dot{y}, \\ \ddot{y} &= -2\,\omega\,\dot{x}, \end{aligned}$$

assuming that $|x|/a \ll 1$ and $|y|/a \ll 1$. You may neglect the gravitational attraction between the satellite and the mass. Show that the mass executes a retrograde (*i.e.*, in the opposite sense to the satellite's orbital rotation) elliptical orbit about the satellite whose period matches that of the satellite's orbit, and whose major and minor axes are in the ratio $2:1$, and are aligned along the y- and x-axes, respectively.

8 Rigid Body Rotation

8.1 Introduction

This chapter examines the rotation of rigid bodies in three dimensions.

8.2 Fundamental Equations

We can think of a rigid body as a collection of a large number of small mass elements which all maintain a fixed spatial relationship with respect to one another. Let there be N elements, and let the ith element be of mass m_i, and instantaneous position vector $\mathbf{r}_i$. The equation of motion of the ith element is written

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \sum_{j=1,N}^{j\neq i} \mathbf{f}_{ij} + \mathbf{F}_i. \tag{8.1}$$

Here, $\mathbf{f}_{ij}$ is the internal force exerted on the ith element by the jth element, and $\mathbf{F}_i$ the external force acting on the ith element. The internal forces $\mathbf{f}_{ij}$ represent the stresses which develop within the body in order to ensure that its various elements maintain a constant spatial relationship with respect to one another. Of course, $\mathbf{f}_{ij} = -\mathbf{f}_{ji}$, by Newton's third law. The external forces represent forces which originate outside the body.

Repeating the analysis of Section 2.7, we can sum Equation (8.1) over all mass elements to obtain

$$M \frac{d^2\mathbf{r}_{cm}}{dt^2} = \mathbf{F}. \tag{8.2}$$

Here, $M = \sum_{i=1,N} m_i$ is the total mass, $\mathbf{r}_{cm}$ the position vector of the center of mass [see Equation (2.27)], and $\mathbf{F} = \sum_{i=1,N} \mathbf{F}_i$ the total external force. It can be seen that the center of mass of a rigid body moves under the action of the external forces like a point particle whose mass is identical with that of the body.

Again repeating the analysis of Section 2.7, we can sum $\mathbf{r}_i\times$ Equation (8.1) over all mass elements to obtain

$$\frac{d\mathbf{L}}{dt} = \mathbf{T}. \tag{8.3}$$

Here, $\mathbf{L} = \sum_{i=1,N} m_i\, \mathbf{r}_i \times d\mathbf{r}_i/dt$ is the total angular momentum of the body (about the origin), and $\mathbf{T} = \sum_{i=1,N} \mathbf{r}_i \times \mathbf{F}_i$ the total external torque (about the origin). Note that the above equation is only valid if the internal forces are *central* in nature. However, this is not a particularly onerous constraint. Equation (8.3) describes how the angular momentum of a rigid body evolves in time under the action of the external torques.

In the following, we shall only consider the *rotational* motion of rigid bodies, since their translational motion is similar to that of point particles [see Equation (8.2)], and, therefore, fairly straightforward in nature.

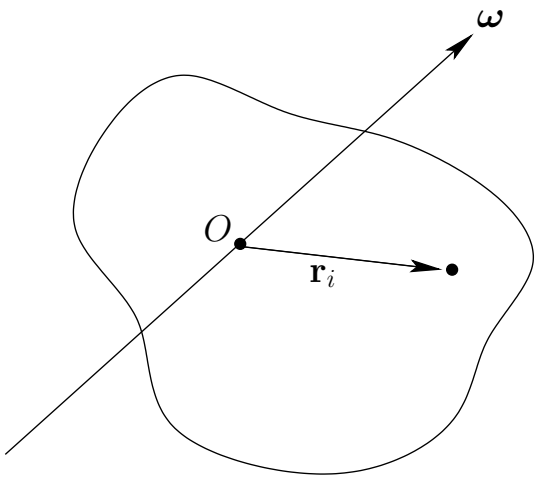


Figure 8.1: *A rigid rotating body.*

8.3 Moment of Inertia Tensor

Consider a rigid body rotating with fixed angular velocity $\boldsymbol{\omega}$ about an axis which passes through the origin—see Figure 8.1. Let $\mathbf{r}_i$ be the position vector of the ith mass element, whose mass is m_i. We expect this position vector to *precess* about the axis of rotation (which is parallel to $\boldsymbol{\omega}$) with angular velocity ω. It, therefore, follows from Equation (A.49) that

$$\frac{d\mathbf{r}_i}{dt} = \boldsymbol{\omega} \times \mathbf{r}_i. \tag{8.4}$$

Thus, the above equation specifies the velocity, $\mathbf{v}_i = d\mathbf{r}_i/dt$, of each mass element as the body rotates with fixed angular velocity $\boldsymbol{\omega}$ about an axis passing through the origin.

The total angular momentum of the body (about the origin) is written

$$\mathbf{L} = \sum_{i=1,N} m_i\, \mathbf{r}_i \times \frac{d\mathbf{r}_i}{dt} = \sum_{i=1,N} m_i\, \mathbf{r}_i \times (\boldsymbol{\omega} \times \mathbf{r}_i) = \sum_{i=1,N} m_i \left[r_i^2\, \boldsymbol{\omega} - (\mathbf{r}_i \cdot \boldsymbol{\omega})\, \mathbf{r}_i \right], \tag{8.5}$$

where use has been made of Equation (8.4), and some standard vector identities (see Section A.10). The above formula can be written as a matrix equation of the form

$$\begin{pmatrix} L_x \\ L_y \\ L_z \end{pmatrix} = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}, \tag{8.6}$$

where

$$I_{xx} = \sum_{i=1,N} (y_i^2 + z_i^2)\, m_i = \int (y^2 + z^2)\, dm, \tag{8.7}$$

$$I_{yy} = \sum_{i=1,N} (x_i^2 + z_i^2)\, m_i = \int (x^2 + z^2)\, dm, \tag{8.8}$$

$$I_{zz} = \sum_{i=1,N} (x_i^2 + y_i^2)\, m_i = \int (x^2 + y^2)\, dm, \tag{8.9}$$

$$I_{xy} = I_{yx} = -\sum_{i=1,N} x_i\, y_i\, m_i = -\int x\, y\, dm, \tag{8.10}$$

$$I_{yz} = I_{zy} = -\sum_{i=1,N} y_i\, z_i\, m_i = -\int y\, z\, dm, \tag{8.11}$$

$$I_{xz} = I_{zx} = -\sum_{i=1,N} x_i\, z_i\, m_i = -\int x\, z\, dm. \tag{8.12}$$

Here, I_{xx} is called the *moment of inertia* about the x-axis, I_{yy} the moment of inertia about the y-axis, I_{xy} the xy *product of inertia*, I_{yz} the yz product of inertia, *etc*. The matrix of the I_{ij} values is known as the *moment of inertia tensor*.[1] Note that each component of the moment of inertia tensor can be written as either a sum over separate mass elements, or as an integral over infinitesimal mass elements. In the integrals, $dm = \rho\, dV$, where ρ is the mass density, and dV a volume element. Equation (8.6) can be written more succinctly as

$$\mathbf{L} = \tilde{\mathbf{I}}\, \boldsymbol{\omega}. \tag{8.13}$$

Here, it is understood that $\mathbf{L}$ and $\boldsymbol{\omega}$ are both *column vectors*, and $\tilde{\mathbf{I}}$ is the *matrix* of the I_{ij} values. Note that $\tilde{\mathbf{I}}$ is a *real symmetric* matrix: *i.e.*, $I_{ij}^* = I_{ij}$ and $I_{ji} = I_{ij}$.

In general, the angular momentum vector, $\mathbf{L}$, obtained from Equation (8.13), points in a different direction to the angular velocity vector, $\boldsymbol{\omega}$. In other words, $\mathbf{L}$ is generally *not parallel* to $\boldsymbol{\omega}$.

Finally, although the above results were obtained assuming a fixed angular velocity, they remain valid at each instant in time if the angular velocity varies.

8.4 Rotational Kinetic Energy

The instantaneous rotational kinetic energy of a rotating rigid body is written

$$K = \frac{1}{2} \sum_{i=1,N} m_i \left(\frac{d\mathbf{r}_i}{dt} \right)^2. \tag{8.14}$$

Making use of Equation (8.4), and some vector identities (see Section A.9), the kinetic energy takes the form

$$K = \frac{1}{2} \sum_{i=1,N} m_i\, (\boldsymbol{\omega} \times \mathbf{r}_i) \cdot (\boldsymbol{\omega} \times \mathbf{r}_i) = \frac{1}{2}\, \boldsymbol{\omega} \cdot \sum_{i=1,N} m_i\, \mathbf{r}_i \times (\boldsymbol{\omega} \times \mathbf{r}_i). \tag{8.15}$$

[1] A tensor is the two-dimensional generalization of a vector. However, for present purposes, we can simply think of a tensor as another name for a matrix.

Hence, it follows from (8.5) that

$$K = \frac{1}{2}\,\boldsymbol{\omega}\cdot\mathbf{L}. \tag{8.16}$$

Making use of Equation (8.13), we can also write

$$K = \frac{1}{2}\,\boldsymbol{\omega}^{\mathrm{T}}\,\tilde{\mathbf{I}}\,\boldsymbol{\omega}. \tag{8.17}$$

Here, $\boldsymbol{\omega}^{\mathrm{T}}$ is the *row vector* of the Cartesian components ω_x, ω_y, ω_z, which is, of course, the transpose (denoted $^{\mathrm{T}}$) of the column vector $\boldsymbol{\omega}$. When written in component form, the above equation yields

$$K = \frac{1}{2}\left(I_{xx}\,\omega_x^2 + I_{yy}\,\omega_y^2 + I_{zz}\,\omega_z^2 + 2\,I_{xy}\,\omega_x\,\omega_y + 2\,I_{yz}\,\omega_y\,\omega_z + 2\,I_{xz}\,\omega_x\,\omega_z\right). \tag{8.18}$$

8.5 Matrix Eigenvalue Theory

It is time to review a little matrix theory. Suppose that $\mathbf{A}$ is a *real symmetric* matrix of dimension n. If follows that $\mathbf{A}^* = \mathbf{A}$ and $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$, where * denotes a complex conjugate, and $^{\mathrm{T}}$ denotes a transpose. Consider the matrix equation

$$\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{x}. \tag{8.19}$$

Any column vector $\mathbf{x}$ which satisfies the above equation is called an *eigenvector* of $\mathbf{A}$. Likewise, the associated number λ is called an *eigenvalue* of $\mathbf{A}$. Let us investigate the properties of the eigenvectors and eigenvalues of a real symmetric matrix.

Equation (8.19) can be rearranged to give

$$(\mathbf{A} - \lambda\,\mathbf{1})\,\mathbf{x} = \mathbf{0}, \tag{8.20}$$

where $\mathbf{1}$ is the unit matrix. The above matrix equation is essentially a set of n homogeneous simultaneous algebraic equations for the n components of $\mathbf{x}$. A well-known property of such a set of equations is that it only has a non-trivial solution when the determinant of the associated matrix is set to zero. Hence, a necessary condition for the above set of equations to have a non-trivial solution is that

$$|\mathbf{A} - \lambda\,\mathbf{1}| = 0. \tag{8.21}$$

The above formula is essentially an nth-order *polynomial* equation for λ. We know that such an equation has n (possibly complex) roots. Hence, we conclude that there are n eigenvalues, and n associated eigenvectors, of the n-dimensional matrix $\mathbf{A}$.

Let us now demonstrate that the n eigenvalues and eigenvectors of the real symmetric matrix $\mathbf{A}$ are all *real*. We have

$$\mathbf{A}\,\mathbf{x}_i = \lambda_i\,\mathbf{x}_i, \tag{8.22}$$

and, taking the transpose and complex conjugate,

$$\mathbf{x}_i^{*\,\mathrm{T}}\,\mathbf{A} = \lambda_i^*\,\mathbf{x}_i^{*\,\mathrm{T}}, \tag{8.23}$$

where $\mathbf{x}_i$ and λ_i are the ith eigenvector and eigenvalue of $\mathbf{A}$, respectively. Left multiplying Equation (8.22) by $\mathbf{x}_i^{*\,\mathrm{T}}$, we obtain

$$\mathbf{x}_i^{*\,\mathrm{T}}\mathbf{A}\,\mathbf{x}_i = \lambda_i\,\mathbf{x}_i^{*\,\mathrm{T}}\mathbf{x}_i. \tag{8.24}$$

Likewise, right multiplying (8.23) by $\mathbf{x}_i$, we get

$$\mathbf{x}_i^{*\,\mathrm{T}}\,\mathbf{A}\,\mathbf{x}_i = \lambda_i^*\,\mathbf{x}_i^{*\,\mathrm{T}}\mathbf{x}_i. \tag{8.25}$$

The difference of the previous two equations yields

$$(\lambda_i - \lambda_i^*)\,\mathbf{x}_i^{*\,\mathrm{T}}\mathbf{x}_i = 0. \tag{8.26}$$

It follows that $\lambda_i = \lambda_i^*$, since $\mathbf{x}_i^{*\,\mathrm{T}}\mathbf{x}_i$ (which is $\mathbf{x}_i^*\cdot\mathbf{x}_i$ in vector notation) is positive definite. Hence, λ_i is real. It immediately follows that $\mathbf{x}_i$ is real.

Next, let us show that two eigenvectors corresponding to two *different* eigenvalues are *mutually orthogonal*. Let

$$\mathbf{A}\,\mathbf{x}_i = \lambda_i\,\mathbf{x}_i, \tag{8.27}$$

$$\mathbf{A}\,\mathbf{x}_j = \lambda_j\,\mathbf{x}_j, \tag{8.28}$$

where $\lambda_i \neq \lambda_j$. Taking the transpose of the first equation and right multiplying by $\mathbf{x}_j$, and left multiplying the second equation by $\mathbf{x}_i^{\mathrm{T}}$, we obtain

$$\mathbf{x}_i^{\mathrm{T}}\,\mathbf{A}\,\mathbf{x}_j = \lambda_i\,\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j, \tag{8.29}$$

$$\mathbf{x}_i^{\mathrm{T}}\,\mathbf{A}\,\mathbf{x}_j = \lambda_j\,\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j. \tag{8.30}$$

Taking the difference of the above two equations, we get

$$(\lambda_i - \lambda_j)\,\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j = 0. \tag{8.31}$$

Since, by hypothesis, $\lambda_i \neq \lambda_j$, it follows that $\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j = 0$. In vector notation, this is the same as $\mathbf{x}_i\cdot\mathbf{x}_j = 0$. Hence, the eigenvectors $\mathbf{x}_i$ and $\mathbf{x}_j$ are mutually orthogonal.

Suppose that $\lambda_i = \lambda_j = \lambda$. In this case, we cannot conclude that $\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j = 0$ by the above argument. However, it is easily seen that any linear combination of $\mathbf{x}_i$ and $\mathbf{x}_j$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ. Hence, it is possible to define two new eigenvectors of $\mathbf{A}$, with the eigenvalue λ, which are mutually orthogonal. For instance,

$$\mathbf{x}_i' = \mathbf{x}_i, \tag{8.32}$$

$$\mathbf{x}_j' = \mathbf{x}_j - \left(\frac{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j}{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_i}\right)\mathbf{x}_i. \tag{8.33}$$

It should be clear that this argument can be generalized to deal with any number of eigenvalues which take the same value.

In conclusion, a real symmetric n-dimensional matrix possesses n *real* eigenvalues, with n associated *real* eigenvectors, which are, or can be chosen to be, *mutually orthogonal*.

8.6 Principal Axes of Rotation

We have seen that the moment of inertia tensor, $\tilde{\mathbf{I}}$, defined in Section 8.3, takes the form of a real symmetric three-dimensional matrix. It therefore follows, from the matrix theory that we have just reviewed, that the moment of inertia tensor possesses *three mutually orthogonal eigenvectors* which are associated with *three real eigenvalues*. Let the ith eigenvector (which can be normalized to be a unit vector) be denoted $\hat{\boldsymbol{\omega}}_i$, and the ith eigenvalue λ_i. It then follows that

$$\tilde{\mathbf{I}}\,\hat{\boldsymbol{\omega}}_i = \lambda_i\,\hat{\boldsymbol{\omega}}_i, \tag{8.34}$$

for $i = 1, 3$.

The directions of the three mutually orthogonal unit vectors $\hat{\boldsymbol{\omega}}_i$ define the three so-called *principal axes of rotation* of the rigid body under investigation. These axes are special because when the body rotates about one of them (*i.e.*, when $\boldsymbol{\omega}$ is parallel to one of them) the angular momentum vector $\mathbf{L}$ becomes *parallel* to the angular velocity vector $\boldsymbol{\omega}$. This can be seen from a comparison of Equation (8.13) and Equation (8.34).

Suppose that we reorient our Cartesian coordinate axes so the they coincide with the mutually orthogonal principal axes of rotation. In this new reference frame, the eigenvectors of $\tilde{\mathbf{I}}$ are the unit vectors, $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$, and the eigenvalues are the moments of inertia about these axes, I_{xx}, I_{yy}, and I_{zz}, respectively. These latter quantities are referred to as the *principal moments of inertia*. Note that the products of inertia are all *zero* in the new reference frame. Hence, in this frame, the moment of inertia tensor takes the form of a *diagonal* matrix: *i.e.*,

$$\tilde{\mathbf{I}} = \begin{pmatrix} I_{xx} & 0 & 0 \\ 0 & I_{yy} & 0 \\ 0 & 0 & I_{zz} \end{pmatrix}. \tag{8.35}$$

Incidentally, it is easy to verify that $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ are indeed the eigenvectors of the above matrix, with the eigenvalues I_{xx}, I_{yy}, and I_{zz}, respectively, and that $\mathbf{L} = \tilde{\mathbf{I}}\,\boldsymbol{\omega}$ is indeed parallel to $\boldsymbol{\omega}$ whenever $\boldsymbol{\omega}$ is directed along $\mathbf{e}_x$, $\mathbf{e}_y$, or $\mathbf{e}_z$.

When expressed in our new coordinate system, Equation (8.13) yields

$$\mathbf{L} = (I_{xx}\,\omega_x,\, I_{yy}\,\omega_y, I_{zz}\,\omega_z)\,, \tag{8.36}$$

whereas Equation (8.18) reduces to

$$K = \frac{1}{2}\left(I_{xx}\,\omega_x^2 + I_{yy}\,\omega_y^2 + I_{zz}\,\omega_z^2\right). \tag{8.37}$$

In conclusion, there are many great simplifications to be had by choosing a coordinate system whose axes coincide with the principal axes of rotation of the rigid body under investigation. But how do we determine the directions of the principal axes in practice?

Well, in general, we have to solve the eigenvalue equation

$$\tilde{\mathbf{I}}\,\hat{\boldsymbol{\omega}} = \lambda\,\hat{\boldsymbol{\omega}}, \tag{8.38}$$

or

$$\begin{pmatrix} I_{xx}-\lambda & I_{xy} & I_{xz} \\ I_{yx} & I_{yy}-\lambda & I_{yz} \\ I_{zx} & I_{zy} & I_{zz}-\lambda \end{pmatrix} \begin{pmatrix} \cos\alpha \\ \cos\beta \\ \cos\gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{8.39}$$

where $\hat{\omega} = (\cos\alpha, \cos\beta, \cos\gamma)$, and $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$. Here, α is the angle the unit eigenvector subtends with the x-axis, β the angle it subtends with the y-axis, and γ the angle it subtends with the z-axis. Unfortunately, the analytic solution of the above matrix equation is generally quite difficult.

Fortunately, however, in many instances the rigid body under investigation possesses some kind of symmetry, so that at least one principal axis can be found by inspection. In this case, the other two principal axes can be determined as follows.

Suppose that the z-axis is known to be a principal axes (at the origin) in some coordinate system. It follows that the two products of inertia I_{xz} and I_{yz} are zero [otherwise, $(0, 0, 1)$ would not be an eigenvector in Equation (8.39)]. The other two principal axes must lie in the x-y plane: *i.e.*, $\cos\gamma = 0$. It then follows that $\cos\beta = \sin\alpha$, since $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$. The first two rows in the matrix equation (8.39) thus reduce to

$$(I_{xx}-\lambda)\cos\alpha + I_{xy}\sin\alpha = 0, \tag{8.40}$$

$$I_{xy}\cos\alpha + (I_{yy}-\lambda)\sin\alpha = 0. \tag{8.41}$$

Eliminating λ between the above two equations, we obtain

$$I_{xy}(1-\tan^2\alpha) = (I_{xx}-I_{yy})\tan\alpha. \tag{8.42}$$

But, $\tan(2\alpha) \equiv 2\tan\alpha/(1-\tan^2\alpha)$. Hence, Equation (8.42) yields

$$\tan(2\alpha) = \frac{2\,I_{xy}}{I_{xx}-I_{yy}}. \tag{8.43}$$

There are two values of α, lying between $-\pi/2$ and $\pi/2$, which satisfy the above equation. These specify the angles, α, that the two mutually orthogonal principal axes lying in the x-y plane make with the x-axis. Hence, we have now determined the directions of all three principal axes. Incidentally, once we have determined the orientation angle, α, of a principal axis, we can then substitute back into Equation (8.40) to obtain the corresponding principal moment of inertia, λ.

As an example, consider a uniform rectangular lamina of mass m and sides a and b which lies in the x-y plane, as shown in Figure 8.2. Suppose that the axis of rotation passes through the origin (*i.e.*, through a corner of the lamina). Since $z = 0$ throughout the lamina, it follows from Equations (8.11) and (8.12) that $I_{xz} = I_{yz} = 0$. Hence, the z-axis is a principal axis. After some straightforward integration, Equations (8.7)–(8.10) yield

$$I_{xx} = \frac{1}{3}\,m\,b^2, \tag{8.44}$$

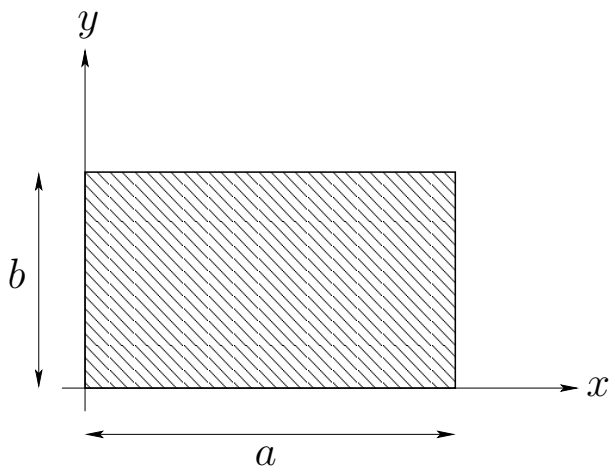


Figure 8.2: *A uniform rectangular laminar.*

$$I_{yy} = \frac{1}{3} m a^2, \tag{8.45}$$

$$I_{xy} = -\frac{1}{4} m a b. \tag{8.46}$$

Thus, it follows from Equation (8.43) that

$$\alpha = \frac{1}{2} \tan^{-1}\left(\frac{3}{2}\frac{a b}{a^2 - b^2}\right). \tag{8.47}$$

The above equation specifies the orientation of the two principal axes that lie in the x-y plane. For the special case where $a = b$, we get $\alpha = \pi/4, 3\pi/4$: *i.e.*, the two in-plane principal axes of a square lamina (at a corner) are parallel to the two diagonals of the lamina.

8.7 Euler's Equations

The fundamental equation of motion of a rotating body [see Equation (8.3)],

$$\mathbf{T} = \frac{d\mathbf{L}}{dt}, \tag{8.48}$$

is only valid in an *inertial* frame. However, we have seen that $\mathbf{L}$ is most simply expressed in a frame of reference whose axes are aligned along the principal axes of rotation of the body. Such a frame of reference *rotates* with the body, and is, therefore, *non-inertial*. Thus, it is helpful to define *two* Cartesian coordinate systems, with the same origins. The first, with coordinates x, y, z, is a fixed inertial frame—let us denote this the *fixed frame*. The second, with coordinates x', y', z', co-rotates with the body in such a manner that the x'-,

y'-, and z'-axes are always pointing along its principal axes of rotation—we shall refer to this as the *body frame*. Since the body frame co-rotates with the body, its instantaneous angular velocity is the same as that of the body. Hence, it follows from the analysis in Section 7.2 that

$$\frac{d\mathbf{L}}{dt} = \frac{d\mathbf{L}}{dt'} + \boldsymbol{\omega} \times \mathbf{L}. \tag{8.49}$$

Here, d/dt is the time derivative in the fixed frame, and d/dt' the time derivative in the body frame. Combining Equations (8.48) and (8.49), we obtain

$$\mathbf{T} = \frac{d\mathbf{L}}{dt'} + \boldsymbol{\omega} \times \mathbf{L}. \tag{8.50}$$

Now, in the body frame let $\mathbf{T} = (T_{x'}, T_{y'}, T_{z'})$ and $\boldsymbol{\omega} = (\omega_{x'}, \omega_{y'}, \omega_{z'})$. It follows that $\mathbf{L} = (I_{x'x'}\,\omega_{x'}, I_{y'y'}\,\omega_{y'}, I_{z'z'}\,\omega_{z'})$, where $I_{x'x'}$, $I_{y'y'}$ and $I_{z'z'}$ are the principal moments of inertia. Hence, in the body frame, the components of Equation (8.50) yield

$$T_{x'} = I_{x'x'}\,\dot{\omega}_{x'} - (I_{y'y'} - I_{z'z'})\,\omega_{y'}\,\omega_{z'}, \tag{8.51}$$

$$T_{y'} = I_{y'y'}\,\dot{\omega}_{y'} - (I_{z'z'} - I_{x'x'})\,\omega_{z'}\,\omega_{x'}, \tag{8.52}$$

$$T_{z'} = I_{z'z'}\,\dot{\omega}_{z'} - (I_{x'x'} - I_{y'y'})\,\omega_{x'}\,\omega_{y'}, \tag{8.53}$$

where $\dot{} = d/dt'$. Here, we have made use of the fact that the moments of inertia of a rigid body are *constant* in time in the co-rotating body frame. The above equations are known as *Euler's equations*.

Consider a rigid body which is constrained to rotate about a fixed axis with *constant* angular velocity. It follows that $\dot{\omega}_{x'} = \dot{\omega}_{y'} = \dot{\omega}_{z'} = 0$. Hence, Euler's equations, (8.51)–(8.53), reduce to

$$T_{x'} = -(I_{y'y'} - I_{z'z'})\,\omega_{y'}\,\omega_{z'}, \tag{8.54}$$

$$T_{y'} = -(I_{z'z'} - I_{x'x'})\,\omega_{z'}\,\omega_{x'}, \tag{8.55}$$

$$T_{z'} = -(I_{x'x'} - I_{y'y'})\,\omega_{x'}\,\omega_{y'}. \tag{8.56}$$

These equations specify the components of the steady (in the body frame) torque exerted on the body by the constraining supports. The steady (in the body frame) angular momentum is written

$$\mathbf{L} = (I_{x'x'}\,\omega_{x'},\, I_{y'y'}\,\omega_{y'},\, I_{z'z'}\,\omega_{z'}). \tag{8.57}$$

It is easily demonstrated that $\mathbf{T} = \boldsymbol{\omega} \times \mathbf{L}$. Hence, the torque is perpendicular to both the angular velocity and the angular momentum vectors. Note that if the axis of rotation is a principal axis then two of the three components of $\boldsymbol{\omega}$ are zero (in the body frame). It follows from Equations (8.54)–(8.56) that all three components of the torque are zero. In other words, *zero* external torque is required to make the body rotate steadily about a *principal axis*.

Suppose that the body is *freely rotating*: *i.e.*, there are no external torques. Furthermore, let the body be *rotationally symmetric* about the z'-axis. It follows that $I_{x'x'} = I_{y'y'} =$

$I_\perp$. Likewise, we can write $I_{z'z'} = I_\parallel$. In general, however, $I_\perp \neq I_\parallel$. Thus, Euler's equations yield

$$I_\perp \frac{d\omega_{x'}}{dt'} + (I_\parallel - I_\perp)\,\omega_{z'}\,\omega_{y'} = 0, \tag{8.58}$$

$$I_\perp \frac{d\omega_{y'}}{dt'} - (I_\parallel - I_\perp)\,\omega_{z'}\,\omega_{x'} = 0, \tag{8.59}$$

$$\frac{d\omega_{z'}}{dt'} = 0. \tag{8.60}$$

Clearly, $\omega_{z'}$ is a constant of the motion. Equation (8.58) and (8.59) can be written

$$\frac{d\omega_{x'}}{dt'} + \Omega\,\omega_{y'} = 0, \tag{8.61}$$

$$\frac{d\omega_{y'}}{dt'} - \Omega\,\omega_{x'} = 0, \tag{8.62}$$

where $\Omega = (I_\parallel/I_\perp - 1)\,\omega_{z'}$. As is easily demonstrated, the solution to the above equations is

$$\omega_{x'} = \omega_\perp \cos(\Omega\,t'), \tag{8.63}$$

$$\omega_{y'} = \omega_\perp \sin(\Omega\,t'), \tag{8.64}$$

where $\omega_\perp$ is a constant. Thus, the projection of the angular velocity vector onto the x'-y' plane has the fixed length $\omega_\perp$, and rotates steadily about the z'-axis with angular velocity Ω. It follows that the length of the angular velocity vector, $\omega = (\omega_{x'}^2 + \omega_{y'}^2 + \omega_{z'}^2)^{1/2}$, is a constant of the motion. Clearly, the angular velocity vector makes some constant angle, α, with the z'-axis, which implies that $\omega_{z'} = \omega\cos\alpha$ and $\omega_\perp = \omega\sin\alpha$. Hence, the components of the angular velocity vector are

$$\omega_{x'} = \omega\,\sin\alpha\,\cos(\Omega\,t'), \tag{8.65}$$

$$\omega_{y'} = \omega\,\sin\alpha\,\sin(\Omega\,t'), \tag{8.66}$$

$$\omega_{z'} = \omega\,\cos\alpha, \tag{8.67}$$

where

$$\Omega = \omega\,\cos\alpha\left(\frac{I_\parallel}{I_\perp} - 1\right). \tag{8.68}$$

We conclude that, in the body frame, the angular velocity vector *precesses* about the symmetry axis (*i.e.*, the z'-axis) with the angular frequency Ω. Now, the components of the angular momentum vector are

$$L_{x'} = I_\perp\,\omega\,\sin\alpha\,\cos(\Omega\,t'), \tag{8.69}$$

$$L_{y'} = I_\perp\,\omega\,\sin\alpha\,\sin(\Omega\,t'), \tag{8.70}$$

$$L_{z'} = I_\parallel\,\omega\,\cos\alpha. \tag{8.71}$$

Thus, in the body frame, the angular momentum vector is also of constant length, and precesses about the symmetry axis with the angular frequency Ω. Furthermore, the angular momentum vector makes a constant angle θ with the symmetry axis, where

$$\tan\theta = \frac{I_\perp}{I_\parallel}\,\tan\alpha. \tag{8.72}$$

Note that the angular momentum vector, the angular velocity vector, and the symmetry axis all lie in the *same plane*: *i.e.*, $\mathbf{e}_{z'}\cdot\mathbf{L}\times\boldsymbol{\omega}=0$, as can easily be verified. Moreover, the angular momentum vector lies between the angular velocity vector and the symmetry axis (*i.e.*, $\theta<\alpha$) for a flattened (or oblate) body (*i.e.*, $I_\perp<I_\parallel$), whereas the angular velocity vector lies between the angular momentum vector and the symmetry axis (*i.e.*, $\theta>\alpha$) for an elongated (or prolate) body (*i.e.*, $I_\perp>I_\parallel$).

Let us now consider the most general motion of a freely rotating *asymmetric* rigid body, as seen in the body frame. Since a freely rotating body experiences no external torques, its angular momentum vector $\mathbf{L}$ is a constant of the motion in the inertial fixed frame. In general, the direction of this vector varies with time in the non-inertial body frame, but its length remains fixed. This can be seen from Equation (8.49): if $d\mathbf{L}/dt=0$ then the scalar product of this equation with $\mathbf{L}$ implies that $dL^2/dt'=0$. It follows from Equation (8.57) that

$$L^2 = I_{x'x'}^2\,\omega_{x'}^2 + I_{y'y'}^2\,\omega_{y'}^2 + I_{z'z'}^2\,\omega_{z'}^2 = \text{constant}. \tag{8.73}$$

The above constraint can also be derived directly from Euler's equations, (8.51)–(8.53), by setting $T_{x'}=T_{y'}=T_{z'}=0$. A freely rotating body subject to no external torques clearly has a constant rotational kinetic energy. Hence, from Equation (8.16),

$$\boldsymbol{\omega}\cdot\mathbf{L} = I_{x'x'}\,\omega_{x'}^2 + I_{y'y'}\,\omega_{y'}^2 + I_{z'z'}\,\omega_{z'}^2 = \text{constant}. \tag{8.74}$$

This constraint can also be derived directly from Euler's equations. We conclude that, in the body frame, the components of $\boldsymbol{\omega}$ must simultaneously satisfy the two constraints (8.73) and (8.74). These constraints are the equations of two ellipsoids whose principal axes coincide with the principal axes of the body, and whose principal radii are in the ratio $I_{x'x'}^{-1} : I_{y'y'}^{-1} : I_{z'z'}^{-1}$ and $I_{x'x'}^{-1/2} : I_{y'y'}^{-1/2} : I_{z'z'}^{-1/2}$, respectively. In general, the intersection of these two ellipsoids is a *closed curve*. Hence, we conclude that the most general motion of a freely rotating asymmetric body, as seen in the body frame, is a form of irregular *precession* in which the tip of the angular velocity vector $\boldsymbol{\omega}$ periodically traces out the aforementioned closed curve. It is easily demonstrated that the tip of the angular momentum vector $\mathbf{L}$ periodically traces out a different closed curve.

8.8 Eulerian Angles

We have seen how we can solve Euler's equations to determine the properties of a rotating body in the co-rotating *body frame*. Let us now investigate how we can determine the same properties in the inertial *fixed frame*.

The fixed frame and the body frame share the same origin. Hence, we can transform from one to the other by means of an appropriate *rotation* of our coordinate axes. In general, if we restrict ourselves to rotations about one of the Cartesian axes, *three* successive rotations are required to transform the fixed frame into the body frame. There are, in fact, many different ways to combined three successive rotations in order to achieve this. In the following, we shall describe the most widely used method, which is due to Euler.

We start in the fixed frame, which has coordinates x, y, z, and unit vectors $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$. Our first rotation is counterclockwise (looking down the axis) through an angle ϕ about the z-axis. The new frame has coordinates x'', y'', z'', and unit vectors $\mathbf{e}_{x''}$, $\mathbf{e}_{y''}$, $\mathbf{e}_{z''}$. According to Equations (A.17)–(A.19), the transformation of coordinates can be represented as follows:

$$\begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{8.75}$$

The angular velocity vector associated with ϕ has the magnitude $\dot{\phi}$, and is directed along $\mathbf{e}_z$ (*i.e.*, along the axis of rotation). Hence, we can write

$$\boldsymbol{\omega}_\phi = \dot{\phi}\,\mathbf{e}_z. \tag{8.76}$$

Clearly, $\dot{\phi}$ is the precession rate about the $\mathbf{e}_z$ axis, as seen in the fixed frame.

The second rotation is counterclockwise (looking down the axis) through an angle θ about the x''-axis. The new frame has coordinates x''', y''', z''', and unit vectors $\mathbf{e}_{x'''}$, $\mathbf{e}_{y'''}$, $\mathbf{e}_{z'''}$. By analogy with Equation (8.75), the transformation of coordinates can be represented as follows:

$$\begin{pmatrix} x''' \\ y''' \\ z''' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix}. \tag{8.77}$$

The angular velocity vector associated with θ has the magnitude $\dot{\theta}$, and is directed along $\mathbf{e}_{x''}$ (*i.e.*, along the axis of rotation). Hence, we can write

$$\boldsymbol{\omega}_\theta = \dot{\theta}\,\mathbf{e}_{x''}. \tag{8.78}$$

The third rotation is counterclockwise (looking down the axis) through an angle ψ about the z'''-axis. The new frame is the body frame, which has coordinates x', y', z', and unit vectors $\mathbf{e}_{x'}$, $\mathbf{e}_{y'}$, $\mathbf{e}_{z'}$. The transformation of coordinates can be represented as follows:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x''' \\ y''' \\ z''' \end{pmatrix}. \tag{8.79}$$

The angular velocity vector associated with ψ has the magnitude $\dot{\psi}$, and is directed along $\mathbf{e}_{z''}$ (*i.e.*, along the axis of rotation). Note that $\mathbf{e}_{z'''} = \mathbf{e}_{z'}$, since the third rotation is about $\mathbf{e}_{z'''}$. Hence, we can write

$$\boldsymbol{\omega}_\psi = \dot{\psi}\,\mathbf{e}_{z'}. \tag{8.80}$$

Clearly, $\dot{\psi}$ is *minus* the precession rate about the $\mathbf{e}_{z'}$ axis, as seen in the body frame.

The full transformation between the fixed frame and the body frame is rather complicated. However, the following results can easily be verified:

$$\mathbf{e}_z = \sin\psi\,\sin\theta\,\mathbf{e}_{x'} + \cos\psi\,\sin\theta\,\mathbf{e}_{y'} + \cos\theta\,\mathbf{e}_{z'}, \tag{8.81}$$

$$\mathbf{e}_{x''} = \cos\psi\,\mathbf{e}_{x'} - \sin\psi\,\mathbf{e}_{y'}. \tag{8.82}$$

It follows from Equation (8.81) that $\mathbf{e}_z\cdot\mathbf{e}_{z'} = \cos\theta$. In other words, θ is the angle of inclination between the z- and z'-axes. Finally, since the total angular velocity can be written

$$\boldsymbol{\omega} = \boldsymbol{\omega}_\phi + \boldsymbol{\omega}_\theta + \boldsymbol{\omega}_\psi, \tag{8.83}$$

Equations (8.76), (8.78), and (8.80)–(8.82) yield

$$\omega_{x'} = \sin\psi\,\sin\theta\,\dot{\phi} + \cos\psi\,\dot{\theta}, \tag{8.84}$$

$$\omega_{y'} = \cos\psi\,\sin\theta\,\dot{\phi} - \sin\psi\,\dot{\theta}, \tag{8.85}$$

$$\omega_{z'} = \cos\theta\,\dot{\phi} + \dot{\psi}. \tag{8.86}$$

The angles ϕ, θ, and ψ are termed *Eulerian angles*. Each has a clear physical interpretation: ϕ is the angle of precession about the $\mathbf{e}_z$ axis in the fixed frame, ψ is minus the angle of precession about the $\mathbf{e}_{z'}$ axis in the body frame, and θ is the angle of inclination between the $\mathbf{e}_z$ and $\mathbf{e}_{z'}$ axes. Moreover, we can express the components of the angular velocity vector $\boldsymbol{\omega}$ in the body frame entirely in terms of the Eulerian angles, and their time derivatives [see Equations (8.84)–(8.86)].

Consider a rigid body which is constrained to rotate about a fixed axis with the constant angular velocity ω. Let the fixed angular velocity vector point along the z-axis. In the previous section, we saw that the angular momentum and the torque were both steady in the body frame. Since there is no precession of quantities in the body frame, it follows that the Eulerian angle ψ is constant. Furthermore, since the angular velocity vector is fixed in the body frame, as well as the fixed frame [as can be seen by applying Equation (8.49) to $\boldsymbol{\omega}$ instead of $\mathbf{L}$], it must subtend a constant angle with the $\mathbf{e}_{z'}$ axis. Hence, the Eulerian angle θ is also constant. It follows from Equations (8.84)–(8.86) that

$$\omega_{x'} = \sin\psi\,\sin\theta\,\dot{\phi}, \tag{8.87}$$

$$\omega_{y'} = \cos\psi\,\sin\theta\,\dot{\phi}, \tag{8.88}$$

$$\omega_{z'} = \cos\theta\,\dot{\phi}, \tag{8.89}$$

which implies that $\omega \equiv (\omega_{x'}^2+\omega_{y'}^2+\omega_{z'}^2)^{1/2} = \dot{\phi}$. In other words, the precession rate, $\dot{\phi}$, in the fixed frame is equal to ω. Hence, in the fixed frame, the constant torque and angular momentum vectors found in the body frame *precess* about the angular velocity vector (*i.e.*, about the z-axis) at the rate ω. As discussed in the previous section, for the special case where the angular velocity vector is *parallel* to one of the principal axes of the body, the

angular momentum vector is parallel to the angular velocity vector, and the torque is *zero*. Thus, in this case, there is no precession in the fixed frame.

Consider a rotating device such as a flywheel or a propeller. If the device is *statically balanced* then its center of mass lies on the axis of rotation. This is desirable since, otherwise, gravity, which effectively acts at the center of mass, exerts a varying torque about the axis of rotation as the device rotates, giving rise to unsteady rotation. If the device is *dynamically balanced* then the axis of rotation is also a principal axis, so that, as the device rotates its angular momentum vector, **L**, remains parallel to the axis of rotation. This is desirable since, otherwise, the angular momentum vector is not parallel to the axis of rotation, and, therefore, precesses around it. Since $d\mathbf{L}/dt$ is equal to the torque, a precessing torque must also be applied to the device (at right-angles to both the axis and **L**). The result is a reaction on the bearings which can give rise to violent vibration and wobbling, even when the device is statically balanced.

Consider a freely rotating body which is rotationally symmetric about one axis (the z'-axis). In the absence of an external torque, the angular momentum vector **L** is a constant of the motion [see Equation (8.3)]. Let **L** point along the z-axis. In the previous section, we saw that the angular momentum vector subtends a constant angle θ with the axis of symmetry: *i.e.*, with the z'-axis. Hence, the time derivative of the Eulerian angle θ is zero. We also saw that the angular momentum vector, the axis of symmetry, and the angular velocity vector are coplanar. Consider an instant in time at which all of these vectors lie in the y'-z' plane. This implies that $\omega_{x'} = 0$. According to the previous section, the angular velocity vector subtends a constant angle α with the symmetry axis. It follows that $\omega_{y'} = \omega \sin\alpha$ and $\omega_{z'} = \omega \cos\alpha$. Equation (8.84) yields $\psi = 0$. Hence, Equation (8.85) yields

$$\omega \sin\alpha = \sin\theta\, \dot{\phi}. \tag{8.90}$$

This can be combined with Equation (8.72) to give

$$\dot{\phi} = \omega \left[1 + \left(\frac{I_\parallel^2}{I_\perp^2} - 1\right) \cos^2\alpha\right]^{1/2}. \tag{8.91}$$

Finally, Equations (8.86), together with (8.72) and (8.90), yields

$$\dot{\psi} = \omega \cos\alpha - \cos\theta\, \dot{\phi} = \omega \cos\alpha \left(1 - \frac{\tan\alpha}{\tan\theta}\right) = \omega \cos\alpha \left(1 - \frac{I_\parallel}{I_\perp}\right). \tag{8.92}$$

A comparison of the above equation with Equation (8.68) gives

$$\dot{\psi} = -\Omega. \tag{8.93}$$

Thus, as expected, $\dot{\psi}$ is minus the precession rate (of the angular momentum and angular velocity vectors) in the body frame. On the other hand, $\dot{\phi}$ is the precession rate (of the angular velocity vector and the symmetry axis) in the fixed frame. Note that $\dot{\phi}$ and Ω are quite dissimilar. For instance, Ω is negative for elongated bodies ($I_\parallel < I_\perp$) whereas $\dot{\phi}$ is

positive definite. It follows that the precession is always in the same sense as L_z in the fixed frame, whereas the precession in the body frame is in the opposite sense to $L_{z'}$ for elongated bodies. We found, in the previous section, that for a *flattened* body the angular momentum vector lies between the angular velocity vector and the symmetry axis. This means that, in the fixed frame, the angular velocity vector and the symmetry axis lie on *opposite* sides of the fixed angular momentum vector. On the other hand, for an *elongated* body we found that the angular velocity vector lies between the angular momentum vector and the symmetry axis. This means that, in the fixed frame, the angular velocity vector and the symmetry axis lie on the *same* side of the fixed angular momentum vector. (Recall that the angular momentum vector, the angular velocity vector, and the symmetry axis, are coplanar.)

As an example, consider the free rotation of a thin disk. It is easily demonstrated (from the perpendicular axis theorem) that

$$I_{\parallel} = 2\, I_{\perp} \tag{8.94}$$

for such a disk. Hence, from Equation (8.68), the precession rate in the body frame is

$$\Omega = \omega\, \cos\alpha. \tag{8.95}$$

According to Equation (8.91), the precession rate in the fixed frame is

$$\dot{\phi} = \omega \left[1 + 3\, \cos^2\alpha\right]^{1/2}. \tag{8.96}$$

In the limit in which α is small (*i.e.*, in which the angular velocity vector is almost parallel to the symmetry axis), we obtain

$$\Omega \simeq \omega, \tag{8.97}$$

$$\dot{\phi} \simeq 2\,\omega. \tag{8.98}$$

Thus, the symmetry axis precesses in the fixed frame at approximately twice the angular speed of rotation. This precession is manifest as a wobbling motion.

It is known that the axis of rotation of the Earth is very slightly inclined to its symmetry axis (which passes through the two geographic poles). The angle α is approximately 0.2 seconds of an arc (which corresponds to a distance of about 6 m on the Earth's surface). It is also known that the ratio of the terrestrial moments of inertia is about $I_{\parallel}/I_{\perp} = 1.00327$, as determined from the Earth's oblateness—see Section 12.7. Hence, from (8.68), the precession rate of the angular velocity vector about the symmetry axis, as viewed in a geostationary reference frame, is

$$\Omega = 0.00327\,\omega, \tag{8.99}$$

giving a precession period of

$$T' = \frac{2\pi}{\Omega} = 305 \text{ days}. \tag{8.100}$$

(Of course, $2\pi/\omega = 1$ day.) The observed period of precession is about 440 days. The disagreement between theory and observation is attributed to the fact that the Earth is not perfectly rigid. The Earth's symmetry axis subtends an angle $\theta \simeq \alpha = 0.2''$ [see (8.72)] with its angular momentum vector, but lies on the opposite side of this vector to the angular velocity vector. This implies that, as viewed from space, the Earth's angular velocity vector is almost parallel to its fixed angular momentum vector, whereas its symmetry axis subtends an angle of $0.2''$ with both vectors, and precesses about them. The (theoretical) precession rate of the Earth's symmetry axis, as seen from space, is given by Equation (8.91):

$$\dot{\phi} = 1.00327\,\omega. \tag{8.101}$$

The associated precession period is

$$T = \frac{2\pi}{\dot{\phi}} = 0.997 \text{ days}. \tag{8.102}$$

The free precession of the Earth's symmetry axis in space, which is known as the *Chandler wobble*, since it was discovered by the American astronomer Seth Chandler in 1891, is superimposed on a much slower forced precession, with a period of about 26,000 years, caused by the small gravitational torque exerted on the Earth by the Sun and the Moon, as a consequence of the Earth's slight oblateness—see Section 12.10.

8.9 Gyroscopic Precession

Let us now study the motion of a *rigid rotationally symmetric top* which is free to turn about a fixed point (without friction), but which is subject to a gravitational torque—see Figure 8.3. Suppose that the z'-axis coincides with the symmetry axis. Let the principal moment of inertia about the symmetry axis be $I_\parallel$, and let the other principal moments both take the value $I_\perp$. Suppose that the z-axis points vertically upward, and let the common origin, O, of the fixed and body frames coincide with the fixed point about which the top turns. Suppose that the center of mass of the top lies a distance l along its symmetry axis from point O, and that the mass of the top is m. Let the symmetry axis of the top subtend an angle θ (which is an Eulerian angle) with the upward vertical.

Consider an instant in time at which the Eulerian angle ψ is zero. This implies that the x'-axis is horizontal [see Equation (8.81)], as shown in the diagram. The gravitational force, which acts at the center of mass, thus exerts a torque $m\,g\,l\sin\theta$ in the x'-direction. Hence, the components of the torque in the body frame are

$$T_{x'} = m\,g\,l\sin\theta, \tag{8.103}$$

$$T_{y'} = 0, \tag{8.104}$$

$$T_{z'} = 0. \tag{8.105}$$

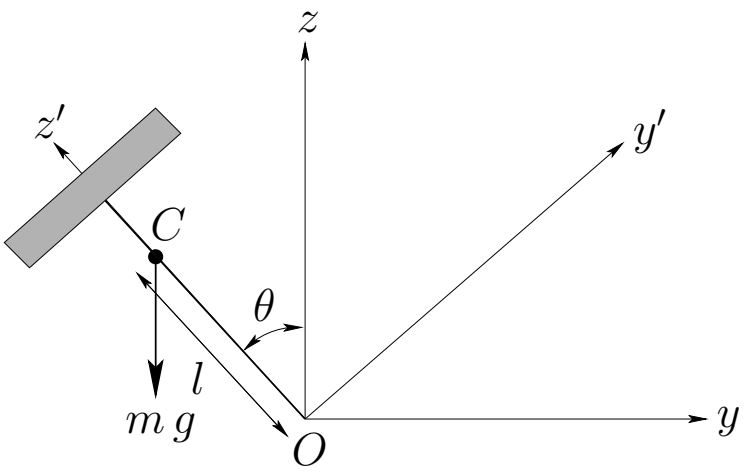


Figure 8.3: *A symmetric top.*

The components of the angular velocity vector in the body frame are given by Equations (8.84)–(8.86). Thus, Euler's equations (8.51)–(8.53) take the form:

$$m\,g\,l\,\sin\theta = I_\perp\,(\ddot\theta - \cos\theta\,\sin\theta\,\dot\phi^{\,2}) + L_\psi\,\sin\theta\,\dot\phi, \tag{8.106}$$

$$0 = I_\perp\,(2\,\cos\theta\,\dot\theta\,\dot\phi + \sin\theta\,\ddot\phi) - L_\psi\,\dot\theta, \tag{8.107}$$

$$0 = \dot L_\psi, \tag{8.108}$$

where

$$L_\psi = I_\parallel\,(\cos\theta\,\dot\phi + \dot\psi) = I_\parallel\,\Omega, \tag{8.109}$$

and $\Omega = \omega_{z'}$ is the angular velocity of the top. Multiplying Equation (8.107) by $\sin\theta$, we obtain

$$\dot L_\phi = 0, \tag{8.110}$$

where

$$L_\phi = I_\perp\,\sin^2\theta\,\dot\phi + L_\psi\,\cos\theta. \tag{8.111}$$

According to Equations (8.108) and (8.110), the two quantities L_ψ and L_ϕ are constants of the motion. These two quantities are the *angular momenta* of the system about the z'- and z-axis, respectively. (To be more exact, they are the generalized momenta conjugate to the coordinates ψ and ϕ, respectively—see Section 9.8.) They are conserved because the gravitational torque has no component along either the z'- or the z-axis. (Alternatively, they are conserved because the Lagrangian of the top does not depend explicitly on the coordinates ψ and θ—see Section 9.8.)

If there are no frictional forces acting on the top then the total energy, $E = K + U$, is also a constant of the motion. Now,

$$E = \frac{1}{2}\,\left(I_\perp\,\omega_{x'}^2 + I_\perp\,\omega_{y'}^2 + I_\parallel\,\omega_{z'}^2\right) + m\,g\,l\,\cos\theta. \tag{8.112}$$

When written in terms of the Eulerian angles (with $\psi = 0$), this becomes

$$E = \frac{1}{2}\left(I_\perp\,\dot\theta^2 + I_\perp\,\sin^2\theta\,\dot\phi^2 + L_\psi^2/I_\parallel\right) + m\,g\,l\,\cos\theta. \tag{8.113}$$

Eliminating $\dot\phi$ between Equations (8.111) and (8.113), we obtain the following differential equation for θ:

$$E = \frac{1}{2}\,I_\perp\,\dot\theta^2 + \frac{(L_\phi - L_\psi\,\cos\theta)^2}{2\,I_\perp\,\sin^2\theta} + \frac{1}{2}\frac{L_\psi^2}{I_\parallel} + m\,g\,l\,\cos\theta. \tag{8.114}$$

Let

$$E' = E - \frac{1}{2}\frac{L_\psi^2}{I_\parallel}, \tag{8.115}$$

and $u = \cos\theta$. It follows that

$$\dot u^2 = 2\,(E' - m\,g\,l\,u)\,(1-u^2)\,I_\perp^{-1} - (L_\phi - L_\psi\,u)^2\,I_\perp^{-2}, \tag{8.116}$$

or

$$\dot u^2 = f(u), \tag{8.117}$$

where $f(u)$ is a *cubic* polynomial. In principal, the above equation can be integrated to give u (and, hence, θ) as a function of t:

$$t = \int_{u_0}^{u} \frac{du'}{\sqrt{f(u')}}. \tag{8.118}$$

Fortunately, we do not have to perform the above integration (which is very ugly) in order to discuss the general properties of the solution to Equation (8.117). It is clear, from Equation (8.118), that $f(u)$ needs to be *positive* in order to obtain a physical solution. Hence, the limits of the motion in θ are determined by the three roots of the equation $f(u) = 0$. Since θ must lie between 0 and $\pi/2$, it follows that u must lie between 0 and 1. It can easily be demonstrated that $f \rightarrow \pm\infty$ as $u \rightarrow \pm\infty$. It can also be shown that the largest root u_3 lies in the region $u_3 > 1$, and the two smaller roots u_1 and u_2 (if they exist) lie in the region $-1 \leq u \leq +1$. It follows that, in the region $-1 \leq u \leq 1$, $f(u)$ is only positive between u_1 and u_2. Figure 8.4 shows a case where u_1 and u_2 lie in the range 0 to 1. The corresponding values of θ—θ_1 and θ_2, say—are then the limits of the vertical motion. The axis of the top oscillates backward and forward between these two values of θ as the top precesses about the vertical axis. This oscillation is called *nutation*. Incidentally, if u_1 becomes negative then the nutation will cause the top to strike the ground (assuming that it is spinning on a level surface).

If there is a double root of $f(u) = 0$ (*i.e.*, if $u_1 = u_2$) then there is no nutation, and the top precesses steadily. However, the criterion for steady precession is most easily obtained directly from Equation (8.106). In the absence of nutation, $\dot\theta = \ddot\theta = 0$. Hence, we obtain

$$m\,g\,l = -I_\perp\,\cos\theta\,\dot\phi^2 + L_\psi\,\dot\phi, \tag{8.119}$$

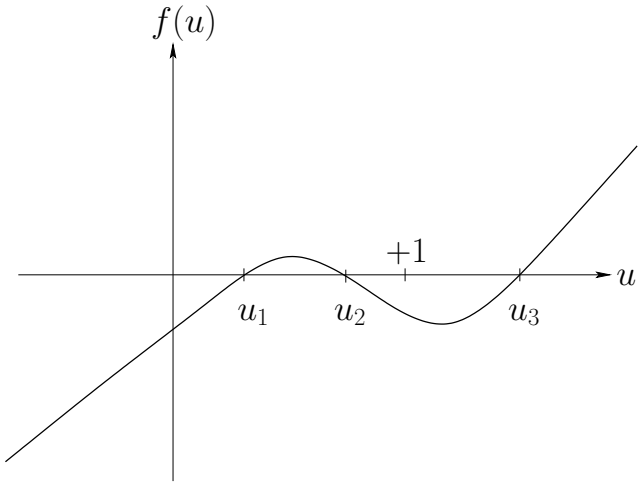


Figure 8.4: *The function* $f(u)$.

or

$$\Omega = \frac{m\,g\,l}{I_\parallel\,\dot{\phi}} + \frac{I_\perp}{I_\parallel}\,\cos\theta\,\dot{\phi}. \tag{8.120}$$

The above equation is the criterion for steady precession. Since the right-hand side of Equation (8.120) possesses a minimum value, which is given by $2\sqrt{m\,g\,l\,I_\perp\,\cos\theta}/I_\parallel$, it follows that

$$\Omega > \Omega_{\rm crit} = \frac{2\sqrt{m\,g\,l\,I_\perp\,\cos\theta}}{I_\parallel} \tag{8.121}$$

is a necessary condition for obtaining steady precession at the inclination angle θ. For $\Omega > \Omega_{\rm crit}$, there are two roots to Equation (8.120), corresponding to a slow and a fast steady precession rate for a given inclination angle θ. If $\Omega \gg \Omega_{\rm crit}$ then these two roots are approximately given by

$$(\dot{\phi})_{\rm slow} \simeq \frac{m\,g\,l}{I_\parallel\,\Omega}, \tag{8.122}$$

$$(\dot{\phi})_{\rm fast} \simeq \frac{I_\parallel\,\Omega}{I_\perp\,\cos\theta}. \tag{8.123}$$

The slower of these two precession rates is the one which is generally observed.

8.10 Rotational Stability

Consider a rigid body for which all of the principal moments of inertia are distinct. Let $I_{z'z'} > I_{y'y'} > I_{x'x'}$. Suppose that the body is rotating freely about one of its principal axes. What happens when the body is slightly disturbed?

Let the body be initially rotating about the x'-axis, so that

$$\boldsymbol{\omega} = \omega_{x'}\,\mathbf{e}_{x'}. \tag{8.124}$$

If we apply a slight perturbation then the angular velocity becomes

$$\boldsymbol{\omega} = \omega_{x'}\,\mathbf{e}_{x'} + \lambda\,\mathbf{e}_{y'} + \mu\,\mathbf{e}_{z'}, \tag{8.125}$$

where λ and μ are both assumed to be small. Euler's equations (8.51)–(8.53) take the form

$$I_{x'x'}\,\dot{\omega}_{x'} - (I_{y'y'} - I_{z'z'})\,\lambda\,\mu = 0, \tag{8.126}$$

$$I_{y'y'}\,\dot{\lambda} - (I_{z'z'} - I_{x'x'})\,\omega_{x'}\,\mu = 0, \tag{8.127}$$

$$I_{z'z'}\,\dot{\mu} - (I_{x'x'} - I_{y'y'})\,\omega_{x'}\,\lambda = 0. \tag{8.128}$$

Since $\lambda\,\mu$ is second-order in small quantities—and, therefore, negligible—the first of the above equations tells us that $\omega_{x'}$ is an approximate constant of the motion. The other two equations can be written

$$\dot{\lambda} = \left[\frac{(I_{z'z'} - I_{x'x'})\,\omega_{x'}}{I_{y'y'}}\right]\mu, \tag{8.129}$$

$$\dot{\mu} = -\left[\frac{(I_{y'y'} - I_{x'x'})\,\omega_{x'}}{I_{z'z'}}\right]\lambda. \tag{8.130}$$

Differentiating the first equation with respect to time, and then eliminating $\dot{\mu}$, we obtain

$$\ddot{\lambda} + \left[\frac{(I_{y'y'} - I_{x'x'})\,(I_{z'z'} - I_{x'x'})}{I_{y'y'}\,I_{z'z'}}\right]\omega_{x'}^2\,\lambda = 0. \tag{8.131}$$

It is easily demonstrated that μ satisfies the same differential equation. Since the term in square brackets in the above equation is *positive*, the equation takes the form of a *simple harmonic equation*, and, thus, has the bounded solution:

$$\lambda = \lambda_0\,\cos(\Omega_{x'}\,t' - \alpha). \tag{8.132}$$

Here, λ_0 and α are constants of integration, and

$$\Omega_{x'} = \left[\frac{(I_{y'y'} - I_{x'x'})\,(I_{z'z'} - I_{x'x'})}{I_{y'y'}\,I_{z'z'}}\right]^{1/2}\omega_{x'}. \tag{8.133}$$

Thus, the body oscillates sinusoidally about its initial state with the angular frequency $\Omega_{x'}$. It follows that the body is *stable* to small perturbations when rotating about the x'-axis, in the sense that the amplitude of such perturbations does not grow in time.

Suppose that the body is initially rotating about the z'-axis, and is subject to a small perturbation. A similar argument to the above allows us to conclude that the body oscillates sinusoidally about its initial state with angular frequency

$$\Omega_{z'} = \left[\frac{(I_{z'z'} - I_{x'x'})\,(I_{z'z'} - I_{y'y'})}{I_{x'x'}\,I_{y'y'}}\right]^{1/2} \omega_{z'}. \tag{8.134}$$

Hence, the body is also stable to small perturbations when rotating about the z'-axis.

Suppose, finally, that the body is initially rotating about the y'-axis, and is subject to a small perturbation, such that

$$\boldsymbol{\omega} = \lambda\,\mathbf{e}_{x'} + \omega_{y'}\,\mathbf{e}_{y'} + \mu\,\mathbf{e}_{z'}. \tag{8.135}$$

It is easily demonstrated that λ satisfies the following differential equation:

$$\ddot{\lambda} - \left[\frac{(I_{y'y'} - I_{x'x'})\,(I_{z'z'} - I_{y'y'})}{I_{x'x'}\,I_{z'z'}}\right] \omega_{y'}^2\,\lambda = 0. \tag{8.136}$$

Note that the term in square brackets is *positive*. Hence, the above equation is *not* the simple harmonic equation. Indeed its solution takes the form

$$\lambda = A\,e^{k\,t'} + B\,e^{-k\,t'}. \tag{8.137}$$

Here, A and B are constants of integration, and

$$k = \left[\frac{(I_{y'y'} - I_{x'x'})\,(I_{z'z'} - I_{y'y'})}{I_{x'x'}\,I_{z'z'}}\right]^{1/2} \omega_{y'}. \tag{8.138}$$

In this case, the amplitude of the perturbation grows exponentially in time. Hence, the body is *unstable* to small perturbations when rotating about the y'-axis.

In conclusion, a rigid body with three distinct principal moments of inertia is stable to small perturbations when rotating about the principal axes with the largest and smallest moments, but is unstable when rotating about the axis with the intermediate moment.

Finally, if two of the principal moments are the same then it can be shown that the body is only stable to small perturbations when rotating about the principal axis whose moment is distinct from the other two.

8.11 Exercises

8.1. Find the principal axes of rotation and the principal moments of inertia for a thin uniform rectangular plate of mass m and dimensions $2a$ by a for rotation about axes passing through (a) the center of mass, and (b) a corner.

8.2. A rigid body having an axis of symmetry rotates freely about a fixed point under no torques. If α is the angle between the symmetry axis and the instantaneous axis of rotation, show that the angle between the axis of rotation and the invariable line (the **L** vector) is

$$\tan^{-1}\left[\frac{(I_\parallel - I_\perp)\,\tan\alpha}{I_\parallel + I_\perp\,\tan^2\alpha}\right]$$

where $I_\parallel$ (the moment of inertia about the symmetry axis) is greater than $I_\perp$ (the moment of inertia about an axis normal to the symmetry axis).

8.3. Since the greatest value of $I_\parallel/I_\perp$ is 2 (symmetrical lamina) show from the previous result that the angle between the angular velocity and angular momentum vectors cannot exceed $\tan^{-1}(1/\sqrt{8}) \simeq 19.5^\circ$. Find the corresponding value of α.

8.4. A thin uniform rod of length l and mass m is constrained to rotate with constant angular velocity ω about an axis passing through the center of the rod, and making an angle α with the rod. Show that the angular momentum about the center of the rod is perpendicular to the rod, and is of magnitude $(m\,l^2\,\omega/12)\,\sin\alpha$. Show that the torque is perpendicular to both the rod and the angular momentum vector, and is of magnitude $(m\,l^2\omega^2/12)\,\sin\alpha\,\cos\alpha$.

8.5. A thin uniform disk of radius a and mass m is constrained to rotate with constant angular velocity ω about an axis passing through its center, and making an angle α with the normal to the disk. Find the angular momentum about the center of the disk, as well as the torque acting on the disk.

8.6. Demonstrate that for an isolated rigid body which possesses an axis of symmetry, and rotates about one of its principal axes, the motion is only stable to small perturbations if the principal axis is that which corresponds to the symmetry axis.

9 Lagrangian Dynamics

9.1 Introduction

This chapter describes an elegant reformulation of the laws of Newtonian dynamics which is due to the French/Italian mathematician Joseph Louis Lagrange. This reformulation is particularly useful for finding the equations of motion of complicated dynamical systems.

9.2 Generalized Coordinates

Let the q_i, for $i = 1, \mathcal{F}$, be a set of coordinates which uniquely specifies the instantaneous configuration of some dynamical system. Here, it is assumed that each of the q_i can vary *independently*. The q_i might be Cartesian coordinates, or polar coordinates, or angles, or some mixture of all three types of coordinate, and are, therefore, termed *generalized coordinates*. A dynamical system whose instantaneous configuration is fully specified by $\mathcal{F}$ independent generalized coordinates is said to have $\mathcal{F}$ *degrees of freedom*. For instance, the instantaneous position of a particle moving freely in three dimensions is completely specified by its three Cartesian coordinates, x, y, and z. Moreover, these coordinates are clearly independent of one another. Hence, a dynamical system consisting of a single particle moving freely in three dimensions has three degrees of freedom. If there are two freely moving particles then the system has six degrees of freedom, and so on.

Suppose that we have a dynamical system consisting of N particles moving freely in three dimensions. This is an $\mathcal{F} = 3\,\mathrm{N}$ degree of freedom system whose instantaneous configuration can be specified by $\mathcal{F}$ Cartesian coordinates. Let us denote these coordinates the x_j, for $j = 1, \mathcal{F}$. Thus, x_1, x_2, x_3 are the Cartesian coordinates of the first particle, x_4, x_5, x_6 the Cartesian coordinates of the second particle, *etc*. Suppose that the instantaneous configuration of the system can also be specified by $\mathcal{F}$ generalized coordinates, which we shall denote the q_i, for $i = 1, \mathcal{F}$. Thus, the q_i might be the spherical coordinates of the particles. In general, we expect the x_j to be functions of the q_i. In other words,

$$x_j = x_j(q_1, q_2, \cdots, q_{\mathcal{F}}, t) \tag{9.1}$$

for $j = 1, \mathcal{F}$. Here, for the sake of generality, we have included the possibility that the functional relationship between the x_j and the q_i might depend on the time, t, explicitly. This would be the case if the dynamical system were subject to time varying constraints. For instance, a system consisting of a particle constrained to move on a surface which is itself moving. Finally, by the chain rule, the variation of the x_j due to a variation of the q_i (at constant t) is given by

$$\delta x_j = \sum_{i=1,\mathcal{F}} \frac{\partial x_j}{\partial q_i}\,\delta q_i, \tag{9.2}$$

for $j = 1, \mathcal{F}$.

9.3 Generalized Forces

The work done on the dynamical system when its Cartesian coordinates change by δx_j is simply

$$\delta W = \sum_{j=1,\mathcal{F}} f_j \, \delta x_j \tag{9.3}$$

Here, the f_j are the Cartesian components of the forces acting on the various particles making up the system. Thus, f_1, f_2, f_3 are the components of the force acting on the first particle, f_4, f_5, f_6 the components of the force acting on the second particle, *etc*. Using Equation (9.2), we can also write

$$\delta W = \sum_{j=1,\mathcal{F}} f_j \sum_{i=1,\mathcal{F}} \frac{\partial x_j}{\partial q_i} \, \delta q_i. \tag{9.4}$$

The above expression can be rearranged to give

$$\delta W = \sum_{i=1,\mathcal{F}} Q_i \, \delta q_i, \tag{9.5}$$

where

$$Q_i = \sum_{j=1,\mathcal{F}} f_j \, \frac{\partial x_j}{\partial q_i}. \tag{9.6}$$

Here, the Q_i are termed *generalized forces*. Note that a generalized force does not necessarily have the dimensions of force. However, the product $Q_i \, q_i$ must have the dimensions of work. Thus, if a particular q_i is a Cartesian coordinate then the associated Q_i is a force. Conversely, if a particular q_i is an angle then the associated Q_i is a torque.

Suppose that the dynamical system in question is *conservative*. It follows that

$$f_j = -\frac{\partial U}{\partial x_j}, \tag{9.7}$$

for $j = 1, \mathcal{F}$, where $U(x_1, x_2, \cdots, x_{\mathcal{F}}, t)$ is the system's potential energy. Hence, according to Equation (9.6),

$$Q_i = -\sum_{j=1,\mathcal{F}} \frac{\partial U}{\partial x_j} \, \frac{\partial x_j}{\partial q_i} = -\frac{\partial U}{\partial q_i}, \tag{9.8}$$

for $i = 1, \mathcal{F}$.

9.4 Lagrange's Equation

The Cartesian equations of motion of our system take the form

$$m_j \, \ddot{x}_j = f_j, \tag{9.9}$$

for $j = 1, \mathcal{F}$, where m_1, m_2, m_3 are each equal to the mass of the first particle, m_4, m_5, m_6 are each equal to the mass of the second particle, *etc.* Furthermore, the kinetic energy of the system can be written

$$K = \frac{1}{2} \sum_{j=1,\mathcal{F}} m_j \, \dot{x}_j^2. \tag{9.10}$$

Now, since $x_j = x_j(q_1, q_2, \cdots, q_{\mathcal{F}}, t)$, we can write

$$\dot{x}_j = \sum_{i=1,\mathcal{F}} \frac{\partial x_j}{\partial q_i} \, \dot{q}_i + \frac{\partial x_j}{\partial t}, \tag{9.11}$$

for $j = 1, \mathcal{F}$. Hence, it follows that $\dot{x}_j = \dot{x}_j(\dot{q}_1, \dot{q}_2, \cdots, \dot{q}_{\mathcal{F}}, q_1, q_2, \cdots, q_{\mathcal{F}}, t)$. According to the above equation,

$$\frac{\partial \dot{x}_j}{\partial \dot{q}_i} = \frac{\partial x_j}{\partial q_i}, \tag{9.12}$$

where we are treating the $\dot{q}_i$ and the q_i as *independent* variables.

Multiplying Equation (9.12) by $\dot{x}_j$, and then differentiating with respect to time, we obtain

$$\frac{d}{dt}\left(\dot{x}_j \frac{\partial \dot{x}_j}{\partial \dot{q}_i}\right) = \frac{d}{dt}\left(\dot{x}_j \frac{\partial x_j}{\partial q_i}\right) = \ddot{x}_j \frac{\partial x_j}{\partial q_i} + \dot{x}_j \frac{d}{dt}\left(\frac{\partial x_j}{\partial q_i}\right). \tag{9.13}$$

Now,

$$\frac{d}{dt}\left(\frac{\partial x_j}{\partial q_i}\right) = \sum_{k=1,\mathcal{F}} \frac{\partial^2 x_j}{\partial q_i \, \partial q_k} \, \dot{q}_k + \frac{\partial^2 x_j}{\partial q_i \, \partial t}. \tag{9.14}$$

Furthermore,

$$\frac{1}{2} \frac{\partial \dot{x}_j^2}{\partial \dot{q}_i} = \dot{x}_j \frac{\partial \dot{x}_j}{\partial \dot{q}_i}, \tag{9.15}$$

and

$$\begin{aligned} \frac{1}{2} \frac{\partial \dot{x}_j^2}{\partial q_i} = \dot{x}_j \frac{\partial \dot{x}_j}{\partial q_i} &= \dot{x}_j \frac{\partial}{\partial q_i}\left(\sum_{k=1,\mathcal{F}} \frac{\partial x_j}{\partial q_k} \, \dot{q}_k + \frac{\partial x_j}{\partial t}\right) \\ &= \dot{x}_j \left(\sum_{k=1,\mathcal{F}} \frac{\partial^2 x_j}{\partial q_i \, \partial q_k} \, \dot{q}_k + \frac{\partial^2 x_j}{\partial q_i \, \partial t}\right) \\ &= \dot{x}_j \frac{d}{dt}\left(\frac{\partial x_j}{\partial q_i}\right), \end{aligned} \tag{9.16}$$

where use has been made of Equation (9.14). Thus, it follows from Equations (9.13), (9.15), and (9.16) that

$$\frac{d}{dt}\left(\frac{1}{2} \frac{\partial \dot{x}_j^2}{\partial \dot{q}_i}\right) = \ddot{x}_j \frac{\partial x_j}{\partial q_i} + \frac{1}{2} \frac{\partial \dot{x}_j^2}{\partial q_i}. \tag{9.17}$$

Let us take the above equation, multiply by m_j, and then sum over all j. We obtain

$$\frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_i}\right) = \sum_{j=1,\mathcal{F}} f_j \frac{\partial x_j}{\partial q_i} + \frac{\partial K}{\partial q_i}, \tag{9.18}$$

where use has been made of Equations (9.9) and (9.10). Thus, it follows from Equation (9.6) that

$$\frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_i}\right) = Q_i + \frac{\partial K}{\partial q_i}. \tag{9.19}$$

Finally, making use of Equation (9.8), we get

$$\frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_i}\right) = -\frac{\partial U}{\partial q_i} + \frac{\partial K}{\partial q_i}. \tag{9.20}$$

It is helpful to introduce a function L, called the *Lagrangian*, which is defined as the difference between the kinetic and potential energies of the dynamical system under investigation:

$$L = K - U. \tag{9.21}$$

Since the potential energy U is clearly independent of the $\dot{q}_i$, it follows from Equation (9.20) that

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0, \tag{9.22}$$

for $i = 1, \mathcal{F}$. This equation is known as *Lagrange's equation*.

According to the above analysis, if we can express the kinetic and potential energies of our dynamical system solely in terms of our generalized coordinates and their time derivatives then we can immediately write down the equations of motion of the system, expressed in terms of the generalized coordinates, using Lagrange's equation, (9.22). Unfortunately, this scheme only works for conservative systems. Let us now consider some examples.

9.5 Motion in a Central Potential

Consider a particle of mass m moving in two dimensions in the central potential $U(r)$. This is clearly a two degree of freedom dynamical system. As described in Section 5.5, the particle's instantaneous position is most conveniently specified in terms of the plane polar coordinates r and θ. These are our two generalized coordinates. According to Equation (5.14), the square of the particle's velocity can be written

$$v^2 = \dot{r}^2 + (r\,\dot{\theta})^2. \tag{9.23}$$

Hence, the Lagrangian of the system takes the form

$$L = \frac{1}{2}\, m\, (\dot{r}^2 + r^2\, \dot{\theta}^2) - U(r). \tag{9.24}$$

Note that

$$\frac{\partial L}{\partial \dot{r}} = m\,\dot{r}, \qquad \frac{\partial L}{\partial r} = m\,r\,\dot{\theta}^2 - \frac{dU}{dr}, \tag{9.25}$$

$$\frac{\partial L}{\partial \dot{\theta}} = m\,r^2\,\dot{\theta}, \qquad \frac{\partial L}{\partial \theta} = 0. \tag{9.26}$$

Now, Lagrange's equation (9.22) yields the equations of motion,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{r}}\right) - \frac{\partial L}{\partial r} = 0, \tag{9.27}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = 0. \tag{9.28}$$

Hence, we obtain

$$\frac{d}{dt}(m\,\dot{r}) - m\,r\,\dot{\theta}^2 + \frac{dU}{dr} = 0, \tag{9.29}$$

$$\frac{d}{dt}\left(m\,r^2\,\dot{\theta}\right) = 0, \tag{9.30}$$

or

$$\ddot{r} - r\,\dot{\theta}^2 = -\frac{dV}{dr}, \tag{9.31}$$

$$r^2\,\dot{\theta} = h, \tag{9.32}$$

where $V = U/m$, and h is a constant. We recognize Equations (9.31) and (9.32) as the equations we derived in Chapter 5 for motion in a central potential. The advantage of the Lagrangian method of deriving these equations is that we avoid having to express the acceleration in terms of the generalized coordinates r and θ.

9.6 Atwood Machines

An Atwood machine consists of two weights, of mass m_1 and m_2, connected by a light inextensible cord of length l, which passes over a pulley of radius $a \ll l$, and moment of inertia I. See Figure 9.1.

Referring to the diagram, we can see that this is a one degree of freedom system whose instantaneous configuration is specified by the coordinate x. Assuming that the cord does not slip with respect to the pulley, the angular velocity of pulley is $\dot{x}/a$. Hence, the kinetic energy of the system is given by

$$K = \frac{1}{2}\,m_1\,\dot{x}^2 + \frac{1}{2}\,m_2\,\dot{x}^2 + \frac{1}{2}\,I\,\frac{\dot{x}^2}{a^2}. \tag{9.33}$$

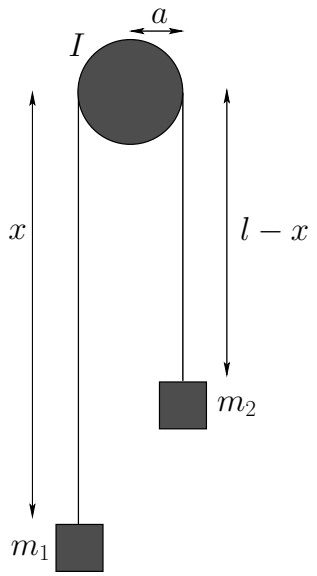


Figure 9.1: *An Atwood machine.*

The potential energy of the system takes the form

$$U = -m_1\, g\, x - m_2\, g\, (l - x). \tag{9.34}$$

It follows that the Lagrangian is written

$$L = \frac{1}{2}\left(m_1 + m_2 + \frac{I}{a^2}\right)\dot{x}^2 + g\,(m_1 - m_2)\,x + \text{const.} \tag{9.35}$$

The equation of motion,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 0, \tag{9.36}$$

thus yields

$$\left(m_1 + m_2 + \frac{I}{a^2}\right)\ddot{x} - g\,(m_1 - m_2) = 0, \tag{9.37}$$

or

$$\ddot{x} = \frac{g\,(m_1 - m_2)}{m_1 + m_2 + I/a^2}, \tag{9.38}$$

which is the correct answer.

Consider the dynamical system drawn in Figure 9.2. This is an Atwood machine in which one of the weights has been replaced by a second Atwood machine with a cord of length l'. The system now has two degrees of freedom, and its instantaneous position is specified by the two coordinates x and x', as shown.

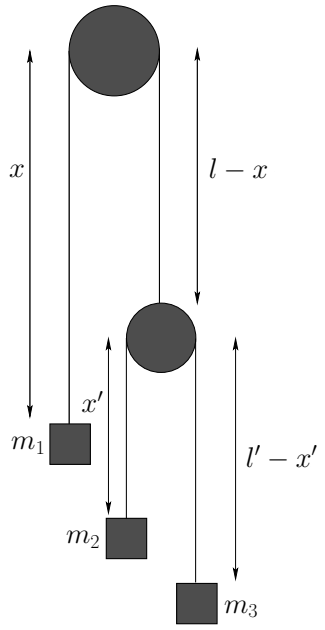


Figure 9.2: *A double Atwood machine.*

For the sake of simplicity, let us neglect the masses of the two pulleys. Thus, the kinetic energy of the system is written

$$K = \frac{1}{2}\, m_1\, \dot{x}^2 + \frac{1}{2}\, m_2\, (-\dot{x} + \dot{x}')^2 + \frac{1}{2}\, m_3\, (-\dot{x} - \dot{x}')^2, \tag{9.39}$$

whereas the potential energy takes the form

$$U = -m_1\, g\, x - m_2\, g\, (l - x + x') - m_3\, g\, (l - x + l' - x'). \tag{9.40}$$

It follows that the Lagrangian of the system is

$$\begin{aligned} L &= \frac{1}{2}\, m_1\, \dot{x}^2 + \frac{1}{2}\, m_2\, (-\dot{x} + \dot{x}')^2 + \frac{1}{2}\, m_3\, (-\dot{x} - \dot{x}')^2 \\ &\quad + g\, (m_1 - m_2 - m_3)\, x + g\, (m_2 - m_3)\, x' + \text{const.} \end{aligned} \tag{9.41}$$

Hence, the equations of motion,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 0, \tag{9.42}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}'}\right) - \frac{\partial L}{\partial x'} = 0, \tag{9.43}$$

yield

$$m_1\, \ddot{x} + m_2\, (\ddot{x} - \ddot{x}') + m_3\, (\ddot{x} + \ddot{x}') - g\, (m_1 - m_2 - m_3) = 0, \tag{9.44}$$

$$m_2\, (-\ddot{x} + \ddot{x}') + m_3\, (\ddot{x} + \ddot{x}') - g\, (m_2 - m_3) = 0. \tag{9.45}$$

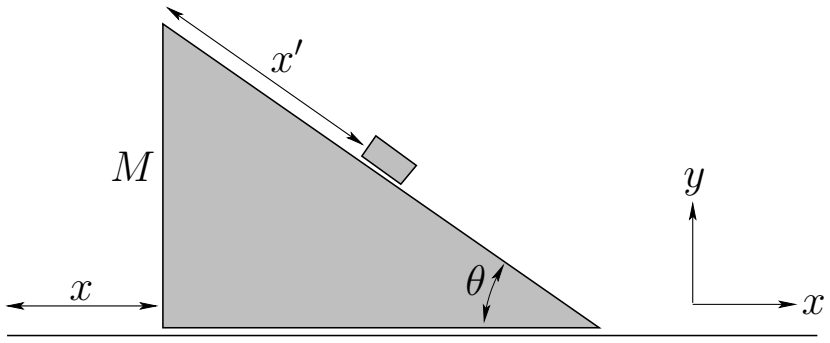


Figure 9.3: *A sliding plane.*

The accelerations $\ddot{x}$ and $\ddot{x}'$ can be obtained from the above two equations via simple algebra.

9.7 Sliding down a Sliding Plane

Consider the case of a particle of mass m sliding down a smooth inclined plane of mass M which is, itself, free to slide on a smooth horizontal surface, as shown in Figure 9.3. This is a two degree of freedom system, so we need two coordinates to specify the configuration. Let us choose x, the horizontal distance of the plane from some reference point, and x', the parallel displacement of the particle from some reference point on the plane.

Defining x- and y-axes, as shown in the diagram, the x- and y-components of the particle's velocity are clearly given by

$$v_x = \dot{x} + \dot{x}' \cos\theta, \tag{9.46}$$

$$v_y = -\dot{x}' \sin\theta, \tag{9.47}$$

respectively, where θ is the angle of inclination of the plane with respect to the horizontal. Thus,

$$v^2 = v_x^2 + v_y^2 = \dot{x}^2 + 2\,\dot{x}\,\dot{x}' \cos\theta + \dot{x}'^2. \tag{9.48}$$

Hence, the kinetic energy of the system takes the form

$$K = \frac{1}{2} M \dot{x}^2 + \frac{1}{2} m (\dot{x}^2 + 2\,\dot{x}\,\dot{x}' \cos\theta + \dot{x}'^2), \tag{9.49}$$

whereas the potential energy is given by

$$U = -m\,g\,x' \sin\theta + \text{const.} \tag{9.50}$$

It follows that the Lagrangian is written

$$L = \frac{1}{2} M \dot{x}^2 + \frac{1}{2} m (\dot{x}^2 + 2\,\dot{x}\,\dot{x}' \cos\theta + \dot{x}'^2) + m\,g\,x' \sin\theta + \text{const.} \tag{9.51}$$

The equations of motion,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 0, \tag{9.52}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}'}\right) - \frac{\partial L}{\partial x'} = 0, \tag{9.53}$$

thus yield

$$M\,\ddot{x} + m\,(\ddot{x} + \ddot{x}'\cos\theta) = 0, \tag{9.54}$$

$$m\,(\ddot{x}' + \ddot{x}\cos\theta) - m\,g\sin\theta = 0. \tag{9.55}$$

Finally, solving for $\ddot{x}$ and $\ddot{x}'$, we obtain

$$\ddot{x} = -\frac{g\sin\theta\cos\theta}{(m+M)/m - \cos^2\theta}, \tag{9.56}$$

$$\ddot{x}' = \frac{g\sin\theta}{1 - m\cos^2\theta/(m+M)}. \tag{9.57}$$

9.8 Generalized Momenta

Consider the motion of a single particle moving in one dimension. The kinetic energy is

$$K = \frac{1}{2}\,m\,\dot{x}^2, \tag{9.58}$$

where m is the mass of the particle, and x its displacement. Now, the particle's linear momentum is $p = m\,\dot{x}$. However, this can also be written

$$p = \frac{\partial K}{\partial \dot{x}} = \frac{\partial L}{\partial \dot{x}}, \tag{9.59}$$

since $L = K - U$, and the potential energy U is independent of $\dot{x}$.

Consider a dynamical system described by $\mathcal{F}$ generalized coordinates q_i, for $i = 1, \mathcal{F}$. By analogy with the above expression, we can define *generalized momenta* of the form

$$p_i = \frac{\partial L}{\partial \dot{q}_i}, \tag{9.60}$$

for $i = 1, \mathcal{F}$. Here, p_i is sometimes called the momentum *conjugate* to the coordinate q_i. Hence, Lagrange's equation (9.22) can be written

$$\frac{dp_i}{dt} = \frac{\partial L}{\partial q_i}, \tag{9.61}$$

for $i = 1, \mathcal{F}$. Note that a generalized momentum does not necessarily have the dimensions of linear momentum.

Suppose that the Lagrangian L does not depend explicitly on some coordinate q_k. It follows from Equation (9.61) that

$$\frac{dp_k}{dt} = \frac{\partial L}{\partial q_k} = 0. \tag{9.62}$$

Hence,

$$p_k = \text{const.} \tag{9.63}$$

The coordinate q_k is said to be *ignorable* in this case. Thus, we conclude that the generalized momentum associated with an ignorable coordinate is a constant of the motion.

For example, in Section 9.5, the Lagrangian (9.24) for a particle moving in a central potential is independent of the angular coordinate θ. Thus, θ is an ignorable coordinate, and

$$p_\theta = \frac{\partial L}{\partial \dot{\theta}} = m\, r^2\, \dot{\theta} \tag{9.64}$$

is a constant of the motion. Of course, p_θ is the angular momentum about the origin. This is conserved because a central force exerts no torque about the origin.

Again, in Section 9.7, the Lagrangian (9.51) for a mass sliding down a sliding slope is independent of the Cartesian coordinate x. It follows that x is an ignorable coordinate, and

$$p_x = \frac{\partial L}{\partial \dot{x}} = M\,\dot{x} + m\,(\dot{x} + \dot{x}'\,\cos\theta) \tag{9.65}$$

is a constant of the motion. Of course, p_x is the total linear momentum in the x-direction. This is conserved because there is no external force acting on the system in the x-direction.

9.9 Spherical Pendulum

Consider a pendulum consisting of a compact mass m on the end of light inextensible string of length l. Suppose that the mass is free to move in any direction (as long as the string remains taut). Let the fixed end of the string be located at the origin of our coordinate system. We can define Cartesian coordinates, (x, y, z), such that the z-axis points vertically upward. We can also define spherical coordinates, (r, θ, ϕ), whose axis points along the $-z$-axis. The latter coordinates are the most convenient, since r is constrained to always take the value l. However, the two angular coordinates, θ and ϕ, are free to vary independently. Hence, this is a two degree of freedom system.

The Cartesian coordinates can be written in terms of the angular coordinates θ and ϕ. In fact,

$$x = l\,\sin\theta\,\cos\phi, \tag{9.66}$$

$$y = l\,\sin\theta\,\sin\phi. \tag{9.67}$$

$$z = -l\,\cos\theta. \tag{9.68}$$

Hence, the potential energy of the system is

$$U = m\,g\,z = -m\,g\,l\,\cos\theta. \tag{9.69}$$

Also,

$$v^2 = \dot{x}^2 + \dot{y}^2 + \dot{z}^2 = l^2\,(\dot{\theta}^2 + \sin^2\theta\,\dot{\phi}^2). \tag{9.70}$$

Thus, the Lagrangian of the system is written

$$L = \frac{1}{2}\,m\,l^2\,(\dot{\theta}^2 + \sin^2\theta\,\dot{\phi}^2) + m\,g\,l\,\cos\theta. \tag{9.71}$$

Note that the Lagrangian is independent of the angular coordinate ϕ. It follows that

$$p_\phi = \frac{\partial L}{\partial \dot{\phi}} = m\,l^2\,\sin^2\theta\,\dot{\phi} \tag{9.72}$$

is a constant of the motion. Of course, p_ϕ is the angular momentum of the system about the z-axis. This is conserved because neither the tension in the string nor the force of gravity exert a torque about the z-axis. Conservation of angular momentum about the z-axis implies that

$$\sin^2\theta\,\dot{\phi} = h, \tag{9.73}$$

where h is a constant.

The equation of motion of the system,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = 0, \tag{9.74}$$

yields

$$\ddot{\theta} + \frac{g}{l}\,\sin\theta - \sin\theta\,\cos\theta\,\dot{\phi}^2 = 0, \tag{9.75}$$

or

$$\ddot{\theta} + \frac{g}{l}\,\sin\theta - h^2\,\frac{\cos\theta}{\sin^3\theta} = 0, \tag{9.76}$$

where use has been made of Equation (9.73).

Suppose that $\phi = \phi_0 = \text{const}$. It follows that $\dot{\phi} = h = 0$. Hence, Equation (9.76) yields

$$\ddot{\theta} + \frac{g}{l}\,\sin\theta = 0. \tag{9.77}$$

This, of course, is the equation of a simple pendulum whose motion is restricted to the vertical plane $\phi = \phi_0$—see Section 3.10.

Suppose that $\theta = \theta_0 = \text{const}$. It follows from Equation (9.73) that $\dot{\phi} = \dot{\phi}_0 = \text{const.}$: *i.e.*, the pendulum bob rotates uniformly in a horizontal plane. According to Equations (9.73) and (9.76),

$$\dot{\phi}_0 = \sqrt{\frac{g}{d}}, \tag{9.78}$$

where $d = l\cos\theta_0$ is the vertical distance of the plane of rotation below the pivot point. This type of pendulum is usually called a *conical pendulum*, since the string attached to the pendulum bob sweeps out a cone as the bob rotates.

Suppose, finally, that the motion is almost conical: *i.e.*, the value of θ remains close to the value θ_0. Let

$$\theta = \theta_0 + \delta\theta. \tag{9.79}$$

Taylor expanding Equation (9.76) to first order in $\delta\theta$, the zeroth order terms cancel out, and we are left with

$$\delta\ddot{\theta} + \dot{\phi}_0^2\,(1 + 3\,\cos^2\theta_0)\,\delta\theta \simeq 0. \tag{9.80}$$

Hence, solving the above equation, we obtain

$$\theta \simeq \theta_0 + \delta\theta_0\,\cos(\Omega\,t), \tag{9.81}$$

where

$$\Omega = \dot{\phi}_0\,\sqrt{1 + 3\,\cos^2\theta_0}. \tag{9.82}$$

Thus, the angle θ executes simple harmonic motion about its mean value θ_0 at the angular frequency Ω.

Now the azimuthal angle, ϕ, increases by

$$\Delta\phi \simeq \dot{\phi}_0\,\frac{\pi}{\Omega} = \frac{\pi}{\sqrt{1 + 3\,\cos^2\theta_0}} \tag{9.83}$$

as the angle of inclination to the vertical, θ, goes between successive maxima and minima. Suppose that θ_0 is small. In this case, $\Delta\phi$ is slightly greater than $\pi/2$. Now if $\Delta\phi$ were exactly $\pi/2$ then the pendulum bob would trace out the outline of a slightly *warped circle*: *i.e.*, something like the outline of a potato chip or a saddle. The fact that $\Delta\phi$ is slightly greater than $\pi/2$ means that this shape *precesses* about the z-axis in the *same* direction as the direction rotation of the bob. The precession rate increases as the angle of inclination θ_0 increases. Suppose, now, that θ_0 is slightly less than $\pi/2$. (Of course, θ_0 can never exceed $\pi/2$). In this case, $\Delta\phi$ is slightly less than π. Now if $\Delta\phi$ were exactly π then the pendulum bob would trace out the outline of a slightly *tilted circle*. The fact that $\Delta\phi$ is slightly less than π means that this shape *precesses* about the z-axis in the *opposite* direction to the direction of rotation of the bob. The precession rate increases as the angle of inclination θ_0 decreases below $\pi/2$.

9.10 Exercises

9.1. A horizontal rod AB rotates with constant angular velocity ω about its midpoint O. A particle P is attached to it by equal strings AP, BP. If θ is the inclination of the plane APB to the vertical, prove that

$$\frac{d^2\theta}{dt^2} - \omega^2\,\sin\theta\,\cos\theta = -\frac{g}{l}\,\sin\theta,$$

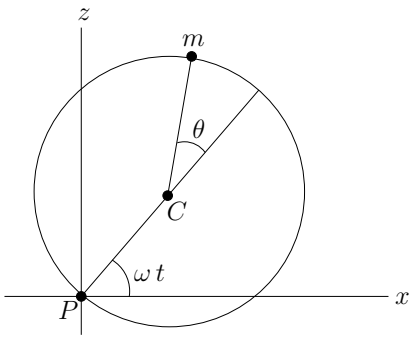
where $l = OP$. Deduce the condition that the vertical position of OP should be stable.

9.2. A double pendulum consists of two simple pendula, with one pendulum suspended from the bob of the other. If the two pendula have equal lengths, l, and have bobs of equal mass, m, and if both pendula are confined to move in the same vertical plane, find Lagrange's equations of motion for the system. Use θ and ϕ—the angles the upper and lower pendulums make with the downward vertical (respectively)—as the generalized coordinates. Do not assume small angles.

9.3. The surface of the Diskworld is a disk of radius R which rotates uniformly about a perpendicular axis passing through its center with angular velocity Ω. Diskworld gravitational acceleration is of magnitude g, and is everywhere directed normal to the disk. Find the Lagrangian of a projectile of mass m using co-rotating cylindrical polar coordinates as the generalized coordinates. What are the momenta conjugate to each coordinate? Are any of these momenta conserved? Find Lagrange's equations of motion for the projectile.

9.4. Find Lagrange's equations of motion for an elastic pendulum consisting of a particle of mass m attached to an elastic string of stiffness k and unstretched length l_0. Assume that the motion takes place in a vertical plane.

9.5. A disk of mass M and radius R rolls without slipping down a plane inclined at an angle α to the horizontal. The disk has a short weightless axle of negligible radius. From this axle is suspended a simple pendulum of length $l < R$ whose bob is of mass m. Assume that the motion of the pendulum takes place in the plane of the disk. Find Lagrange's equations of motion of the system.



9.6. A vertical circular hoop of radius a is rotated in a vertical plane about a point P on its circumference at the constant angular velocity ω. A bead of mass m slides without friction on the hoop. Find the kinetic energy, the potential energy, the Lagrangian, and Largrange's equation of motion of the bead, respectively, in terms of the angular coordinate θ shown in the above diagram. Here, x is a horizontal Cartesian coordinate, z a vertical Cartesian coordinate, and C the center of the hoop.

9.7. Consider a spherical pendulum of length l. Suppose that the string is initially horizontal, and the bob is rotating horizontally with tangental velocity v. Demonstrate that, at its lowest subsequent point, the bob will have fallen a vertical height $l\,e^{-u}$, where

$$\sinh u = \frac{v^2}{4\,g\,l}.$$

Show that if v^2 is large compared to $4\,g\,l$ then this result becomes approximately $2\,g\,l^2/v^2$.

9.8. The kinetic energy of a rotating rigid object with an axis of symmetry can be written

$$K = \frac{1}{2}\left[I_\perp\,\dot{\theta}^{\,2} + (I_\perp\,\sin^2\theta + I_\parallel\,\cos^2\theta)\,\dot{\phi}^{\,2} + 2\,I_\parallel\,\cos\theta\,\dot{\phi}\,\dot{\psi} + I_\parallel\,\dot{\psi}^{\,2}\right],$$

where $I_\parallel$ is the moment of inertia about the symmetry axis, $I_\perp$ is the moment of inertia about an axis perpendicular to the symmetry axis, and θ, ϕ, ψ are the three Euler angles. Suppose that the object is rotating freely. Find the momenta conjugate to the Euler angles. Which of these momenta are conserved? Find Lagrange's equations of motion for the system. Demonstrate that if the system is precessing steadily (which implies that θ, $\dot{\phi}$, and $\dot{\psi}$ are constants) then

$$\dot{\psi} = \left(\frac{I_\perp - I_\parallel}{I_\parallel}\right)\cos\theta\,\dot{\phi}.$$

9.9. Consider a nonconservative system in which the dissipative forces take the form $f_i = -k_i\,\dot{x}_i$, where the x_i are Cartesian coordinates, and the k_i are all positive. Demonstrate that the dissipative forces can be incorporated into the Lagrangian formalism provided that Lagrange's equations of motion are modified to read

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} + \frac{\partial R}{\partial \dot{q}_i} = 0,$$

where

$$R = \frac{1}{2}\sum_i k_i\,\dot{x}_i^2$$

is termed the Rayleigh Dissipation Function.

10 Hamiltonian Dynamics

10.1 Introduction

This chapter investigates the application of variational principles to Newtonian dynamics.

10.2 Calculus of Variations

It is a well-known fact, first enunciated by Archimedes, that the shortest distance between two points in a plane is a straight-line. However, suppose that we wish to demonstrate this result from first principles. Let us consider the length, l, of various curves, $y(x)$, which run between two fixed points, A and B, in a plane, as illustrated in Figure 10.1. Now, l takes the form

$$l = \int_A^B [dx^2 + dy^2]^{1/2} = \int_a^b [1 + y'^2(x)]^{1/2}\, dx, \tag{10.1}$$

where $y' \equiv dy/dx$. Note that l is a function of the function $y(x)$. In mathematics, a function of a function is termed a *functional*.

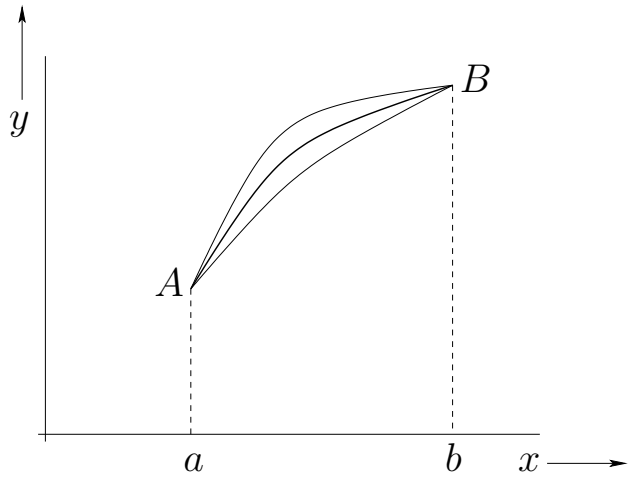


Figure 10.1: *Different paths between points* A *and* B.

Now, in order to find the shortest path between points A and B, we need to *minimize* the functional l with respect to small variations in the function $y(x)$, subject to the constraint that the end points, A and B, remain fixed. In other words, we need to solve

$$\delta l = 0. \tag{10.2}$$

The meaning of the above equation is that if $y(x) \rightarrow y(x) + \delta y(x)$, where $\delta y(x)$ is small, then the *first-order variation* in l, denoted δl, vanishes. In other words, $l \rightarrow l + \mathcal{O}(\delta y^2)$. The

particular function $y(x)$ for which $\delta I = 0$ obviously yields an *extremum* of I (*i.e.*, either a maximum or a minimum). Hopefully, in the case under consideration, it yields a minimum of I.

Consider a general functional of the form

$$I = \int_a^b F(y, y', x)\, dx, \tag{10.3}$$

where the end points of the integration are fixed. Suppose that $y(x) \rightarrow y(x) + \delta y(x)$. The first-order variation in I is written

$$\delta I = \int_a^b \left(\frac{\partial F}{\partial y}\, \delta y + \frac{\partial F}{\partial y'}\, \delta y' \right) dx, \tag{10.4}$$

where $\delta y' = d(\delta y)/dx$. Setting δI to zero, we obtain

$$\int_a^b \left(\frac{\partial F}{\partial y}\, \delta y + \frac{\partial F}{\partial y'}\, \delta y' \right) dx = 0. \tag{10.5}$$

This equation must be satisfied for all possible small perturbations $\delta y(x)$.

Integrating the second term in the integrand of the above equation by parts, we get

$$\int_a^b \left[\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'} \right) \right] \delta y\, dx + \left[\frac{\partial F}{\partial y'}\, \delta y \right]_a^b = 0. \tag{10.6}$$

Now, if the end points are fixed then $\delta y = 0$ at $x = a$ and $x = b$. Hence, the last term on the left-hand side of the above equation is zero. Thus, we obtain

$$\int_a^b \left[\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'} \right) \right] \delta y\, dx = 0. \tag{10.7}$$

The above equation must be satisfied for *all* small perturbations $\delta y(x)$. The only way in which this is possible is for the expression enclosed in square brackets in the integral to be zero. Hence, the functional I attains an extremum value whenever

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'} \right) - \frac{\partial F}{\partial y} = 0. \tag{10.8}$$

This condition is known as the *Euler-Lagrange equation*.

Let us consider some special cases. Suppose that F does not explicitly depend on y. It follows that $\partial F/\partial y = 0$. Hence, the Euler-Lagrange equation (10.8) simplifies to

$$\frac{\partial F}{\partial y'} = \text{const}. \tag{10.9}$$

Next, suppose that F does not depend explicitly on x. Multiplying Equation (10.8) by y', we obtain

$$y' \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - y' \frac{\partial F}{\partial y} = 0. \tag{10.10}$$

However,

$$\frac{d}{dx}\left(y' \frac{\partial F}{\partial y'}\right) = y' \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) + y'' \frac{\partial F}{\partial y'}. \tag{10.11}$$

Thus, we get

$$\frac{d}{dx}\left(y' \frac{\partial F}{\partial y'}\right) = y' \frac{\partial F}{\partial y} + y'' \frac{\partial F}{\partial y'}. \tag{10.12}$$

Now, if F is not an explicit function of x then the right-hand side of the above equation is the total derivative of F, namely dF/dx. Hence, we obtain

$$\frac{d}{dx}\left(y' \frac{\partial F}{\partial y'}\right) = \frac{dF}{dx}, \tag{10.13}$$

which yields

$$y' \frac{\partial F}{\partial y'} - F = \text{const}. \tag{10.14}$$

Returning to the case under consideration, we have $F = \sqrt{1 + y'^2}$, according to Equation (10.1) and (10.3). Hence, F is not an explicit function of y, so Equation (10.9) yields

$$\frac{\partial F}{\partial y'} = \frac{y'}{\sqrt{1 + y'^2}} = c, \tag{10.15}$$

where c is a constant. So,

$$y' = \frac{c}{\sqrt{1 - c^2}} = \text{const}. \tag{10.16}$$

Of course, $y' =$ constant is the equation of a *straight-line*. Thus, the shortest distance between two fixed points in a plane is indeed a straight-line.

10.3 Conditional Variation

Suppose that we wish to find the function $y(x)$ which maximizes or minimizes the functional

$$I = \int_a^b F(y, y', x)\, dx, \tag{10.17}$$

subject to the *constraint* that the value of

$$J = \int_a^b G(y, y', x)\, dx \tag{10.18}$$

remains constant. We can achieve our goal by finding an extremum of the new functional $K = I + \lambda J$, where $\lambda(x)$ is an undetermined function. We know that $\delta J = 0$, since the value of J is fixed, so if $\delta K = 0$ then $\delta I = 0$ as well. In other words, finding an extremum of K is equivalent to finding an extremum of I. Application of the Euler-Lagrange equation yields

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} + \left[\frac{d}{dx}\left(\frac{\partial [\lambda\, G]}{\partial y'}\right) - \frac{\partial [\lambda\, G]}{\partial y}\right] = 0. \tag{10.19}$$

In principle, the above equation, together with the constraint (10.18), yields the functions $\lambda(x)$ and $y(x)$. Incidentally, λ is generally termed a *Lagrange multiplier*. If F and G have no explicit x-dependence then λ is usually a *constant*.

As an example, consider the following famous problem. Suppose that a uniform chain of fixed length l is suspended by its ends from two equal-height fixed points which are a distance a apart, where $a < l$. What is the equilibrium configuration of the chain?

Suppose that the chain has the uniform density per unit length ρ. Let the x- and y-axes be horizontal and vertical, respectively, and let the two ends of the chain lie at $(\pm a/2, 0)$. The equilibrium configuration of the chain is specified by the function $y(x)$, for $-a/2 \leq x \leq +a/2$, where $y(x)$ is the vertical distance of the chain below its end points at horizontal position x. Of course, $y(-a/2) = y(+a/2) = 0$.

According to the discussion in Section 3.2, the stable equilibrium state of a conservative dynamical system is one which *minimizes* the system's potential energy. Now, the potential energy of the chain is written

$$U = -\rho\, g \int y\, ds = -\rho\, g \int_{-a/2}^{a/2} y\, [1 + y'^2]^{1/2}\, dx, \tag{10.20}$$

where $ds = \sqrt{dx^2 + dy^2}$ is an element of length along the chain, and g is the acceleration due to gravity. Hence, we need to minimize U with respect to small variations in $y(x)$. However, the variations in $y(x)$ must be such as to conserve the fixed length of the chain. Hence, our minimization procedure is subject to the constraint that

$$l = \int ds = \int_{-a/2}^{a/2} [1 + y'^2]^{1/2}\, dx \tag{10.21}$$

remains constant.

It follows, from the above discussion, that we need to minimize the functional

$$K = U + \lambda\, l = \int_{-a/2}^{a/2} (-\rho\, g\, y + \lambda)\, [1 + y'^2]^{1/2}\, dx, \tag{10.22}$$

where λ is an, as yet, undetermined constant. Since the integrand in the functional does not depend explicitly on x, we have from Equation (10.14) that

$$y'^2\, (-\rho\, g\, y + \lambda)\, [1 + y'^2]^{-1/2} - (-\rho\, g\, y + \lambda)\, [1 + y'^2]^{1/2} = k, \tag{10.23}$$

where k is a constant. This expression reduces to

$$y'^2 = \left(\lambda' + \frac{y}{h}\right)^2 - 1, \tag{10.24}$$

where $\lambda' = \lambda/k$, and $h = -k/\rho\, g$.

Let

$$\lambda' + \frac{y}{h} = -\cosh z. \tag{10.25}$$

Making this substitution, Equation (10.24) yields

$$\frac{dz}{dx} = -h^{-1}. \tag{10.26}$$

Hence,

$$z = -\frac{x}{h} + c, \tag{10.27}$$

where c is a constant. It follows from Equation (10.25) that

$$y(x) = -h\,[\lambda' + \cosh(-x/h + c)]. \tag{10.28}$$

The above solution contains three undetermined constants, h, λ', and c. We can eliminate two of these constants by application of the boundary conditions $y(\pm a/2) = 0$. This yields

$$\lambda' + \cosh(\mp a/2\,h + c) = 0. \tag{10.29}$$

Hence, $c = 0$, and $\lambda' = -\cosh(a/2\,h)$. It follows that

$$y(x) = h\,[\cosh(a/2\,h) - \cosh(x/h)]. \tag{10.30}$$

The final unknown constant, h, is determined via the application of the constraint (10.21). Thus,

$$l = \int_{-a/2}^{a/2} [1 + y'^2]^{1/2}\, dx = \int_{-a/2}^{a/2} \cosh(x/h)\, dx = 2\,h\, \sinh(a/2\,h). \tag{10.31}$$

Hence, the equilibrium configuration of the chain is given by the curve (10.30), which is known as a *catenary*, where the parameter h satisfies

$$\frac{l}{2\,h} = \sinh\left(\frac{a}{2\,h}\right). \tag{10.32}$$

10.4 Multi-Function Variation

Suppose that we wish to maximize or minimize the functional

$$I = \int_a^b F(y_1, y_2, \cdots, y_{\mathcal{F}}, y_1', y_2', \cdots, y_{\mathcal{F}}', x)\, dx. \tag{10.33}$$

Here, the integrand F is now a functional of the $\mathcal{F}$ independent functions $y_i(x)$, for $i = 1, \mathcal{F}$. A fairly straightforward extension of the analysis in Section 10.2 yields $\mathcal{F}$ separate Euler-Lagrange equations,

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y_i'}\right) - \frac{\partial F}{\partial y_i} = 0, \tag{10.34}$$

for $i = 1, \mathcal{F}$, which determine the $\mathcal{F}$ functions $y_i(x)$. If F does not explicitly depend on the function y_k then the kth Euler-Lagrange equation simplifies to

$$\frac{\partial F}{\partial y_k'} = \text{const}. \tag{10.35}$$

Likewise, if F does not explicitly depend on x then all $\mathcal{F}$ Euler-Lagrange equations simplify to

$$y_i' \frac{\partial F}{\partial y_i'} - F = \text{const}, \tag{10.36}$$

for $i = 1, \mathcal{F}$.

10.5 Hamilton's Principle

We saw, in Chapter 9, that we can specify the instantaneous configuration of a conservative dynamical system with $\mathcal{F}$ degrees of freedom in terms of $\mathcal{F}$ independent generalized coordinates q_i, for $i = 1, \mathcal{F}$. Let $K(q_1, q_2, \cdots, q_{\mathcal{F}}, \dot{q}_1, \dot{q}_2, \cdots, \dot{q}_{\mathcal{F}}, t)$ and $U(q_1, q_2, \cdots, q_{\mathcal{F}}, t)$ represent the kinetic and potential energies of the system, respectively, expressed in terms of these generalized coordinates. Here, $\dot{} \equiv d/dt$. The Lagrangian of the system is defined

$$L(q_1, q_2, \cdots, q_{\mathcal{F}}, \dot{q}_1, \dot{q}_2, \cdots, \dot{q}_{\mathcal{F}}, t) = K - U. \tag{10.37}$$

Finally, the $\mathcal{F}$ Lagrangian equations of motion of the system take the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0, \tag{10.38}$$

for $i = 1, \mathcal{F}$.

Note that the above equations of motion have exactly the same mathematical form as the Euler-Lagrange equations (10.34). Indeed, it is clear, from Section 10.4, that the $\mathcal{F}$ Lagrangian equations of motion (10.38) can all be derived from a single equation: namely,

$$\delta \int_{t_1}^{t_2} L(q_1, q_2, \cdots, q_{\mathcal{F}}, \dot{q}_1, \dot{q}_2, \cdots, \dot{q}_{\mathcal{F}}, t)\, dt = 0. \tag{10.39}$$

In other words, the motion of the system in a given time interval is such as to maximize or minimize the time integral of the Lagrangian, which is known as the *action integral*. Thus, the laws of Newtonian dynamics can be summarized in a single statement:

> The motion of a dynamical system in a given time interval is such as to maximize or minimize the action integral.

(In practice, the action integral is almost always minimized.) This statement is known as *Hamilton's principle*, and was first formulated in 1834 by the Irish mathematician William Hamilton.

10.6 Constrained Lagrangian Dynamics

Suppose that we have a dynamical system described by two generalized coordinates, q_1 and q_2. Suppose, further, that q_1 and q_2 are *not* independent variables. In other words, q_1 and q_2 are connected via some constraint equation of the form

$$f(q_1, q_2, t) = 0, \tag{10.40}$$

where f is some function of three variables. This type of constraint is called a *holonomic*. [A general holonomic constraint is of the form $f(q_1, q_2, \cdots, q_{\mathcal{F}}, t) = 0$.] Let $L(q_1, q_2, \dot{q}_1, \dot{q}_2, t)$ be the Lagrangian. How do we write the Lagrangian equations of motion of the system?

Well, according to Hamilton's principle,

$$\delta \int_{t_1}^{t_2} L\, dt = \int_{t_1}^{t_2} \left\{ \left[\frac{\partial L}{\partial q_1} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_1}\right)\right] \delta q_1 + \left[\frac{\partial L}{\partial q_2} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_2}\right)\right] \delta q_2 \right\} dt = 0. \tag{10.41}$$

However, at any given instant in time, δq_1 and δq_2 are not independent. Indeed, Equation (10.40) yields

$$\delta f = \frac{\partial f}{\partial q_1}\,\delta q_1 + \frac{\partial f}{\partial q_2}\,\delta q_2 = 0 \tag{10.42}$$

at a fixed time. Eliminating δq_2 from Equation (10.41), we obtain

$$\int_{t_1}^{t_2} \left\{ \left[\frac{\partial L}{\partial q_1} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_1}\right)\right] \frac{1}{\partial f/\partial q_1} - \left[\frac{\partial L}{\partial q_2} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_2}\right)\right] \frac{1}{\partial f/\partial q_2} \right\} \delta q_1\, dt = 0. \tag{10.43}$$

This equation must be satisfied for all possible perturbations $\delta q_1(t)$, which implies that the term enclosed in curly brackets is zero. Hence, we obtain

$$\frac{\partial L/\partial q_1 - (d/dt)\,(\partial L/\partial \dot{q}_1)}{\partial f/\partial q_1} = \frac{\partial L/\partial q_2 - (d/dt)\,(\partial L/\partial \dot{q}_2)}{\partial f/\partial q_2}. \tag{10.44}$$

One obvious way in which we can solve this equation is to separately set both sides equal to the same function of time, which we shall denote $-\lambda(t)$. It follows that the Lagrangian

equations of motion of the system can be written

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_1}\right) - \frac{\partial L}{\partial q_1} - \lambda(t)\,\frac{\partial f}{\partial q_1} = 0, \tag{10.45}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_2}\right) - \frac{\partial L}{\partial q_2} - \lambda(t)\,\frac{\partial f}{\partial q_2} = 0. \tag{10.46}$$

In principle, the above two equations can be solved, together with the constraint equation (10.40), to give $q_1(t)$, $q_2(t)$, and the so-called *Lagrange multiplier* $\lambda(t)$. Equation (10.45) can be rewritten

$$\frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_1}\right) - \frac{\partial K}{\partial q_1} = -\frac{\partial U}{\partial q_1} + \lambda(t)\,\frac{\partial f}{\partial q_1}. \tag{10.47}$$

Now, the generalized force conjugate to the generalized coordinate q_1 is [see Equation (9.8)]

$$Q_1 = -\frac{\partial U}{\partial q_1}. \tag{10.48}$$

By analogy, it seems clear from Equation (10.47) that the generalized *constraint force* [*i.e.*, the generalized force responsible for maintaining the constraint (10.40)] conjugate to q_1 takes the form

$$\tilde{Q}_1 = \lambda(t)\,\frac{\partial f}{\partial q_1}, \tag{10.49}$$

with a similar expression for the generalized constraint force conjugate to q_2.

Suppose, now, that we have a dynamical system described by $\mathcal{F}$ generalized coordinates q_i, for $i = 1, \mathcal{F}$, which is subject to the holonomic constraint

$$f(q_1, q_2, \cdots, q_{\mathcal{F}}, t) = 0. \tag{10.50}$$

A simple extension of above analysis yields following the Lagrangian equations of motion of the system,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} - \lambda(t)\,\frac{\partial f}{\partial q_i} = 0, \tag{10.51}$$

for $i = 1, \mathcal{F}$. As before,

$$\tilde{Q}_i = \lambda(t)\,\frac{\partial f}{\partial q_i} \tag{10.52}$$

is the generalized constraint force conjugate to q_i. Finally, the generalization to multiple holonomic constraints is straightforward.

Consider the following example. A cylinder of radius a rolls without slipping down a plane inclined at an angle θ to the horizontal. Let x represent the downward displacement of the center of mass of the cylinder parallel to the surface of the plane, and let ϕ represent the angle of rotation of the cylinder about its symmetry axis. The fact that the cylinder is rolling without slipping implies that x and ϕ are interrelated via the well-known constraint

$$f(x, \phi) = x - a\,\phi = 0. \tag{10.53}$$

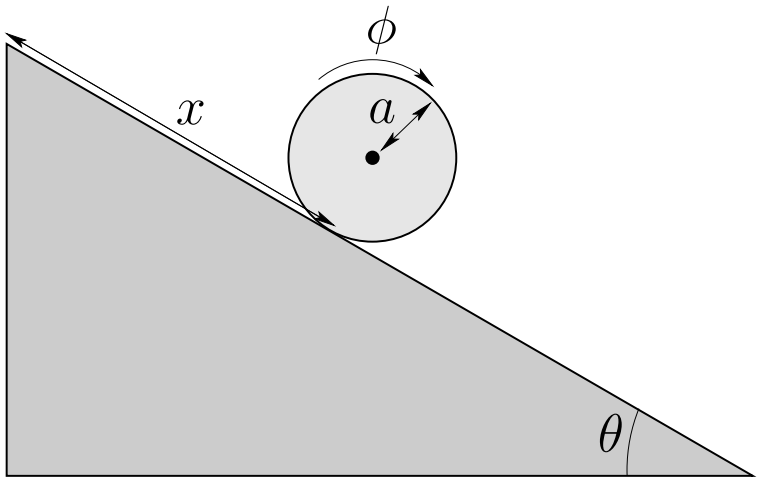


Figure 10.2: *A cylinder rolling down an inclined plane.*

The Lagrangian of the cylinder takes the form

$$L = \frac{1}{2}\, m\, \dot{x}^2 + \frac{1}{2}\, I\, \dot{\phi}^2 + m\, g\, x\, \sin\theta, \tag{10.54}$$

where m is the cylinder's mass, I its moment of inertia, and g the acceleration due to gravity.

Note that $\partial f/\partial x = 1$ and $\partial f/\partial \phi = -a$. Hence, Equation (10.51) yields the following Lagrangian equations of motion:

$$m\,\ddot{x} - m\, g\, \sin\theta - \lambda = 0, \tag{10.55}$$

$$I\,\ddot{\phi} + \lambda\, a = 0. \tag{10.56}$$

Equations (10.53), (10.55), and (10.56) can be solved to give

$$\ddot{x} = \frac{g\, \sin\theta}{1 + I/m\, a^2}, \tag{10.57}$$

$$a\,\ddot{\phi} = \frac{g\, \sin\theta}{1 + I/m\, a^2}, \tag{10.58}$$

$$\lambda = -\frac{m\, g\, \sin\theta}{1 + m\, a^2/I}. \tag{10.59}$$

The generalized constraint force conjugate to x is

$$\tilde{Q}_x = \lambda\, \frac{\partial f}{\partial x} = -\frac{m\, g\, \sin\theta}{1 + m\, a^2/I}. \tag{10.60}$$

This represents the frictional force acting parallel to the plane which impedes the downward acceleration of the cylinder, causing it to be less than the standard value $m\, g\, \sin\theta$.

The generalized constraint force conjugate to ϕ is

$$\tilde{Q}_\phi = \lambda \frac{\partial f}{\partial \phi} = \frac{m\, g\, a\, \sin\theta}{1 + m\, a^2/I}. \tag{10.61}$$

This represents the frictional torque acting on the cylinder which forces the cylinder to rotate in such a manner that the constraint (10.53) is always satisfied.

Consider a second example. A bead of mass m slides without friction on a vertical circular hoop of radius a. Let r be the radial coordinate of the bead, and let θ be its angular coordinate, with the lowest point on the hoop corresponding to $\theta = 0$. Both coordinates are measured relative to the center of the hoop. Now, the bead is constrained to slide along the wire, which implies that

$$f(r,\theta) = r - a = 0. \tag{10.62}$$

Note that $\partial f/\partial r = 1$ and $\partial f/\partial\theta = 0$. The Lagrangian of the system takes the form

$$L = \frac{1}{2}\, m\, (\dot{r}^2 + r^2\, \dot{\theta}^2) + m\, g\, r\, \cos\theta. \tag{10.63}$$

Hence, according to Equation (10.51), the Lagrangian equations of motion of the system are written

$$m\,\ddot{r} - m\, r\, \dot{\theta}^2 - m\, g\, \cos\theta - \lambda = 0, \tag{10.64}$$

$$m\, r^2\, \ddot{\theta} + m\, g\, r\, \sin\theta = 0. \tag{10.65}$$

The second of these equations can be integrated (by multiplying by $\dot{\theta}$), subject to the constraint (10.62), to give

$$\dot{\theta}^2 = \frac{2\, g}{a}\, \cos\theta + c, \tag{10.66}$$

where c is a constant. Let v_0 be the tangential velocity of the bead at the bottom of the hoop (*i.e.*, at $\theta = 0$). It follows that

$$\dot{\theta}^2 = \frac{2\, g}{a}\, (\cos\theta - 1) + \frac{v_0^2}{a^2}. \tag{10.67}$$

Equations (10.62), (10.64), and (10.67) can be combined to give

$$\lambda = -m \left(3\, g\, \cos\theta - 2\, g + \frac{v_0^2}{a} \right). \tag{10.68}$$

Finally, the constraint force conjugate to r is given by

$$\tilde{Q}_r = \lambda \frac{\partial f}{\partial r} = -m \left(3\, g\, \cos\theta - 2\, g + \frac{v_0^2}{a} \right). \tag{10.69}$$

This represents the radial reaction exerted on the bead by the hoop. Of course, there is no constraint force conjugate to θ (since $\partial f/\partial\theta = 0$) because the bead slides without friction.

10.7 Hamilton's Equations

Consider a dynamical system with $\mathcal{F}$ degrees of freedom which is described by the generalized coordinates q_i, for $i = 1, \mathcal{F}$. Suppose that neither the kinetic energy, K, nor the potential energy, U, depend explicitly on the time, t. Now, in conventional dynamical systems, the potential energy is generally independent of the $\dot{q}_i$, whereas the kinetic energy takes the form of a *homogeneous quadratic function* of the $\dot{q}_i$. In other words,

$$K = \sum_{i,j=1,\mathcal{F}} m_{ij}\, \dot{q}_i\, \dot{q}_j, \tag{10.70}$$

where the m_{ij} depend on the q_i, but not on the $\dot{q}_i$. It is easily demonstrated from the above equation that

$$\sum_{i=1,\mathcal{F}} \dot{q}_i \frac{\partial K}{\partial \dot{q}_i} = 2\, K. \tag{10.71}$$

Recall, from Section 9.8, that generalized momentum conjugate to the ith generalized coordinate is defined

$$p_i = \frac{\partial L}{\partial \dot{q}_i} = \frac{\partial K}{\partial \dot{q}_i}, \tag{10.72}$$

where $L = K - U$ is the Lagrangian of the system, and we have made use of the fact that U is independent of the $\dot{q}_i$. Consider the function

$$H = \sum_{i=1,\mathcal{F}} \dot{q}_i\, p_i - L = \sum_{i=1,\mathcal{F}} \dot{q}_i\, p_i - K + U. \tag{10.73}$$

If all of the conditions discussed above are satisfied then Equations (10.71) and (10.72) yield

$$H = K + U. \tag{10.74}$$

In other words, the function H is equal to the *total energy* of the system.

Consider the variation of the function H. We have

$$\delta H = \sum_{i=1,\mathcal{F}} \left(\delta \dot{q}_i\, p_i + \dot{q}_i\, \delta p_i - \frac{\partial L}{\partial \dot{q}_i}\, \delta \dot{q}_i - \frac{\partial L}{\partial q_i}\, \delta q_i \right). \tag{10.75}$$

The first and third terms in the bracket cancel, because $p_i = \partial L/\partial \dot{q}_i$. Furthermore, since Lagrange's equation can be written $\dot{p}_i = \partial L/\partial q_i$ (see Section 9.8), we obtain

$$\delta H = \sum_{i=1,\mathcal{F}} (\dot{q}_i\, \delta p_i - \dot{p}_i\, \delta q_i). \tag{10.76}$$

Suppose, now, that we can express the total energy of the system, H, solely as a function of the q_i and the p_i, with no explicit dependence on the $\dot{q}_i$. In other words, suppose that we

can write $H = H(q_i, p_i)$. When the energy is written in this fashion it is generally termed the *Hamiltonian* of the system. The variation of the Hamiltonian function takes the form

$$\delta H = \sum_{i=1,\mathcal{F}} \left(\frac{\partial H}{\partial p_i}\,\delta p_i + \frac{\partial H}{\partial q_i}\,\delta q_i \right). \tag{10.77}$$

A comparison of the previous two equations yields

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \tag{10.78}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}, \tag{10.79}$$

for $i = 1, \mathcal{F}$. These $2\mathcal{F}$ first-order differential equations are known as *Hamilton's equations*. Hamilton's equations are often a useful alternative to Lagrange's equations, which take the form of $\mathcal{F}$ second-order differential equations.

Consider a one-dimensional harmonic oscillator. The kinetic and potential energies of the system are written $K = (1/2)\,m\,\dot{x}^2$ and $U = (1/2)\,k\,x^2$, where x is the displacement, m the mass, and $k > 0$. The generalized momentum conjugate to x is

$$p = \frac{\partial K}{\partial \dot{x}} = m\,\dot{x}. \tag{10.80}$$

Hence, we can write

$$K = \frac{1}{2}\,\frac{p^2}{m}. \tag{10.81}$$

So, the Hamiltonian of the system takes the form

$$H = K + U = \frac{1}{2}\,\frac{p^2}{m} + \frac{1}{2}\,k\,x^2. \tag{10.82}$$

Thus, Hamilton's equations, (10.78) and (10.79), yield

$$\dot{x} = \frac{\partial H}{\partial p} = \frac{p}{m}, \tag{10.83}$$

$$\dot{p} = -\frac{\partial H}{\partial x} = -k\,x. \tag{10.84}$$

Of course, the first equation is just a restatement of Equation (10.80), whereas the second is Newton's second law of motion for the system.

Consider a particle of mass m moving in the central potential $U(r)$. In this case,

$$K = \frac{1}{2}\,m\,(\dot{r}^2 + r^2\,\dot{\theta}^2), \tag{10.85}$$

where r, θ are polar coordinates. The generalized momenta conjugate to r and θ are

$$p_r = \frac{\partial K}{\partial \dot{r}} = m\,\dot{r}, \tag{10.86}$$

$$p_\theta = \frac{\partial K}{\partial \dot{\theta}} = m\,r^2\,\dot{\theta}, \tag{10.87}$$

respectively. Hence, we can write

$$K = \frac{1}{2\,m}\left(p_r^2 + \frac{p_\theta^2}{r^2}\right). \tag{10.88}$$

Thus, the Hamiltonian of the system takes the form

$$H = \frac{1}{2\,m}\left(p_r^2 + \frac{p_\theta^2}{r^2}\right) + U(r). \tag{10.89}$$

In this case, Hamilton's equations yield

$$\dot{r} = \frac{\partial H}{\partial p_r} = \frac{p_r}{m}, \tag{10.90}$$

$$\dot{\theta} = \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{m\,r^2}, \tag{10.91}$$

which are just restatements of Equations (10.86) and (10.87), respectively, as well as

$$\dot{p}_r = -\frac{\partial H}{\partial r} = \frac{p_\theta^2}{m\,r^3} - \frac{\partial U}{\partial r}, \tag{10.92}$$

$$\dot{p}_\theta = -\frac{\partial H}{\partial \theta} = 0. \tag{10.93}$$

The last equation implies that

$$\frac{p_\theta}{m} = r^2\,\dot{\theta} = h, \tag{10.94}$$

where h is a constant. This can be combined with Equation (10.92) to give

$$\frac{\dot{p}_r}{m} = \ddot{r} = \frac{h^2}{r^3} - \frac{\partial V}{\partial r}, \tag{10.95}$$

where $V = U/m$. Of course, Equations (10.94) and (10.95) are the conventional equations of motion for a particle moving in a central potential—see Chapter 5.

10.8 Exercises

10.1. A particle of mass m is placed at the top of a smooth vertical hoop of radius a. Calculate the reaction of the hoop on the particle as it slides down the hoop by means of the method of Lagrange multipliers. Find the height at which the particle falls off the hoop.

10.2. A uniform disk of mass m and radius a has a light string wrapped around its circumference with one end of the string attached to a fixed support. The disk is allowed to fall under gravity, unwinding the string as it falls. Solve the problem using the method of Lagrange multipliers. What is the tension in the string?

10.3. Consider two particles of masses m_1 and m_2. Let m_1 be constrained to move on a circle of radius a in the $z = 0$ plane, centered at $x = y = 0$. (Here, z measures vertical height). Let m_2 be constrained to move on a circle of radius a in the $z = c$ plane, centered on $x = y = 0$. A light spring of spring constant k and unstreatched length c is attached between the particles. Find the Lagrangian of the system. Solve the problem using Lagrange multipliers and give a physical interpretation for each multiplier.

10.4. Find the Hamiltonian of a particle of mass m constrained to move under gravity on the inside of a sphere of radius a. Use the standard spherical polar coordinates θ and ϕ as your generalized coordinates, where the axis of the coordinates points vertically downward. Find Hamilton's equations of motion for the system.

10.5. A particle of mass m is subject to a central attractive force given by

$$\mathbf{f} = -\frac{k\, e^{-\beta t}}{r^2}\, \mathbf{e}_r,$$

where k and β are positive constants. Find the Hamiltonian of the particle. Compare the Hamiltonian to the total energy of the particle. Is the energy of the particle conserved?

11 Coupled Oscillations

11.1 Introduction

This chapter examines the motion of a many degree of freedom dynamical system which is perturbed from some equilibrium state.

11.2 Equilibrium State

Consider an $\mathcal{F}$ degree of freedom dynamical system described by the generalized coordinates q_i, for $i = 1, \mathcal{F}$. Suppose that the kinetic energy K and the potential energy U are not explicit functions of time. This implies that the system in question is *isolated*: *i.e.*, it is not subject to any external forces or time-varying constraints. In virtually all dynamical systems of interest, the kinetic energy can be expressed as a *quadratic form*: *i.e.*,

$$K = \frac{1}{2} \sum_{i,j=1,\mathcal{F}} m_{ij}(q_1, q_2, \cdots, q_{\mathcal{F}})\, \dot{q}_i\, \dot{q}_j. \tag{11.1}$$

Without loss of generality, we can specify that the weights m_{ij} in the above form are invariant under interchange of the indices i and j: *i.e.*,

$$m_{ij} = m_{ji}. \tag{11.2}$$

Finally, the potential energy is written $U = U(q_1, q_2, \cdots, q_{\mathcal{F}})$.

Suppose that $q_i = q_{i\,0}$, for $i = 1, \mathcal{F}$, corresponds to an *equilibrium state* of the system. It follows that $q_i = q_{i\,0}$ and $\dot{q}_i = \ddot{q}_i = 0$, for $i = 1, \mathcal{F}$, should be a possible solution of the equations of motion. Now, Lagrange's equations of motion for the system take the form [see Equation (9.22)]

$$\frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_i}\right) - \frac{\partial K}{\partial q_i} + \frac{\partial U}{\partial q_i} = 0, \tag{11.3}$$

for $i = 1, \mathcal{F}$. Here, we have made use of the definition $L = K - U$, and the fact that U is independent of the $\dot{q}_i$. Now, it is clear, from an examination of Equation (11.1), that every component making up the first two terms in the above equation contains either a $\dot{q}_j$ or a $\ddot{q}_j$, for some j. But, we can set all of the generalized velocities and accelerations to zero in an equilibrium state of the system. Hence, the first two terms in the above equation are zero, and the condition for equilibrium reduces to

$$Q_{i\,0} = -\frac{\partial U(q_{1\,0}, q_{2\,0}, \cdots, q_{\mathcal{F}\,0})}{\partial q_i} = 0, \tag{11.4}$$

for $i = 1, \mathcal{F}$. In other words, $q_i = q_{i\,0}$, for $i = 1, \mathcal{F}$, is an equilibrium state provided that all of the generalized forces, Q_i [see Equation (9.8)], evaluated at $q_i = q_{i\,0}$, are zero. Let us suppose that this is the case.

11.3 Stability Equations

It is evident that if our system is initialized in some equilibrium state, with all of the $\dot{q}_i$ set to zero, then it will remain in this state for ever. But what happens if the system is slightly perturbed from the equilibrium state?

Let

$$q_i = q_{i\,0} + \delta q_i, \tag{11.5}$$

for $i = 1, \mathcal{F}$, where the δq_i are *small*. To lowest order in δq_i, the kinetic energy (11.1) can be written

$$K \simeq \frac{1}{2} \sum_{i,j=1,\mathcal{F}} M_{ij}\, \delta\dot{q}_i\, \delta\dot{q}_j, \tag{11.6}$$

where

$$M_{ij} = m_{ij}(q_{1\,0}, q_{2\,0}, \cdots, q_{\mathcal{F}\,0}), \tag{11.7}$$

and

$$M_{ij} = M_{ji}. \tag{11.8}$$

Note that the weights M_{ij} in the quadratic form (11.6) are now *constants*.

Taylor expanding the potential energy function about the equilibrium state, up to second-order in the δq_i, we obtain

$$U \simeq U_0 - \sum_{i=1,\mathcal{F}} Q_{i\,0}\, \delta q_i - \frac{1}{2} \sum_{i,j=1,\mathcal{F}} G_{ij}\, \delta q_i\, \delta q_j, \tag{11.9}$$

where $U_0 = U(q_{1\,0}, q_{2\,0}, \cdots, q_{\mathcal{F}\,0})$, the $Q_{i\,0}$ are specified in Equation (11.4), and

$$G_{ij} = -\frac{\partial^2 U(q_{1\,0}, q_{2\,0}, \cdots, q_{\mathcal{F}\,0})}{\partial q_i\, \partial q_j}. \tag{11.10}$$

Now, we can set U_0 to zero without loss of generality. Moreover, according to Equation (11.4), the $Q_{i\,0}$ are all zero. Hence, the expression (11.9) reduces to

$$U \simeq -\frac{1}{2} \sum_{i,j=1,\mathcal{F}} G_{ij}\, \delta q_i\, \delta q_j. \tag{11.11}$$

Note that, since $\partial^2 U/\partial q_i\, \partial q_j \equiv \partial^2 U/\partial q_j\, \partial q_i$, the constants weights G_{ij} in the above quadratic form are invariant under interchange of the indices i and j: *i.e.*,

$$G_{ij} = G_{ji}. \tag{11.12}$$

With K and U specified by the quadratic forms (11.6) and (11.11), respectively, Lagrange's equations of motion (11.3) reduce to

$$\sum_{j=1,\mathcal{F}} (M_{ij}\, \delta\ddot{q}_j - G_{ij}\, \delta q_j) = 0, \tag{11.13}$$

for $i = 1, \mathcal{F}$. Note that the above coupled differential equations are *linear* in the δq_i. It follows that the solutions are *superposable*. Let us search for solutions of the above equations in which all of the perturbed coordinates δq_i have a common time variation of the form

$$\delta q_i(t) = \delta q_i \, e^{\gamma t}, \tag{11.14}$$

for $i = 1, \mathcal{F}$. Now, Equations (11.13) are a set of $\mathcal{F}$ second-order differential equations. Hence, the most general solution contains $2\mathcal{F}$ arbitrary constants of integration. Thus, if we can find sufficient independent solutions of the form (11.14) to Equations (11.13) that the superposition of these solutions contains $2\mathcal{F}$ arbitrary constants then we can be sure that we have found the most general solution. Equations (11.13) and (11.14) yield

$$\sum_{j=1,\mathcal{F}} (G_{ij} - \gamma^2 M_{ij}) \, \delta q_j = 0, \tag{11.15}$$

which can be written more succinctly as a matrix equation:

$$(\mathbf{G} - \gamma^2 \mathbf{M}) \, \delta \mathbf{q} = \mathbf{0}. \tag{11.16}$$

Here, $\mathbf{G}$ is the real [see Equation (11.10)] symmetric [see Equation (11.12)] $\mathcal{F} \times \mathcal{F}$ matrix of the G_{ij} values. Furthermore, $\mathbf{M}$ is the real [see Equation (11.1)] symmetric [see Equation (11.8)] $\mathcal{F} \times \mathcal{F}$ matrix of the M_{ij} values. Finally, $\delta \mathbf{q}$ is the $1 \times \mathcal{F}$ vector of the δq_i values, and $\mathbf{0}$ is a null vector.

11.4 More Matrix Eigenvalue Theory

Equation (11.16) takes the form of a matrix eigenvalue equation:

$$(\mathbf{G} - \lambda \mathbf{M}) \, \mathbf{x} = \mathbf{0}. \tag{11.17}$$

Here, $\mathbf{G}$ and $\mathbf{M}$ are both real symmetric matrices, whereas λ is termed the *eigenvalue*, and $\mathbf{x}$ the associated *eigenvector*. The above matrix equation is essentially a set of $\mathcal{F}$ homogeneous simultaneous algebraic equations for the components of $\mathbf{x}$. As is well-known, a necessary condition for such a set of equations to possess a non-trivial solution is that the *determinant* of the matrix must be zero: *i.e.*,

$$|\mathbf{G} - \lambda \mathbf{M}| = 0. \tag{11.18}$$

The above formula reduces to an $\mathcal{F}$th-order *polynomial* equation for λ. Hence, we conclude that Equation (11.17) is satisfied by $\mathcal{F}$ eigenvalues, and $\mathcal{F}$ associated eigenvectors.

We can easily demonstrate that the eigenvalues are all *real*. Suppose that λ_k and $\mathbf{x}_k$ are the kth eigenvalue and eigenvector, respectively. Then we have

$$\mathbf{G} \, \mathbf{x}_k = \lambda_k \, \mathbf{M} \, \mathbf{x}_k. \tag{11.19}$$

Taking the transpose and complex conjugate of the above equation, and right multiplying by $\mathbf{x}_k$, we obtain

$$\mathbf{x}_k^{*\,\mathrm{T}}\,\mathbf{G}^{*\,\mathrm{T}}\,\mathbf{x}_k = \lambda_k^*\,\mathbf{x}_k^{*\,\mathrm{T}}\,\mathbf{M}^{*\,\mathrm{T}}\,\mathbf{x}_k. \tag{11.20}$$

Here, $^{\mathrm{T}}$ denotes a transpose, and * a complex conjugate. However, since $\mathbf{G}$ and $\mathbf{M}$ are both real symmetric matrices, it follows that $\mathbf{G}^{*\,\mathrm{T}} = \mathbf{G}$ and $\mathbf{M}^{*\,\mathrm{T}} = \mathbf{M}$. Hence,

$$\mathbf{x}_k^{*\,\mathrm{T}}\,\mathbf{G}\,\mathbf{x}_k = \lambda_k^*\,\mathbf{x}_k^{*\,\mathrm{T}}\,\mathbf{M}\,\mathbf{x}_k. \tag{11.21}$$

Next, left multiplying Equation (11.19) by $\mathbf{x}_k^{*\,\mathrm{T}}$, we obtain

$$\mathbf{x}_k^{*\,\mathrm{T}}\,\mathbf{G}\,\mathbf{x}_k = \lambda_k\,\mathbf{x}_k^{*\,\mathrm{T}}\,\mathbf{M}\,\mathbf{x}_k. \tag{11.22}$$

Taking the difference between the above two expressions, we get

$$(\lambda_k^* - \lambda_k)\,\mathbf{x}_k^{*\,\mathrm{T}}\,\mathbf{M}\,\mathbf{x}_k = 0. \tag{11.23}$$

Since $\mathbf{x}_k^{*\,\mathrm{T}}\,\mathbf{M}\,\mathbf{x}_k$ is not generally zero, except in the trivial case where $\mathbf{x}_k$ is a null vector, we conclude that $\lambda_k^* = \lambda_k$ for all k. In other words, the eigenvalues are all real. It immediately follows that the eigenvectors can also be chosen to be all real.

Consider two distinct eigenvalues, λ_k and λ_l, with the associated eigenvectors $\mathbf{x}_k$ and $\mathbf{x}_l$, respectively. We have

$$\mathbf{G}\,\mathbf{x}_k = \lambda_k\,\mathbf{M}\,\mathbf{x}_k, \tag{11.24}$$

$$\mathbf{G}\,\mathbf{x}_l = \lambda_l\,\mathbf{M}\,\mathbf{x}_l. \tag{11.25}$$

Right multiplying the transpose of Equation (11.24) by $\mathbf{x}_l$, and left multiplying Equation (11.25) by $\mathbf{x}_k^{\mathrm{T}}$, we obtain

$$\mathbf{x}_k^{\mathrm{T}}\,\mathbf{G}\,\mathbf{x}_l = \lambda_k\,\mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\mathbf{x}_l, \tag{11.26}$$

$$\mathbf{x}_k^{\mathrm{T}}\,\mathbf{G}\,\mathbf{x}_l = \lambda_l\,\mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\mathbf{x}_l. \tag{11.27}$$

Taking the difference between the above two expressions, we get

$$(\lambda_k - \lambda_l)\,\mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\mathbf{x}_l = 0. \tag{11.28}$$

Hence, we conclude that

$$\mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\mathbf{x}_l = 0, \tag{11.29}$$

provided $\lambda_k \neq \lambda_l$. In other words, two eigenvectors corresponding to two different eigenvalues are "orthogonal" to one another (in the sense specified in the above equation). Moreover, it is easily demonstrated that different eigenvectors corresponding to the same eigenvalue can be chosen in such a manner that they are also orthogonal to one another—see Section 8.5. Thus, we conclude that all of the eigenvectors are *mutually orthogonal*. Since Equation (11.17) only specifies the directions, and not the lengths, of the eigenvectors, we are free to normalize our eigenvectors such that

$$\mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\mathbf{x}_l = \delta_{kl}, \tag{11.30}$$

where $\delta_{kl} = 1$ when $k = l$, and $\delta_{kl} = 0$ otherwise. Note, finally, that since the $\mathbf{x}_k$, for $k = 1, \mathcal{F}$, are mutually orthogonal, they are *independent* (*i.e.*, one eigenvector cannot be expressed as a linear combination of the others), and completely span $\mathcal{F}$-dimensional vector space.

11.5 Normal Modes

It follows from Equation (11.14) and (11.15), plus the mathematical results contained in the previous section, that the most general solution to Equation (11.13) can be written

$$\delta\mathbf{q}(t) = \sum_{k=1,\mathcal{F}} \delta\mathbf{q}_k(t), \tag{11.31}$$

where

$$\delta\mathbf{q}_k(t) = \left(\alpha_k\, e^{+\sqrt{\lambda_k}\, t} + \beta_k\, e^{-\sqrt{\lambda_k}\, t}\right) \mathbf{x}_k. \tag{11.32}$$

Here, the λ_k and the $\mathbf{x}_k$ are the eigenvalues and eigenvectors obtained by solving Equation (11.17). Moreover, the α_k and β_k are arbitrary constants. Finally, we have made use of the fact that the two roots of $\gamma^2 = \lambda_k$ are $\gamma = \pm\sqrt{\lambda_k}$.

According to Equation (11.31), the most general perturbed motion of the system consists of a *linear combination* of $\mathcal{F}$ different modes. These modes are generally termed *normal modes*, since they are mutually orthogonal (because the $\mathbf{x}_k$ are mutually orthogonal). Furthermore, it follows from the independence of the $\mathbf{x}_k$ that the normal modes are also *independent* (*i.e.*, one mode cannot be expressed as a linear combination of the others). The kth normal mode has a specific pattern of motion which is specified by the kth eigenvector, $\mathbf{x}_k$. Moreover, the kth mode has a specific time variation which is determined by the associated eigenvalue, λ_k. Recall that λ_k is *real*. Hence, there are only two possibilities. Either λ_k is *positive*, in which case we can write

$$\delta\mathbf{q}_k(t) = \left(\alpha_k\, e^{+\gamma_k\, t} + \beta_k\, e^{-\gamma_k\, t}\right) \mathbf{x}_k, \tag{11.33}$$

where $\lambda_k = \gamma_k^2$, or λ_k is *negative*, in which case we can write

$$\delta\mathbf{q}_k(t) = \left(\alpha_k\, e^{+i\,\omega_k\, t} + \beta_k\, e^{-i\,\omega_k\, t}\right) \mathbf{x}_k, \tag{11.34}$$

where $\lambda_k = -\omega_k^2$. In other words, if λ_k is positive then the kth normal mode *grows* or decays secularly in time, whereas if λ_k is negative then the kth normal mode *oscillates* in time. Obviously, if the system possesses one or more normal modes which grow secularly in time then the equilibrium about which we originally expanded the equations of motion must be an *unstable* equilibrium. On the other hand, if all of the normal modes oscillate in time then the equilibrium is *stable*. Thus, we conclude that whilst Equation (11.4) is the condition for the existence of an equilibrium state in a many degree of freedom system, the condition for the equilibrium to be *stable* is that all of the eigenvalues of the stability equation (11.17) must be *negative*.

The arbitrary constants α_k and β_k appearing in expression (11.32) are determined from the *initial conditions*. Thus, if $\delta\mathbf{q}^{(0)} = \delta\mathbf{q}(t=0)$ and $\delta\dot{\mathbf{q}}^{(0)} = \delta\dot{\mathbf{q}}(t=0)$ then it is easily demonstrated from Equations (11.30)–(11.32) that

$$\mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\delta\mathbf{q}^{(0)} = \alpha_k + \beta_k, \tag{11.35}$$

and

$$\mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\delta\dot{\mathbf{q}}^{(0)} = \sqrt{\lambda_k}\,(\alpha_k - \beta_k). \tag{11.36}$$

Hence,

$$\alpha_k = \frac{\mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\delta\mathbf{q}^{(0)} + \mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\delta\dot{\mathbf{q}}^{(0)}/\sqrt{\lambda_k}}{2}, \tag{11.37}$$

$$\beta_k = \frac{\mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\delta\mathbf{q}^{(0)} - \mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\delta\dot{\mathbf{q}}^{(0)}/\sqrt{\lambda_k}}{2}. \tag{11.38}$$

Note, finally, that since there are $2\mathcal{F}$ arbitrary constants (two for each of the $\mathcal{F}$ normal modes), we can be sure that Equation (11.31) represents the most general solution to Equation (11.13).

11.6 Normal Coordinates

Since the eigenvectors $\mathbf{x}_k$, for $k = 1, \mathcal{F}$, span $\mathcal{F}$-dimensional vector space, we can always write the displacement vector $\delta\mathbf{q}$ as some linear combination of the $\mathbf{x}_k$: *i.e.*,

$$\delta\mathbf{q}(t) = \sum_{k=1,\mathcal{F}} \eta_k(t)\,\mathbf{x}_k. \tag{11.39}$$

We can regard the $\eta_k(t)$ as a new set of generalized coordinates, since specifying the η_k is equivalent to specifying the δq_k (and, hence, the q_k). The η_k are usually termed *normal coordinates*. According to Equations (11.30) and (11.39), the normal coordinates can be written in terms of the δq_k as

$$\eta_k = \mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\delta\mathbf{q}. \tag{11.40}$$

Let us now try to express K, U, and the equations of motion in terms of the η_k.

The kinetic energy can be written

$$K = \frac{1}{2}\,\delta\dot{\mathbf{q}}^{\mathrm{T}}\,\mathbf{M}\,\delta\dot{\mathbf{q}}, \tag{11.41}$$

where use has been made of Equation (11.6). It follows from (11.39) that

$$K = \frac{1}{2}\sum_{k,l=1,\mathcal{F}} \dot{\eta}_k\,\dot{\eta}_l\,\mathbf{x}_k^{\mathrm{T}}\,\mathbf{M}\,\mathbf{x}_l. \tag{11.42}$$

Finally, making use of the orthonormality condition (11.30), we obtain

$$K = \frac{1}{2} \sum_{k=1,\mathcal{F}} \dot{\eta}_k^2. \tag{11.43}$$

Hence, the kinetic energy K takes the form of a *diagonal* quadratic form when expressed in terms of the normal coordinates.

The potential energy can be written

$$U = -\frac{1}{2}\, \delta\mathbf{q}^{\mathrm{T}}\, \mathbf{G}\, \delta\mathbf{q}, \tag{11.44}$$

where use has been made of Equations (11.11). It follows from (11.39) that

$$U = -\frac{1}{2} \sum_{k,l=1,\mathcal{F}} \eta_k\, \eta_l\, \mathbf{x}_k^{\mathrm{T}}\, \mathbf{G}\, \mathbf{x}_l. \tag{11.45}$$

Finally, making use of Equation (11.19) and the orthonormality condition (11.30), we obtain

$$U = -\frac{1}{2} \sum_{k=1,\mathcal{F}} \lambda_k\, \eta_k^2. \tag{11.46}$$

Hence, the potential energy U also takes the form of a *diagonal* quadratic form when expressed in terms of the normal coordinates.

Writing Lagrange's equations of motion in terms of the normal coordinates, we obtain [*cf.*, Equation (11.3)]

$$\frac{d}{dt}\left(\frac{\partial K}{\partial \dot{\eta}_k}\right) - \frac{\partial K}{\partial \eta_k} + \frac{\partial U}{\partial \eta_k} = 0, \tag{11.47}$$

for $k = 1, \mathcal{F}$. Thus, it follows from Equations (11.43) and (11.46) that

$$\ddot{\eta}_k = \lambda_k\, \eta_k, \tag{11.48}$$

for $k = 1, \mathcal{F}$. In other words, Lagrange's equations reduce to a set of $\mathcal{F}$ *uncoupled* simple harmonic equations when expressed in terms of the normal coordinates. The solutions to the above equations are obvious: *i.e.*,

$$\eta_k(t) = \alpha_k\, e^{+\sqrt{\lambda_k}\, t} + \beta_k\, e^{-\sqrt{\lambda_k}\, t}, \tag{11.49}$$

where α_k and β_k are arbitrary constants. Hence, it is clear from Equations (11.39) and (11.49) that the most general solution to the perturbed equations of motion is indeed given by Equations (11.31) and (11.32).

In conclusion, the equations of motion of a many degree of freedom dynamical system which is slightly perturbed from an equilibrium state take a particularly simple form when expressed in terms of the normal coordinates. Each normal coordinate specifies the instantaneous displacement of an independent mode of oscillation (or secular growth) of the system. Moreover, each normal coordinate oscillates at a characteristic frequency (or grows at a characteristic rate), and is completely unaffected by the other coordinates.

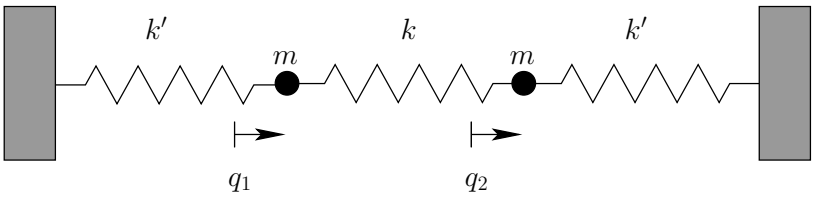


Figure 11.1: *Two spring-coupled masses.*

11.7 Spring-Coupled Masses

Consider the two degree of freedom dynamical system pictured in Figure 11.1. In this system, two point objects of mass m are free to move in one dimension. Furthermore, the masses are connected together by a spring of spring constant k, and are also each attached to fixed supports via springs of spring constant k'.

Let q_1 and q_2 be the displacements of the first and second masses, respectively, from the equilibrium state. It follows that the extensions of the left-hand, middle, and right-hand springs are q_1, $q_2 - q_1$, and $-q_2$, respectively. The kinetic energy of the system takes the form

$$K = \frac{m}{2}\,(\dot{q}_1^{\,2} + \dot{q}_2^{\,2}), \tag{11.50}$$

whereas the potential energy is written

$$U = \frac{1}{2}\left[k'\,q_1^2 + k\,(q_2 - q_1)^2 + k'\,q_2^2\right]. \tag{11.51}$$

The above expression can be rearranged to give

$$U = \frac{1}{2}\left[(k + k')\,q_1^2 - 2\,k\,q_1\,q_2 + (k + k')\,q_2^2\right]. \tag{11.52}$$

A comparison of Equations (11.50) and (11.52) with the standard forms (11.6) and (11.11) yields the following expressions for the mass matrix, **M**, and the force matrix, **G**:

$$\mathbf{M} = \begin{pmatrix} m & 0 \\ 0 & m \end{pmatrix}, \tag{11.53}$$

$$\mathbf{G} = \begin{pmatrix} -k - k' & k \\ k & -k - k' \end{pmatrix}. \tag{11.54}$$

Now, the equation of motion of the system takes the form [see Equation (11.17)]

$$(\mathbf{G} - \lambda\,\mathbf{M})\,\mathbf{x} = \mathbf{0}, \tag{11.55}$$

where $\mathbf{x}$ is the column vector of the q_1 and q_2 values. The solubility condition for the above equation is

$$|\mathbf{G} - \lambda\,\mathbf{M}| = 0, \tag{11.56}$$

or

$$\begin{vmatrix} -k-k'-\lambda m & k \\ k & -k-k'-\lambda m \end{vmatrix} = 0, \tag{11.57}$$

which yields the following quadratic equation for the eigenvalue λ:

$$m^2\lambda^2 + 2m(k+k')\lambda + k'(k'+2k) = 0. \tag{11.58}$$

The two roots of the above equation are

$$\lambda_1 = -\frac{k'}{m}, \tag{11.59}$$

$$\lambda_2 = -\frac{(2k+k')}{m}. \tag{11.60}$$

The fact that the roots are negative implies that both normal modes are *oscillatory* in nature: *i.e.*, the original equilibrium is *stable*. The characteristic oscillation frequencies of the modes are

$$\omega_1 = \sqrt{-\lambda_1} = \sqrt{\frac{k'}{m}}, \tag{11.61}$$

$$\omega_2 = \sqrt{-\lambda_2} = \sqrt{\frac{2k+k'}{m}}. \tag{11.62}$$

Now, the first row of Equation (11.55) gives

$$\frac{q_1}{q_2} = \frac{k}{k+k'+\lambda m}. \tag{11.63}$$

Moreover, Equations (11.30) and (11.53) yield the following normalization condition for the eigenvectors:

$$\mathbf{x}_k^{\mathrm{T}}\,\mathbf{x}_k = m^{-1}, \tag{11.64}$$

for $k = 1, 2$. It follows that the two eigenvectors are

$$\mathbf{x}_1 = (2m)^{-1/2}(1,\ \ 1), \tag{11.65}$$

$$\mathbf{x}_2 = (2m)^{-1/2}(1,-1). \tag{11.66}$$

According to Equations (11.61)–(11.62) and (11.65)–(11.66), our two degree of freedom system possesses two normal modes. The first mode oscillates at the frequency ω_1, and is a purely *symmetric* mode: *i.e.*, $q_1 = q_2$. Note that such a mode does not stretch the middle spring. Hence, ω_1 is independent of k. In fact, ω_1 is simply the characteristic oscillation frequency of a mass m on the end of a spring of spring constant k'. The second mode oscillates at the frequency ω_2, and is a purely *anti-symmetric* mode: *i.e.*, $q_1 = -q_2$. Since such a mode stretches the middle spring, the second mode experiences a greater restoring force than the first, and hence has a higher oscillation frequency: *i.e.*, $\omega_2 > \omega_1$.

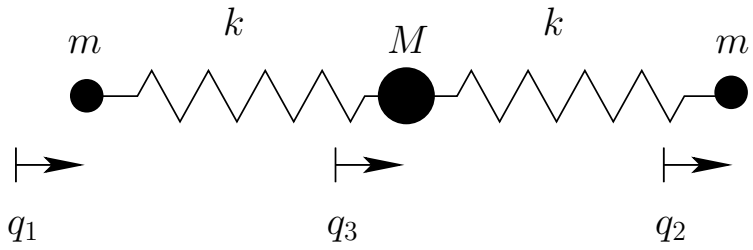


Figure 11.2: *A model triatomic molecule.*

Note, finally, from Equations (11.40) and (11.53), that the normal coordinates of the system are:

$$\eta_1 = \sqrt{\frac{m}{2}}\,(q_1+q_2), \tag{11.67}$$

$$\eta_2 = \sqrt{\frac{m}{2}}\,(q_1-q_2). \tag{11.68}$$

When expressed in terms of these normal coordinates, the kinetic and potential energies of the system reduce to

$$K = \frac{1}{2}\,(\dot\eta_1^2+\dot\eta_2^2), \tag{11.69}$$

$$U = \frac{1}{2}\,(\omega_1^2\,\eta_1^2+\omega_2^2\,\eta_2^2), \tag{11.70}$$

respectively.

11.8 Triatomic Molecule

Consider the simple model of a linear triatomic molecule (*e.g.*, carbon dioxide) illustrated in Figure 11.2. The molecule consists of a central atom of mass M flanked by two identical atoms of mass m. The atomic bonds are represented as springs of spring constant k. The linear displacements of the flanking atoms are q_1 and q_2, whilst that of the central atom is q_3. Let us investigate the linear modes of oscillation our model molecule.

The kinetic energy of the molecule is written

$$K = \frac{m}{2}\,(\dot q_1^2+\dot q_2^2)+\frac{M}{2}\,\dot q_3^2, \tag{11.71}$$

whereas the potential energy takes the form

$$U = \frac{k}{2}\,(q_3-q_1)^2+\frac{k}{2}\,(q_2-q_3)^2. \tag{11.72}$$

Clearly, we have a three degree of freedom dynamical system. However, we can reduce this to a two degree of freedom system by only considering *oscillatory* modes of motion,

and, hence, neglecting *translational* modes. We can achieve this by demanding that the center of mass of the system remains stationary. In other words, we require that

$$m\,(q_1+q_2)+M\,q_3=0. \tag{11.73}$$

This constraint can be rearranged to give

$$q_3=-\frac{m}{M}\,(q_1+q_2). \tag{11.74}$$

Eliminating q_3 from Equations (11.71) and (11.72), we obtain

$$K=\frac{m}{2}\left[(1+\alpha)\,\dot{q}_1^{\,2}+2\,\alpha\,\dot{q}_1\,\dot{q}_2+(1+\alpha)\,\dot{q}_2^{\,2}\right], \tag{11.75}$$

and

$$U=\frac{k}{2}\left[(1+2\,\alpha+2\,\alpha^2)\,q_1^2+4\,\alpha\,(1+\alpha)\,q_1\,q_2+(1+2\,\alpha+2\,\alpha^2)\,q_2^2\right], \tag{11.76}$$

respectively, where $\alpha=m/M$.

A comparison of the above expressions with the standard forms (11.6) and (11.11) yields the following expressions for the mass matrix, **M**, and the force matrix, **G**:

$$\mathbf{M} = m\begin{pmatrix} 1+\alpha & \alpha \\ \alpha & 1+\alpha \end{pmatrix}, \tag{11.77}$$

$$\mathbf{G} = -k\begin{pmatrix} 1+2\,\alpha+2\,\alpha^2 & 2\,\alpha\,(1+\alpha) \\ 2\,\alpha\,(1+\alpha) & 1+2\,\alpha+2\,\alpha^2 \end{pmatrix}. \tag{11.78}$$

Now, the equation of motion of the system takes the form [see Equation (11.17)]

$$(\mathbf{G}-\lambda\,\mathbf{M})\,\mathbf{x}=\mathbf{0}, \tag{11.79}$$

where **x** is the column vector of the q_1 and q_2 values. The solubility condition for the above equation is

$$|\mathbf{G}-\lambda\,\mathbf{M}|=0, \tag{11.80}$$

which yields the following quadratic equation for the eigenvalue λ:

$$(1+2\,\alpha)\,\left[m^2\,\lambda^2+2\,m\,k\,(1+\alpha)\,\lambda+k^2\,(1+2\alpha)\right]=0. \tag{11.81}$$

The two roots of the above equation are

$$\lambda_1 = -\frac{k}{m}, \tag{11.82}$$

$$\lambda_2 = -\frac{k\,(1+2\alpha)}{m}. \tag{11.83}$$

The fact that the roots are negative implies that both normal modes are indeed *oscillatory* in nature. The characteristic oscillation frequencies are

$$\omega_1 = \sqrt{-\lambda_1} = \sqrt{\frac{k}{m}}, \tag{11.84}$$

$$\omega_2 = \sqrt{-\lambda_2} = \sqrt{\frac{k\,(1+2\,\alpha)}{m}}. \tag{11.85}$$

Equation (11.79) can now be solved, subject to the normalization condition (11.30), to give the two eigenvectors:

$$\mathbf{x}_1 = (2\,m)^{-1/2}\,(1,-1), \tag{11.86}$$

$$\mathbf{x}_2 = (2\,m)^{-1/2}\,(1+2\,\alpha)^{-1/2}\,(1,\ \ 1). \tag{11.87}$$

Thus, we conclude from Equations (11.74) and (11.84)–(11.87) that our model molecule possesses two normal modes of oscillation. The first mode oscillates at the frequency ω_1, and is an *anti-symmetric* mode in which $q_1 = -q_2$ and $q_3 = 0$. In other words, in this mode of oscillation, the two end atoms move in opposite directions whilst the central atom remains stationary. The second mode oscillates at the frequency ω_2, and is a mixed symmetry mode in which $q_1 = q_2$ but $q_3 = -2\,\alpha\,q_1$. In other words, in this mode of oscillation, the two end atoms move in the same direction whilst the central atom moves in the opposite direction.

Finally, it is easily demonstrated that the normal coordinates of the system are

$$\eta_1 = \sqrt{\frac{m}{2}}\,(q_1 - q_2), \tag{11.88}$$

$$\eta_2 = \sqrt{\frac{m\,(1+2\,\alpha)}{k}}\,(q_1 + q_2). \tag{11.89}$$

When expressed in terms of these coordinates, K and U reduce to

$$K = \frac{1}{2}\,(\dot\eta_1^2 + \dot\eta_2^2), \tag{11.90}$$

$$U = \frac{1}{2}\,(\omega_1^2\,\eta_1^2 + \omega_2^2\,\eta_2^2), \tag{11.91}$$

respectively.

11.9 Exercises

11.1. A particle of mass m is attached to a rigid support by means of a spring of spring constant k. At equilibrium, the spring hangs vertically downward. An identical oscillator is added to this system, the spring of the former being attached to the mass of the latter. Calculate the characteristic frequencies for one-dimensional vertical oscillations, and describe the associated normal modes.

11.2. A simple pendulum consists of a bob of mass m suspended by an inextensible light string of length l. From the bob of this pendulum, a second identical pendulum is suspended. Consider the case of small angle oscillations in the same vertical plane. Calculate the characteristic frequencies, and describe the associated normal modes.

11.3. A thin hoop of radius a and mass m oscillates in its own plane (which is constrained to be vertical) hanging from a single fixed point. A small mass m slides without friction along the hoop. Consider the case of small oscillations. Calculate the characteristic frequencies, and describe the associated normal modes.

11.4. A simple pendulum of mass m and length l is suspended from a block of mass M which is constrained to slide along a frictionless horizontal track. Consider the case of small oscillations. Calculate the characteristic frequencies, and describe the associated normal modes.

11.5. A thin uniform rod of mass m and length a is suspended from one end by a light string of length l. The other end of the string is attached to a fixed support. Consider the case of small oscillations in a vertical plane. Calculate the characteristic frequencies, and describe the associated normal modes.

11.6. A triatomic molecule consists of three atoms of equal mass. Each atom is attached to the other two atoms via identical chemical bonds. The equilibrium state of the molecule is such that the masses are at the vertices of an equilateral triangle of side a. Modeling the chemical bonds as springs of spring constant k, and only considering motion in the plane of the molecule, find the vibrational frequencies and normal modes of the molecule. Exclude translational and rotational modes.

12 Gravitational Potential Theory

12.1 Introduction

This chapter employs gravitational potential theory in combination with Newtonian dynamics to examine various interesting phenomena in the Solar System.

12.2 Gravitational Potential

Consider two point masses, m and m', located at position vectors $\mathbf{r}$ and $\mathbf{r}'$, respectively. According to Section 5.3, the acceleration $\mathbf{g}$ of mass m due to the gravitational force exerted on it by mass m' takes the form

$$\mathbf{g} = G\, m' \frac{(\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3}. \tag{12.1}$$

Now, the x-component of this acceleration is written

$$g_x = G\, m' \frac{(x' - x)}{[(x' - x)^2 + (y' - y)^2 + (z' - z)^2]^{3/2}}, \tag{12.2}$$

where $\mathbf{r} = (x, y, z)$ and $\mathbf{r}' = (x', y', z')$. However, as is easily demonstrated,

$$\frac{(x' - x)}{[(x' - x)^2 + (y' - y)^2 + (z' - z)^2]^{3/2}} \equiv \frac{\partial}{\partial x}\left(\frac{1}{[(x' - x)^2 + (y' - y)^2 + (z' - z)^2]^{1/2}}\right). \tag{12.3}$$

Hence,

$$g_x = G\, m' \frac{\partial}{\partial x}\left(\frac{1}{|\mathbf{r}' - \mathbf{r}|}\right), \tag{12.4}$$

with analogous expressions for g_y and g_z. It follows that

$$\mathbf{g} = -\nabla\Phi, \tag{12.5}$$

where

$$\Phi = -\frac{G\, m'}{|\mathbf{r}' - \mathbf{r}|} \tag{12.6}$$

is termed the *gravitational potential*. Of course, we can only write $\mathbf{g}$ in the form (12.5) because gravity is a *conservative* force—see Chapter 2. Note that gravitational potential, Φ, is directly related to gravitational potential energy, U. In fact, the potential energy of mass m is $U = m\,\Phi$.

Now, it is well-known that gravity is a *superposable* force. In other words, the gravitational force exerted on some test mass by a collection of point masses is simply the sum of the forces exerted on the test mass by each point mass taken in isolation. It follows that the gravitational potential generated by a collection of point masses at a certain location in space is the sum of the potentials generated at that location by each point mass taken in isolation. Hence, using Equation (12.6), if there are N point masses, m_i (for $i = 1, N$), located at position vectors $\mathbf{r}_i$, then the gravitational potential generated at position vector $\mathbf{r}$ is simply

$$\Phi(\mathbf{r}) = -G \sum_{i=1,N} \frac{m_i}{|\mathbf{r}_i - \mathbf{r}|}. \tag{12.7}$$

Suppose, finally, that, instead of having a collection of point masses, we have a *continuous* mass distribution. In other words, let the mass at position vector $\mathbf{r}'$ be $\rho(\mathbf{r}')\, d^3\mathbf{r}'$, where $\rho(\mathbf{r}')$ is the local mass density, and $d^3\mathbf{r}'$ a volume element. Summing over all space, and taking the limit $d^3\mathbf{r}' \rightarrow 0$, Equation (12.7) yields

$$\Phi(\mathbf{r}) = -G \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|}\, d^3\mathbf{r}', \tag{12.8}$$

where the integral is taken over all space. This is the general expression for the gravitational potential, $\Phi(\mathbf{r})$, generated by a continuous mass distribution, $\rho(\mathbf{r})$.

12.3 Axially Symmetric Mass Distributions

At this point, it is convenient to adopt standard spherical coordinates, (r, θ, ϕ), aligned along the z-axis. These coordinates are related to regular Cartesian coordinates as follows (see Section A.17):

$$x = r \sin\theta \cos\phi, \tag{12.9}$$

$$y = r \sin\theta \sin\phi, \tag{12.10}$$

$$z = r \cos\theta. \tag{12.11}$$

Consider an *axially symmetric* mass distribution: *i.e.*, a $\rho(\mathbf{r})$ which is *independent* of the azimuthal angle, ϕ. We would expect such a mass distribution to generated an axially symmetric gravitational potential, $\Phi(r, \theta)$. Hence, without loss of generality, we can set $\phi = 0$ when evaluating Φ from Equation (12.8). In fact, given that $d^3\mathbf{r}' = r'^2 \sin\theta'\, dr'\, d\theta'\, d\phi'$ in spherical coordinates, this equation yields

$$\Phi(r, \theta) = -G \int_0^\infty \int_0^\pi \int_0^{2\pi} \frac{r'^2\, \rho(r', \theta')\, \sin\theta'}{|\mathbf{r} - \mathbf{r}'|}\, dr'\, d\theta'\, d\phi', \tag{12.12}$$

with the right-hand side evaluated at $\phi = 0$. However, since $\rho(r', \theta')$ is independent of ϕ', the above equation can also be written

$$\Phi(r, \theta) = -2\pi G \int_0^\infty \int_0^\pi r'^2\, \rho(r', \theta')\, \sin\theta' \, \langle |\mathbf{r} - \mathbf{r}'|^{-1} \rangle\, dr'\, d\theta', \tag{12.13}$$

where $\langle\cdots\rangle$ denotes an average over the azimuthal angle, ϕ'.

Now,

$$|\mathbf{r}'-\mathbf{r}|^{-1} = (r^2 - 2\,\mathbf{r}\cdot\mathbf{r}' + r'^2)^{-1/2}, \tag{12.14}$$

and

$$\mathbf{r}\cdot\mathbf{r}' = r\,r'\,F, \tag{12.15}$$

where (at $\phi = 0$)

$$F = \sin\theta\,\sin\theta'\,\cos\phi' + \cos\theta\,\cos\theta'. \tag{12.16}$$

Hence,

$$|\mathbf{r}'-\mathbf{r}|^{-1} = (r^2 - 2\,r\,r'\,F + r'^2)^{-1/2}. \tag{12.17}$$

Suppose that $r > r'$. In this case, we can expand $|\mathbf{r}'-\mathbf{r}|^{-1}$ as a convergent power series in r'/r, to give

$$|\mathbf{r}'-\mathbf{r}|^{-1} = \frac{1}{r}\left[1 + \left(\frac{r'}{r}\right) F + \frac{1}{2}\left(\frac{r'}{r}\right)^2 (3\,F^2 - 1) + \mathcal{O}\left(\frac{r'}{r}\right)^3\right]. \tag{12.18}$$

Let us now average this expression over the azimuthal angle, ϕ'. Since $\langle 1\rangle = 1$, $\langle\cos\phi'\rangle = 0$, and $\langle\cos^2\phi'\rangle = 1/2$, it is easily seen that

$$\begin{aligned}\langle F\rangle &= \cos\theta\,\cos\theta', &(12.19)\\ \langle F^2\rangle &= \frac{1}{2}\,\sin^2\theta\,\sin^2\theta' + \cos^2\theta\,\cos^2\theta'\\ &= \frac{1}{3} + \frac{2}{3}\left(\frac{3}{2}\,\cos^2\theta - \frac{1}{2}\right)\left(\frac{3}{2}\,\cos^2\theta' - \frac{1}{2}\right). &(12.20)\end{aligned}$$

Hence,

$$\begin{aligned}\left\langle|\mathbf{r}'-\mathbf{r}|^{-1}\right\rangle &= \frac{1}{r}\left[1 + \left(\frac{r'}{r}\right)\cos\theta\,\cos\theta'\right. &(12.21)\\ &\left. + \left(\frac{r'}{r}\right)^2\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)\left(\frac{3}{2}\cos^2\theta' - \frac{1}{2}\right) + \mathcal{O}\left(\frac{r'}{r}\right)^3\right].\end{aligned}$$

Now, the well-known *Legendre polynomials*, $P_n(x)$, are defined

$$P_n(x) = \frac{1}{2^n\,n!}\,\frac{d^n}{dx^n}\left[(x^2-1)^n\right], \tag{12.22}$$

for $n = 0, \infty$. It follows that

$$\begin{aligned}P_0(x) &= 1, &(12.23)\\ P_1(x) &= x, &(12.24)\\ P_2(x) &= \frac{1}{2}\,(3\,x^2 - 1), &(12.25)\end{aligned}$$

etc. The Legendre polynomials are *mutually orthogonal*: *i.e.*,

$$\int_{-1}^{1} P_n(x)\, P_m(x)\, dx = \int_0^{\pi} P_n(\cos\theta)\, P_m(\cos\theta)\, \sin\theta\, d\theta = \frac{\delta_{nm}}{n+1/2}. \tag{12.26}$$

Here, δ_{nm} is 1 if $n = m$, and 0 otherwise. The Legendre polynomials also form a *complete set*: *i.e.*, any well-behaved function of x can be represented as a weighted sum of the $P_n(x)$. Likewise, any well-behaved (even) function of θ can be represented as a weighted sum of the $P_n(\cos\theta)$.

A comparison of Equation (12.21) and Equations (12.23)–(12.25) makes it reasonably clear that, when $r > r'$, the complete expansion of $\langle |\mathbf{r}' - \mathbf{r}|^{-1} \rangle$ is

$$\left\langle |\mathbf{r}' - \mathbf{r}|^{-1} \right\rangle = \frac{1}{r} \sum_{n=0,\infty} \left(\frac{r'}{r}\right)^n P_n(\cos\theta)\, P_n(\cos\theta'). \tag{12.27}$$

Similarly, when $r < r'$, we can expand in powers of r/r' to obtain

$$\left\langle |\mathbf{r}' - \mathbf{r}|^{-1} \right\rangle = \frac{1}{r'} \sum_{n=0,\infty} \left(\frac{r}{r'}\right)^n P_n(\cos\theta)\, P_n(\cos\theta'). \tag{12.28}$$

It follows from Equations (12.13), (12.27), and (12.28) that

$$\Phi(r,\theta) = \sum_{n=0,\infty} \Phi_n(r)\, P_n(\cos\theta), \tag{12.29}$$

where

$$\begin{aligned} \Phi_n(r) &= -\frac{2\pi G}{r^{n+1}} \int_0^r \int_0^{\pi} r'^{\,n+2}\, \rho(r',\theta')\, P_n(\cos\theta')\, \sin\theta'\, dr'\, d\theta' \\ &\quad -2\pi G\, r^n \int_r^{\infty} \int_0^{\pi} r'^{\,1-n}\, \rho(r',\theta')\, P_n(\cos\theta')\, \sin\theta'\, dr'\, d\theta'. \end{aligned} \tag{12.30}$$

Now, given that the $P_n(\cos\theta)$ form a complete set, we can always write

$$\rho(r,\theta) = \sum_{n=0,\infty} \rho_n(r)\, P_n(\cos\theta). \tag{12.31}$$

This expression can be inverted, with the aid of Equation (12.26), to give

$$\rho_n(r) = (n+1/2) \int_0^{\pi} \rho(r,\theta)\, P_n(\cos\theta)\, \sin\theta\, d\theta. \tag{12.32}$$

Hence, Equation (12.30) reduces to

$$\Phi_n(r) = -\frac{2\pi G}{(n+1/2)\, r^{n+1}} \int_0^r r'^{\,n+2}\, \rho_n(r')\, dr' - \frac{2\pi G\, r^n}{n+1/2} \int_r^{\infty} r'^{\,1-n}\, \rho_n(r')\, dr'. \tag{12.33}$$

Thus, we now have a general expression for the gravitational potential, $\Phi(r,\theta)$, generated by any axially symmetric mass distribution, $\rho(r,\theta)$.

12.4 Potential Due to a Uniform Sphere

Let us calculate the gravitational potential generated by a sphere of uniform mass density γ and radius a. Expressing $\rho(r,\theta)$ in the form (12.31), it is clear that

$$\rho_0(r) = \begin{cases} \gamma & \text{for } r \leq a \\ 0 & \text{for } r > a \end{cases}, \tag{12.34}$$

with $\rho_n(r) = 0$ for $n > 0$. Thus, from (12.33),

$$\Phi_0(r) = -\frac{4\pi G\gamma}{r}\int_0^r r'^2\,dr' - 4\pi G\gamma\int_r^a r'\,dr' \tag{12.35}$$

for $r \leq a$, and

$$\Phi_0(r) = -\frac{4\pi G\gamma}{r}\int_0^a r'^2\,dr' \tag{12.36}$$

for $r > a$, with $\Phi_n(r) = 0$ for $n > 0$. Hence,

$$\Phi(r) = -\frac{2\pi G\gamma}{3}(3a^2 - r^2) = -G M\frac{(3a^2 - r^2)}{2a^3} \tag{12.37}$$

for $r \leq a$, and

$$\Phi(r) = -\frac{4\pi G\gamma}{3}\frac{a^3}{r} = -\frac{G M}{r} \tag{12.38}$$

for $r > a$. Here, $M = (4\pi/3)\,a^3\gamma$ is the total mass of the sphere.

According to Equation (12.38), the gravitational potential *outside* a uniform sphere of mass M is the same as that generated by a point mass M located at the sphere's center. It turns out that this is a general result for *any* finite spherically symmetric mass distribution. Indeed, from the previous analysis, it is clear that $\rho(r) = \rho_0(r)$ and $\Phi(r) = \Phi_0(r)$ for a spherically symmetric mass distribution. Suppose that the mass distribution extends out to $r = a$. It immediately follows, from Equation (12.33), that

$$\Phi(r) = -\frac{G}{r}\int_0^a 4\pi r'^2\,\rho(r')\,dr' = -\frac{G M}{r} \tag{12.39}$$

for $r > a$, where M is the total mass of the distribution. We, thus, conclude that Newton's laws of motion, in their primitive form, apply not only to point masses, but also to extended spherically symmetric masses. In fact, this is something which we have implicitly assumed all along in this book.

According to Equation (12.37), the gravitational potential *inside* a uniform sphere is *quadratic* in r. This implies that if a narrow shaft were drilled though the center of the sphere then a test mass, m, moving in this shaft would experience a gravitational force acting toward the center which scales *linearly* in r. In fact, the force in question is given

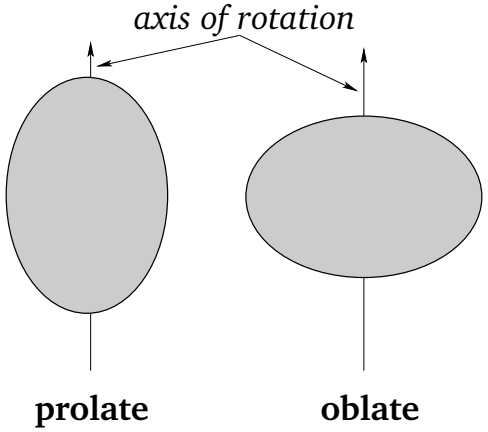


Figure 12.1: *Prolate and oblate spheroids.*

by $f_r = -m\,\partial\Phi/\partial r = -(G\,m\,M/a^3)\,r$. It follows that a test mass dropped into the shaft executes simple harmonic motion about the center of the sphere with period

$$T = 2\pi\sqrt{\frac{a}{g}}, \tag{12.40}$$

where $g = G\,M/a^2$ is the gravitational acceleration at the sphere's surface.

12.5 Potential Outside a Uniform Spheroid

Let us now calculate the gravitational potential generated outside a *spheroid* of uniform mass density γ and mean radius a. A spheroid is the solid body produced by rotating an ellipse about a major or minor axis. Let the axis of rotation coincide with the z-axis, and let the outer boundary of the spheroid satisfy

$$r = a_\theta(\theta) = a\left[1 - \frac{2}{3}\,\epsilon\,P_2(\cos\theta)\right], \tag{12.41}$$

where ϵ is the termed the *ellipticity*. Here, we are assuming that $|\epsilon| \ll 1$, so that the spheroid is very close to being a sphere. If $\epsilon > 0$ then the spheroid is slightly squashed along its symmetry axis, and is termed *oblate*. Likewise, if $\epsilon < 0$ then the spheroid is slightly elongated along its axis, and is termed *prolate*—see Figure 12.1. Of course, if $\epsilon = 0$ then the spheroid reduces to a sphere.

Now, according to Equation (12.29) and (12.30), the gravitational potential generated *outside* an axially symmetric mass distribution can be written

$$\Phi(r,\theta) = \sum_{n=0,\infty} J_n\,\frac{P_n(\cos\theta)}{r^{n+1}}, \tag{12.42}$$

where

$$J_n = -2\pi\,G\int\int r^{2+n}\,\rho(r,\theta)\,P_n(\cos\theta)\,\sin\theta\,dr\,d\theta. \tag{12.43}$$

Here, the integral is taken over the whole cross-section of the distribution in r–θ space.

It follows that for a uniform spheroid

$$J_n = -2\pi\,G\,\gamma\int_0^\pi P_n(\cos\theta)\int_0^{a_\theta(\theta)} r^{2+n}\,dr\,\sin\theta\,d\theta \tag{12.44}$$

Hence,

$$J_n = -\frac{2\pi\,G\,\gamma}{(3+n)}\int_0^\pi P_n(\cos\theta)\,a_\theta^{3+n}(\theta)\,\sin\theta\,d\theta, \tag{12.45}$$

giving

$$J_n \simeq -\frac{2\pi\,G\,\gamma\,a^{3+n}}{(3+n)}\int_0^\pi P_n(\cos\theta)\left[P_0(\cos\theta)-\frac{2}{3}\,(3+n)\,\epsilon\,P_2(\cos\theta)\right]\sin\theta\,d\theta, \tag{12.46}$$

to first-order in ϵ. It is thus clear, from Equation (12.26), that, to first-order in ϵ, the only non-zero J_n are

$$J_0 = -\frac{4\pi\,G\,\gamma\,a^3}{3} = -G\,M, \tag{12.47}$$

$$J_2 = \frac{8\pi\,G\,\gamma\,a^5\,\epsilon}{15} = \frac{2}{5}\,G\,M\,a^2\,\epsilon, \tag{12.48}$$

since $M = (4\pi/3)\,a^3\,\gamma$.

Thus, the gravitational potential outside a uniform spheroid of total mass M, mean radius a, and ellipticity ϵ, is

$$\Phi(r,\theta) = -\frac{G\,M}{r}+\frac{2}{5}\frac{G\,M\,a^2}{r^3}\,\epsilon\,P_2(\cos\theta)+\mathcal{O}(\epsilon^2). \tag{12.49}$$

In particular, the gravitational potential on the surface of the spheroid is

$$\Phi(a_\theta,\theta) = -\frac{G\,M}{a_\theta}+\frac{2}{5}\frac{G\,M\,a^2}{a_\theta^3}\,\epsilon\,P_2(\cos\theta)+\mathcal{O}(\epsilon^2), \tag{12.50}$$

which yields

$$\Phi(a_\theta,\theta) \simeq -\frac{G\,M}{a}\left[1+\frac{4}{15}\,\epsilon\,P_2(\cos\theta)+\mathcal{O}(\epsilon^2)\right], \tag{12.51}$$

where use has been made of Equation (12.41).

Consider a self-gravitating spheroid of mass M, mean radius a, and ellipticity ϵ: *e.g.*, a star, or a planet. Assuming, for the sake of simplicity, that the spheroid is composed of uniform density incompressible fluid, the gravitational potential on its surface is given by Equation (12.51). However, the condition for an equilibrium state is that the potential be

constant over the surface. If this is not the case then there will be gravitational forces acting *tangential* to the surface. Such forces cannot be balanced by internal pressure, which only acts *normal* to the surface. Hence, from (12.51), it is clear that the condition for equilibrium is $\epsilon = 0$. In other words, the equilibrium configuration of a self-gravitating mass is a *sphere*. Deviations from this configuration can only be caused by forces in addition to self-gravity and internal pressure: *e.g.*, centrifugal forces due to rotation, or tidal forces due to orbiting masses.

12.6 Rotational Flattening

Let us consider the equilibrium configuration of a self-gravitating spheroid, composed of uniform density incompressible fluid, which is *rotating* steadily about some fixed axis. Let M be the total mass, a the mean radius, ϵ the ellipticity, and Ω the angular velocity of rotation. Furthermore, let the axis of rotation coincide with the axis of symmetry, which is assumed to run along the z-axis.

Let us transform to a non-inertial frame of reference which co-rotates with the spheroid about the z-axis, and in which the spheroid consequently appears to be stationary. From Chapter 7, the problem is now analogous to that of a non-rotating spheroid, except that the surface acceleration is written $\mathbf{g} = \mathbf{g}_g + \mathbf{g}_c$, where $\mathbf{g}_g = -\nabla\Phi(r,\theta)$ is the gravitational acceleration, and $\mathbf{g}_c$ the *centrifugal* acceleration. The latter acceleration is of magnitude $r\sin\theta\,\Omega^2$, and is everywhere directed away from the axis of rotation (see Figure 12.2 and Chapter 7). The acceleration thus has components

$$\mathbf{g}_c = r\,\Omega^2\,\sin\theta\,(\sin\theta,\,\cos\theta,\,0) \tag{12.52}$$

in spherical polar coordinates. It follows that $\mathbf{g}_c = -\nabla\chi$, where

$$\chi(r,\theta) = -\frac{\Omega^2\,r^2}{2}\,\sin^2\theta = -\frac{\Omega^2\,r^2}{3}\,[1 - P_2(\cos\theta)] \tag{12.53}$$

can be thought of as a sort of centrifugal potential. Hence, the total surface acceleration is

$$\mathbf{g} = -\nabla(\Phi + \chi). \tag{12.54}$$

As before, the criterion for an equilibrium state is that the surface lie at a constant total potential, so as to eliminate tangential surface forces which cannot be balanced by internal pressure. Hence, assuming that the surface satisfies Equation (12.41), the equilibrium configuration is specified by

$$\Phi(a_\theta,\theta) + \chi(a_\theta,\theta) = c, \tag{12.55}$$

where c is a constant. It follows from Equations (12.51) and (12.53) that, to first-order in ϵ and Ω^2,

$$-\frac{G\,M}{a}\left[1 + \frac{4}{15}\,\epsilon\,P_2(\cos\theta)\right] - \frac{\Omega^2\,a^2}{3}\,[1 - P_2(\cos\theta)] \simeq c, \tag{12.56}$$

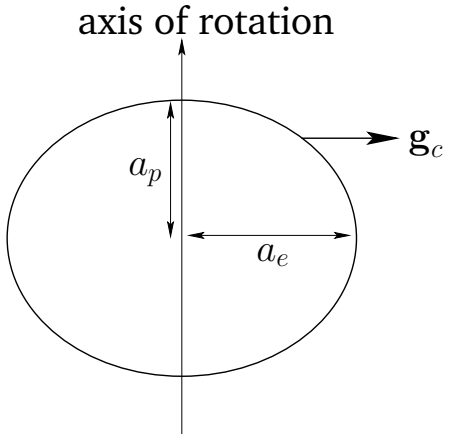


Figure 12.2: *Rotational flattening.*

which yields

$$\epsilon = \frac{5}{4}\,\frac{\Omega^2\,a^3}{G\,M}. \tag{12.57}$$

We conclude, from the above expression, that the equilibrium configuration of a (relatively slowly) rotating self-gravitating mass distribution is an *oblate spheroid*: *i.e.*, a sphere which is slightly flattened along its axis of rotation. The degree of flattening is proportional to the square of the rotation rate. Now, from (12.41), the mean radius of the spheroid is a, the radius at the poles (*i.e.*, along the axis of rotation) is $a_p = a\,(1 - 2\,\epsilon/3)$, and the radius at the equator (*i.e.*, perpendicular to the axis of rotation) is $a_e = a\,(1 + \epsilon/3)$—see Figure 12.2. Hence, the degree of rotational flattening can be written

$$\frac{a_e - a_p}{a} = \epsilon = \frac{5}{4}\,\frac{\Omega^2\,a^3}{G\,M}. \tag{12.58}$$

Now, for the Earth, $a = 6.37 \times 10^6$ m, $\Omega = 7.27 \times 10^{-5}$ rad./s, and $M = 5.97 \times 10^{24}$ kg. Thus, we predict that

$$\epsilon = 0.00429, \tag{12.59}$$

corresponding to a difference between equatorial and polar radii of

$$\Delta a = a_e - a_p = \epsilon\,a = 27.3\,\text{km}. \tag{12.60}$$

In fact, the observed degree of flattening of the Earth is $\epsilon = 0.00335$, corresponding to a difference between equatorial and polar radii of 21.4 km. The main reason that our analysis has overestimated the degree of rotational flattening of the Earth is that it models the terrestrial interior as a uniform density incompressible fluid. In reality, the Earth's core is much denser than its crust (see Exercise 12.1).

For Jupiter, $a = 6.92 \times 10^7$ m, $\Omega = 1.76 \times 10^{-4}$ rad./s, and $M = 1.90 \times 10^{27}$ kg. Hence, we predict that

$$\epsilon = 0.101. \tag{12.61}$$

Note that this degree of flattening is much larger than that of the Earth, due to Jupiter's relatively large radius (about 10 times that of Earth), combined with its relatively short rotation period (about 0.4 days). In fact, the polar flattening of Jupiter is clearly apparent from images of this planet. The observed degree of polar flattening of Jupiter is actually $\epsilon = 0.065$. Our estimate of ϵ is probably slightly too large because Jupiter, which is mostly gaseous, has a mass distribution which is strongly concentrated at its core (see Exercise 12.1).

12.7 McCullough's Formula

According to Equations (12.43) and (12.48), if the Earth is modeled as spheroid of uniform density γ then its ellipticity is given by

$$\epsilon = -\int r^2 \gamma\, P_2(\cos\theta)\, d^3\mathbf{r} \Big/ I_0 = -\frac{1}{2}\int r^2 \gamma \left(3\cos^2\theta - 1\right) d^3\mathbf{r} \Big/ I_0, \quad (12.62)$$

where the integral is over the whole volume of the Earth, and $I_0 = (2/5)\, M\, a^2$ would be the Earth's moment of inertia were it exactly spherical. Now, the Earth's moment of inertia about its axis of rotation is given by

$$I_\parallel = \int (x^2 + y^2)\, \gamma\, d^3\mathbf{r} = \int r^2 \gamma\, (1 - \cos^2\theta)\, d^3\mathbf{r}. \quad (12.63)$$

Here, use has been made of Equations (12.9)–(12.11). Likewise, the Earth's moment of inertia about an axis perpendicular to its axis of rotation (and passing through the Earth's center) is

$$\begin{aligned} I_\perp &= \int (y^2 + z^2)\, \gamma\, d^3\mathbf{r} = \int r^2 \gamma\, (\sin^2\theta\, \sin^2\phi + \cos^2\theta)\, d^3\mathbf{r} \\ &= \int r^2 \gamma \left(\frac{1}{2}\sin^2\theta + \cos^2\theta\right) d^3\mathbf{r} = \frac{1}{2}\int r^2 \gamma\, (1 + \cos^2\theta)\, d^3\mathbf{r}, \end{aligned} \quad (12.64)$$

since the average of $\sin^2\phi$ is $1/2$ for an axisymmetric mass distribution. It follows from the above three equations that

$$\epsilon = \frac{I_\parallel - I_\perp}{I_0} \simeq \frac{I_\parallel - I_\perp}{I_\parallel}. \quad (12.65)$$

This result, which is known as *McCullough's formula*, demonstrates that the Earth's ellipticity is directly related to the difference between its principle moments of inertia. It turns out that McCullough's formula holds for *any* axially symmetric mass distribution, and not just a spheroidal distribution with uniform density. Finally, McCullough's formula can be combined with Equation (12.49) to give

$$\Phi(r,\theta) = -\frac{G\, M}{r} + \frac{G\, (I_\parallel - I_\perp)}{r^3}\, P_2(\cos\theta). \quad (12.66)$$

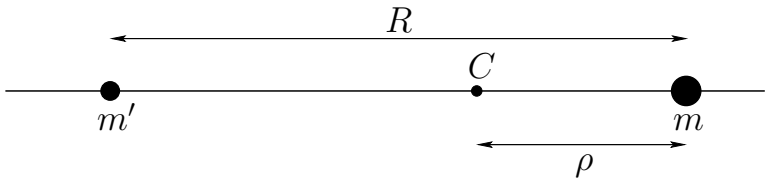


Figure 12.3: *Two orbiting masses.*

This is the general expression for the gravitational potential generated outside an axially symmetric mass distribution. The first term on the right-hand side is the *monopole* gravitational field which would be generated if all of the mass in the distribution were concentrated at its center of mass, whereas the second term is the *quadrupole* field generated by any deviation from spherical symmetry in the distribution.

12.8 Tidal Elongation

Consider two point masses, m and m', executing circular orbits about their common center of mass, C, with angular velocity ω. Let R be the distance between the masses, and ρ the distance between point C and mass m—see Figure 12.3. We know, from Section 6.3, that

$$\omega^2 = \frac{G\,M}{R^3}, \tag{12.67}$$

and

$$\rho = \frac{m'}{M}\,R, \tag{12.68}$$

where $M = m + m'$.

Let us transform to a non-inertial frame of reference which *rotates*, about an axis perpendicular to the orbital plane and passing through C, at the angular velocity ω. In this reference frame, both masses appear to be *stationary*. Consider mass m. In the rotating frame, this mass experiences a gravitational acceleration

$$a_g = \frac{G\,m'}{R^2} \tag{12.69}$$

directed *toward* the center of mass, and a centrifugal acceleration (see Chapter 7)

$$a_c = \omega^2\,\rho \tag{12.70}$$

directed *away from* the center of mass. However, it is easily demonstrated, using Equations (12.67) and (12.68), that

$$a_c = a_g. \tag{12.71}$$

In other words, the gravitational and centrifugal accelerations *balance*, as must be the case if mass m is to remain stationary in the rotating frame. Let us investigate how this balance is affected if the masses m and m' have *finite* spatial extents.

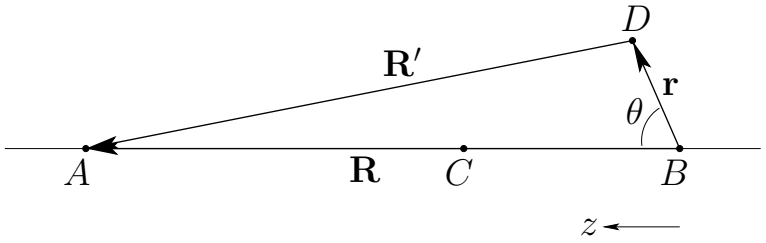


Figure 12.4: *Calculation of tidal forces.*

Let the center of the mass distribution m' lie at A, the center of the mass distribution m at B, and the center of mass at C—see Figure 12.4. We wish to calculate the centrifugal and gravitational accelerations at some point D in the vicinity of point B. It is convenient to adopt spherical coordinates, centered on point B, and aligned such that the z-axis coincides with the line BA.

Let us assume that the mass distribution m is orbiting around C, but is *not* rotating about an axis passing through its center, in order to exclude rotational flattening from our analysis. If this is the case then it is easily seen that each constituent point of m executes circular motion of angular velocity ω and radius ρ—see Figure 12.5. Hence, each constituent point experiences the *same* centrifugal acceleration: *i.e.*,

$$\mathbf{g}_c = -\omega^2\,\rho\,\mathbf{e}_z. \tag{12.72}$$

It follows that

$$\mathbf{g}_c = -\nabla\chi, \tag{12.73}$$

where

$$\chi = \omega^2\,\rho\,z \tag{12.74}$$

is the centrifugal potential, and $z = r\,\cos\theta$. The centrifugal potential can also be written

$$\chi = \frac{G\,m'}{R}\frac{r}{R}\,P_1(\cos\theta). \tag{12.75}$$

The gravitational acceleration at point D due to mass m' is given by

$$\mathbf{g}_g = -\nabla\Phi', \tag{12.76}$$

where the gravitational potential takes the form

$$\Phi' = -\frac{G\,m'}{R'}. \tag{12.77}$$

Here, R' is the distance between points A and D. Note that the gravitational potential generated by the mass distribution m' is the same as that generated by an equivalent point mass at A, as long as the distribution is spherically symmetric, which we shall assume to be the case.

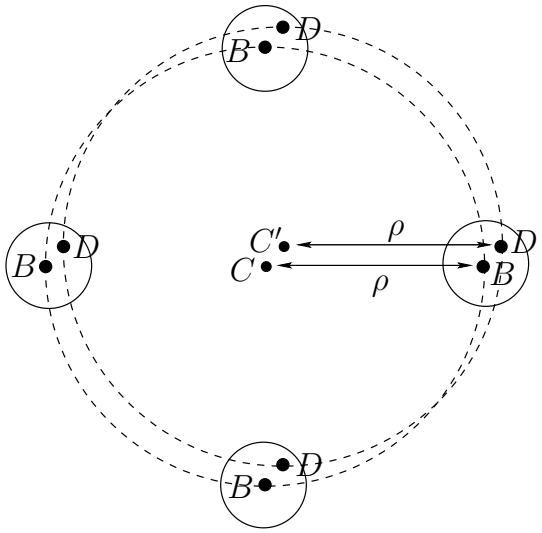


Figure 12.5: *The center* B *of the mass distribution* m *orbits about the center of mass* C *in a circle of radius* ρ*. If the mass distribution is non-rotating then a non-central point* D *must maintain a constant spatial relationship to* B*. It follows that point* D *orbits some point* C′*, which has the same spatial relationship to* C *that* D *has to* B*, in a circle of radius* ρ*.*

Now,

$$\mathbf{R}' = \mathbf{R} - \mathbf{r}, \tag{12.78}$$

where $\mathbf{R}'$ is the vector $\overrightarrow{DA}$, and $\mathbf{R}$ the vector $\overrightarrow{BA}$—see Figure 12.4. It follows that

$$R'^{-1} = \left(R^2 - 2\,\mathbf{R}\cdot\mathbf{r} + r^2\right)^{-1/2} = \left(R^2 - 2\,R\,r\,\cos\theta + r^2\right)^{-1/2}. \tag{12.79}$$

Expanding in powers of r/R, we obtain

$$R'^{-1} = \frac{1}{R}\sum_{n=0,\infty}\left(\frac{r}{R}\right)^n P_n(\cos\theta). \tag{12.80}$$

Hence,

$$\Phi' \simeq -\frac{G\,m'}{R}\left[1 + \frac{r}{R}\,P_1(\cos\theta) + \frac{r^2}{R^2}\,P_2(\cos\theta)\right] \tag{12.81}$$

to second-order in r/R.

Adding χ and Φ', we obtain

$$\chi + \Phi' \simeq -\frac{G\,m'}{R}\left[1 + \frac{r^2}{R^2}\,P_2(\cos\theta)\right] \tag{12.82}$$

to second-order in r/R. Note that $\chi+\Phi'$ is the potential due to the net *external* force acting on the mass distribution m. This potential is constant up to first-order in r/R, because the first-order variations in χ and Φ' cancel one another. The cancellation is a manifestation

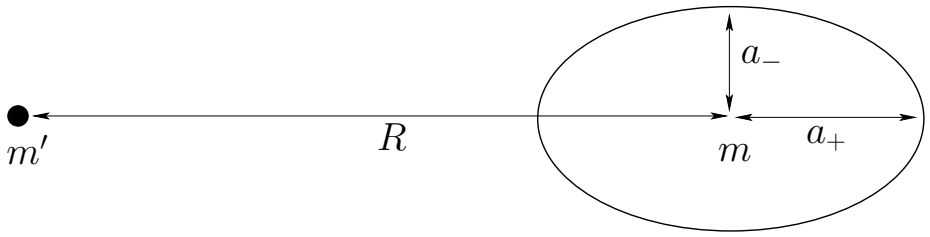


Figure 12.6: *Tidal elongation.*

of the balance between the centrifugal and gravitational accelerations in the equivalent point mass problem discussed above. However, this balance is only exact at the *center* of the mass distribution m. Away from the center, the centrifugal acceleration remains constant, whereas the gravitational acceleration increases with increasing z. Hence, at positive z, the gravitational acceleration is larger than the centrifugal, giving rise to a net acceleration in the $+z$-direction. Likewise, at negative z, the centrifugal acceleration is larger than the gravitational, giving rise to a net acceleration in the $-z$-direction. It follows that the mass distribution m is subject to a residual acceleration, represented by the second-order variation in Equation (12.82), which acts to *elongate* it along the z-axis. This effect is known as *tidal elongation*.

In order to calculate the tidal elongation of the mass distribution m we need to add the potential, $\chi + \Phi'$, due to the external forces, to the gravitational potential, Φ, generated by the distribution itself. Assuming that the mass distribution is spheroidal with mass m, mean radius a, and ellipticity ϵ, it follows from Equations (12.41), (12.51), and (12.82) that the total surface potential is given by

$$\begin{aligned}\chi + \Phi' + \Phi &\simeq -\frac{G\,m}{a} - \frac{G\,m'}{R} \\ &\quad -\frac{4}{15}\,\frac{G\,m}{a}\,\epsilon\,P_2(\cos\theta) - \frac{G\,m'\,a^2}{R^3}\,P_2(\cos\theta),\end{aligned} \tag{12.83}$$

where we have treated ϵ and a/R as small quantities. As before, the condition for equilibrium is that the total potential be constant over the surface of the spheroid. Hence, we obtain

$$\epsilon = -\frac{15}{4}\,\frac{m'}{m}\left(\frac{a}{R}\right)^3 \tag{12.84}$$

as our prediction for the ellipticity induced in a self-gravitating spherical mass distribution of total mass m and radius a by a second mass, m', which is in a circular orbit of radius R about the distribution. Thus, if a_+ is the maximum radius of the distribution, and a_- the minimum radius (see Figure 12.6), then

$$\frac{a_+ - a_-}{a} = -\epsilon = \frac{15}{4}\,\frac{m'}{m}\left(\frac{a}{R}\right)^3. \tag{12.85}$$

Consider the tidal elongation of the Earth due to the Moon. In this case, we have $a = 6.37 \times 10^6\,\mathrm{m}$, $R = 3.84 \times 10^8\,\mathrm{m}$, $m = 5.97 \times 10^{24}\,\mathrm{kg}$, and $m' = 7.35 \times 10^{22}\,\mathrm{kg}$. Hence, we calculate that $-\epsilon = 2.1 \times 10^{-7}$, or

$$\Delta a = a_+ - a_- = -\epsilon\, a = 1.34\,\mathrm{m}. \tag{12.86}$$

We, thus, predict that tidal forces due to the Moon cause the Earth to elongate along the axis joining its center to the Moon by about 1.3 meters. Since water is obviously more fluid than rock (especially on relatively short time-scales) most of this elongation takes place in the oceans rather than in the underlying land. Hence, the oceans rise, relative to the land, in the region of the Earth closest to the Moon, and also in the region furthest away. Since the Earth is rotating, whilst the tidal bulge of the oceans remains relatively stationary, the Moon's tidal force causes the ocean at a given point on the Earth's surface to rise and fall, by about a meter, *twice* daily, giving rise to the phenomenon known as the *tides*.

Consider the tidal elongation of the Earth due to the Sun. In this case, we have $a = 6.37 \times 10^6\,\mathrm{m}$, $R = 1.50 \times 10^{11}\,\mathrm{m}$, $m = 5.97 \times 10^{24}\,\mathrm{kg}$, and $m' = 1.99 \times 10^{30}\,\mathrm{kg}$. Hence, we calculate that $-\epsilon = 9.6 \times 10^{-8}$, or

$$\Delta a = a_+ - a_- = -\epsilon\, a = 0.61\,\mathrm{m}. \tag{12.87}$$

Thus, the tidal elongation due to the Sun is about half that due to the Moon. It follows that the tides are particularly high when the Sun, the Earth, and the Moon lie approximately in a straight-line, so that the tidal effects of the Sun and the Moon reinforce one another. This occurs at a new moon, or at a full moon. These type of tides are called *spring tides* (note that the name has nothing to do with the season). Conversely, the tides are particularly low when the Sun, the Earth, and the Moon form a right-angle, so that the tidal effects of the Sun and the Moon partially cancel one another. These type of tides are called *neap tides*. Generally speaking, we would expect two spring tides and two neap tides per month.

In reality, the amplitude of the tides varies significantly from place to place on the Earth's surface, due to the presence of the continents, which impede the flow of the oceanic tidal bulge around the Earth. Moreover, there is a time-lag of approximately 12 minutes between the Moon being directly overhead (or directly below) and high tide, because of the finite inertia of the oceans. Similarly, the time-lag between a spring tide and a full moon, or a new moon, can be up to 2 days.

12.9 Roche Radius

Consider a spherical moon of mass m and radius a which is in a circular orbit of radius R about a spherical planet of mass m' and radius a'. (Strictly speaking, the moon and the planet execute circular orbits about their common center of mass. However, if the planet is much more massive than the moon then the center of mass lies close to the planet's center.) According to the analysis in the previous section, a constituent element of the

moon experiences a force per unit mass, due to the gravitational field of the planet, which takes the form

$$\mathbf{g}' = -\nabla(\chi + \Phi'), \tag{12.88}$$

where

$$\chi + \Phi' = -\frac{G\,m'}{R^3}\,(z^2 - x^2/2 - y^2/2) + \text{const.} \tag{12.89}$$

Here, (x, y, z) is a Cartesian coordinate system whose origin is the center of the moon, and whose z-axis always points toward the center of the planet. It follows that

$$\mathbf{g}' = \frac{2\,G\,m'}{R^3}\left(-\frac{x}{2}\,\mathbf{e}_x - \frac{y}{2}\,\mathbf{e}_y + z\,\mathbf{e}_z\right). \tag{12.90}$$

This so-called *tidal force* is generated by the spatial variation of the planet's gravitational field over the interior of the moon, and acts to elongate the moon along an axis joining its center to that of the planet, and to compress it in any direction perpendicular to this axis. Note that the magnitude of the tidal force increases strongly as the radius, R, of the moon's orbit decreases. Now, if the tidal force becomes sufficiently strong then it can overcome the moon's self-gravity, and thereby rip the moon apart. It follows that there is a minimum radius, generally referred to as the *Roche radius*, at which a moon can orbit a planet without being destroyed by tidal forces.

Let us derive an expression for the Roche radius. Consider a small mass element at the point on the surface of the moon which lies closest to the planet, and at which the tidal force is consequently largest (*i.e.*, $x = y = 0$, $z = a$). According to Equation (12.90), the mass experiences an upward (from the moon's surface) tidal acceleration due to the gravitational attraction of the planet of the form

$$\mathbf{g}' = \frac{2\,G\,m'\,a}{R^3}\,\mathbf{e}_z. \tag{12.91}$$

The mass also experiences a downward gravitational acceleration due to the gravitational influence of the moon which is written

$$\mathbf{g} = -\frac{G\,m}{a^2}\,\mathbf{e}_z. \tag{12.92}$$

Thus, the effective surface gravity at the point in question is

$$g_{\rm eff} = \frac{G\,m}{a^2}\left(1 - 2\,\frac{m'}{m}\,\frac{a^3}{R^3}\right). \tag{12.93}$$

Note that if $R < R_c$, where

$$R_c = \left(2\,\frac{m'}{m}\right)^{1/3} a, \tag{12.94}$$

then the effective gravity is *negative*. In other words, the tidal force due to the planet is strong enough to overcome surface gravity and lift objects off the moon's surface. If this

is the case, and the tensile strength of the moon is negligible, then it is fairly clear that the tidal force will start to break the moon apart. Hence, R_c is the Roche radius. Now, $m'/m = (\rho'/\rho)\,(a'/a)^3$, where ρ and ρ' are the mean mass densities of the moon and planet, respectively. Thus, the above expression for the Roche radius can also be written

$$R_c = 1.41 \left(\frac{\rho'}{\rho}\right)^{1/3} a'. \tag{12.95}$$

The above calculation is somewhat inaccurate, since it fails to take into account the inevitable distortion of the moon's shape in the presence of strong tidal forces. (In fact, the calculation assumes that the moon always remains spherical.) A more accurate calculation, which treats the moon as a self-gravitating incompressible fluid, yields

$$R_c = 2.44 \left(\frac{\rho'}{\rho}\right)^{1/3} a'. \tag{12.96}$$

It follows that if the planet and the moon have the same mean density then the Roche radius is 2.44 times the planet's radius. Note that small orbital bodies such as rocks, or even very small moons, can survive intact within the Roche radius because they are held together by internal tensile forces rather than gravitational attraction. However, this mechanism becomes progressively less effective as the size of the body in question increases. Not surprisingly, virtually all large planetary moons occurring in the Solar System have orbital radii which exceed the relevant Roche radius, whereas virtually all planetary ring systems (which consist of myriads of small orbiting rocks) have radii which lie inside the relevant Roche radius.

12.10 Precession and Forced Nutation of the Earth

Consider the Earth-Sun system—see Figure 12.7. From a geocentric viewpoint, the Sun orbits the Earth counter-clockwise (looking from the north), once per year, in an approximately circular orbit of radius $a_s = 1.50 \times 10^{11}$ m. In astronomy, the plane of the Sun's apparent orbit relative to the Earth is known as the *ecliptic plane*. Let us define *non-rotating* Cartesian coordinates, centered on the Earth, which are such that the x- and y-axes lie in the ecliptic plane, and the z-axis is normal to this plane (in the sense that the Earth's north pole lies at positive z). It follows that the z-axis is directed toward a point in the sky (located in the constellation Draco) known as the *north ecliptic pole*. In the following, we shall treat the $Oxyz$ system as inertial. This is a reasonable approximation because the orbital acceleration of the Earth is much smaller than the acceleration due to its diurnal rotation. It is convenient to parameterize the instantaneous position of the Sun in terms of a counter-clockwise (looking from the north) azimuthal angle λ_s that is zero on the positive x-axis—see Figure 12.7.

Let $\boldsymbol{\omega}$ be the Earth's angular velocity vector due to its daily rotation. This vector makes an angle θ with the z-axis, where $\theta = 23.44^\circ$ is the mean inclination of the ecliptic to the

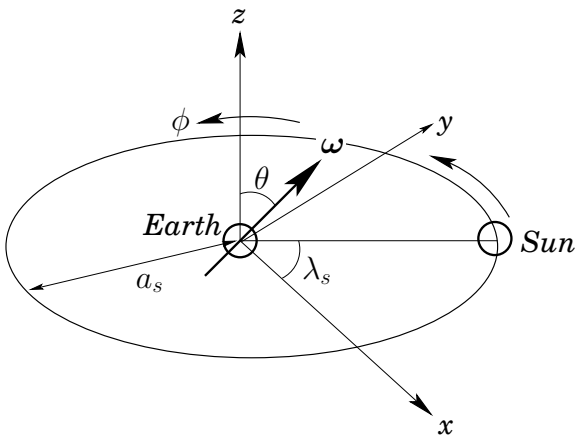


Figure 12.7: *The Earth-Sun system.*

Earth's equatorial plane. Suppose that the projection of $\boldsymbol{\omega}$ onto the ecliptic plane subtends an angle ϕ with the y-axis, where ϕ is measured in a counter-clockwise (looking from the north) sense—see Figure 12.7. The orientation of the Earth's axis of rotation (which is, of course, parallel to $\boldsymbol{\omega}$) is thus determined by the two angles θ and ϕ. Note, however, that these two angles are also *Euler angles*, in the sense given in Chapter 8. Let us examine the Earth-Sun system at an instant in time, $t = 0$, when $\phi = 0$: *i.e.*, when $\boldsymbol{\omega}$ lies in the y-z plane. At this particular instant, the x-axis points towards the so-called *vernal equinox*, which is defined as the point in the sky where the ecliptic plane crosses the projection of the Earth's equator (*i.e.*, the plane normal to $\boldsymbol{\omega}$) from south to north. A counter-clockwise (looking from the north) angle in the ecliptic plane that is zero at the vernal equinox is generally known as an *ecliptic longitude*. Thus, λ_s is the Sun's ecliptic longitude.

According to Equation (12.66), the potential energy of the Earth-Sun system is written

$$U = M_s\,\Phi = -\frac{G\,M_s\,M}{a_s} + \frac{G\,M_s\,(I_\parallel - I_\perp)}{a_s^3}\,P_2[\cos(\gamma_s)], \tag{12.97}$$

where M_s is the mass of the Sun, M the mass of the Earth, $I_\parallel$ the Earth's moment of inertia about its axis of rotation, $I_\perp$ the Earth's moment of inertia about an axis lying in its equatorial plane, and $P_2(x) = (1/2)\,(3\,x^2 - 1)$. Furthermore, γ_s is the angle subtended between $\boldsymbol{\omega}$ and $\mathbf{r}_s$, where $\mathbf{r}_s$ is the position vector of the Sun relative to the Earth.

It is easily demonstrated that (with $\phi = 0$)

$$\boldsymbol{\omega} = \omega\,(0,\,\sin\theta,\,\cos\theta), \tag{12.98}$$

and

$$\mathbf{r}_s = a_s\,(\cos\lambda_s,\,\sin\lambda_s,\,0). \tag{12.99}$$

Hence,

$$\cos\gamma_s = \frac{\boldsymbol{\omega}\cdot\mathbf{r}_s}{|\boldsymbol{\omega}|\,|\mathbf{r}_s|} = \sin\theta\,\sin\lambda_s, \tag{12.100}$$

giving

$$U = -\frac{G\,M_s\,M}{a_s} + \frac{G\,M_s\,(I_\parallel - I_\perp)}{2\,a_s^3}\,(3\,\sin^2\theta\,\sin^2\lambda_s - 1). \tag{12.101}$$

Now, we are primarily interested in the motion of the Earth's axis of rotation over timescales that are *much longer* than a year, so we can average the above expression over the Sun's orbit to give

$$U = -\frac{G\,M_s\,M}{a_s} + \frac{G\,M_s\,(I_\parallel - I_\perp)}{2\,a_s^3}\,\left(\frac{3}{2}\,\sin^2\theta - 1\right) \tag{12.102}$$

(since the average of $\sin^2\lambda_s$ over a year is $1/2$). Thus, we obtain

$$U = U_0 - \epsilon\,\alpha_s\,\cos(2\,\theta), \tag{12.103}$$

where U_0 is a constant, and

$$\alpha_s = \frac{3}{8}\,I_\parallel\,n_s^2. \tag{12.104}$$

Here,

$$\epsilon = \frac{I_\parallel - I_\perp}{I_\parallel} = 0.00335 \tag{12.105}$$

is the Earth's ellipticity, and

$$n_s = \frac{d\lambda_s}{dt} = \left(\frac{G\,M_s}{a_s^3}\right)^{1/2} \tag{12.106}$$

the Sun's apparent orbital angular velocity.

According to Section 8.9, the rotational kinetic energy of the Earth can be written

$$K = \frac{1}{2}\left(I_\perp\,\dot\theta^2 + I_\perp\,\sin^2\theta\,\dot\phi^2 + I_\parallel\,\omega^2\right), \tag{12.107}$$

where the Earth's angular velocity

$$\omega = \cos\theta\,\dot\phi + \dot\psi \tag{12.108}$$

is a constant of the motion. Here, ψ is the third Euler angle. Hence, the Earth's Lagrangian takes the form

$$\mathcal{L} = K - U = \frac{1}{2}\left(I_\perp\,\dot\theta^2 + I_\perp\,\sin^2\theta\,\dot\phi^2 + I_\parallel\,\omega^2\right) + \epsilon\,\alpha_s\,\cos(2\,\theta) \tag{12.109}$$

where any constant terms have been neglected. One equation of motion which can immediately be derived from this Lagrangian is

$$\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot\theta}\right) - \frac{\partial\mathcal{L}}{\partial\theta} = 0, \tag{12.110}$$

which reduces to

$$I_\perp\,\ddot{\theta} - \frac{\partial \mathcal{L}}{\partial \theta} = 0. \tag{12.111}$$

Consider *steady precession* of the Earth's rotational axis, which is characterized by $\dot{\theta} = 0$, with both $\dot{\phi}$ and $\dot{\psi}$ constant. It follows, from the above equation, that such motion must satisfy the constraint

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0. \tag{12.112}$$

Thus, we obtain

$$\frac{1}{2}\,I_\perp\,\sin(2\,\theta)\,\dot{\phi}^{\,2} - I_\parallel\,\sin\theta\,\omega\,\dot{\phi} - 2\,\epsilon\,\alpha_s\,\sin(2\,\theta) = 0, \tag{12.113}$$

where use has been made of Equations (12.108) and (12.109). Now, as can easily be verified after the fact, $|\dot{\phi}| \ll \omega$, so the above equation reduces to

$$\dot{\phi} \simeq -\frac{4\,\epsilon\,\alpha_s\,\cos\theta}{I_\parallel\,\omega} = \Omega_\phi, \tag{12.114}$$

which can be integrated to give

$$\phi \simeq -\Omega_\phi\,t, \tag{12.115}$$

where

$$\Omega_\phi = \frac{3}{2}\,\frac{\epsilon\,n_s^{\,2}}{\omega}\,\cos\theta, \tag{12.116}$$

and use has been made of Equation (12.104). According to the above expression, the mutual interaction between the Sun and the quadrupole gravitational field generated by the Earth's slight oblateness causes the Earth's axis of rotation to *precess* steadily about the normal to the ecliptic plane at the rate $-\Omega_\phi$. The fact that $-\Omega_\phi$ is negative implies that the precession is in the *opposite* direction to the direction of the Earth's rotation and the Sun's apparent orbit about the Earth. Incidentally, the interaction causes a precession of the Earth's rotational axis, rather than the plane of the Sun's orbit, because the Earth's axial moment of inertia is much less than the Sun's orbital moment of inertia. The precession period in years is given by

$$T_\phi(\mathrm{yr}) = \frac{n_s}{\Omega_\phi} = \frac{2\,T_s(\mathrm{day})}{3\,\epsilon\,\cos\theta}, \tag{12.117}$$

where $T_s(\mathrm{day}) = \omega/n_s = 365.24$ is the Sun's orbital period in days. Thus, given that $\epsilon = 0.00335$ and $\theta = 23.44^\circ$, we obtain

$$T_\phi \simeq 79,200 \text{ years}. \tag{12.118}$$

Unfortunately, the observed precession period of the Earth's axis of rotation about the normal to the ecliptic plane is approximately 25,800 years, so something is clearly missing from our model. It turns out that the missing factor is the influence of the *Moon*.

Using analogous arguments to those given above, the potential energy of the Earth-Moon system can be written

$$U = -\frac{G\,M_m\,M}{a_m} + \frac{G\,M_m\,(I_\parallel - I_\perp)}{a_m^3}\,P_2[\cos(\gamma_m)], \tag{12.119}$$

where M_m is the lunar mass, and a_m the radius of the Moon's (approximately circular) orbit. Furthermore, γ_m is the angle subtended between $\boldsymbol{\omega}$ and $\mathbf{r}_m$, where

$$\boldsymbol{\omega} = \omega\ (-\sin\theta\,\sin\phi,\,\sin\theta\,\cos\phi,\,\cos\theta) \tag{12.120}$$

is the Earth's angular velocity vector, and $\mathbf{r}_m$ is the position vector of the Moon relative to the Earth. Here, for the moment, we have retained the ϕ dependence in our expression for $\boldsymbol{\omega}$ (since we shall presently differentiate by ϕ, before setting $\phi = 0$). Now, the Moon's orbital plane is actually slightly inclined to the ecliptic plane, the angle of inclination being $\iota_m = 5.16^\circ$. Hence, we can write

$$\mathbf{r}_m \simeq a_m\ (\cos\lambda_m,\,\sin\lambda_m,\,\iota_m\,\sin(\lambda_m - \varpi_n))\,, \tag{12.121}$$

to first order in ι_m, where λ_m is the Moon's ecliptic longitude, and ϖ_n is the ecliptic longitude of the lunar *ascending node*, which is defined as the point on the lunar orbit where the Moon crosses the ecliptic plane from south to north. Of course, λ_m increases at the rate n_m, where

$$n_m = \frac{d\lambda_m}{dt} \simeq \left(\frac{G\,M}{a_m^3}\right)^{1/2} \tag{12.122}$$

is the Moon's orbital angular velocity. It turns out that the lunar ascending node precesses steadily, in the opposite direction to the Moon's orbital rotation, in such a manner that it completes a full circuit every 18.6 years. This precession is caused by the perturbing influence of the Sun—see Chapter 14. It follows that

$$\frac{d\varpi_n}{dt} = -\Omega_n, \tag{12.123}$$

where $2\pi/\Omega_n = 18.6$ years. Now, from (12.120) and (12.121),

$$\cos\gamma_m = \frac{\boldsymbol{\omega}\cdot\mathbf{r}_m}{|\boldsymbol{\omega}|\,|\mathbf{r}_m|} = \sin\theta\,\sin(\lambda_m - \phi) + \iota_m\,\cos\theta\,\sin(\lambda_m - \varpi_n), \tag{12.124}$$

so (12.119) yields

$$\begin{aligned} U \;\simeq\; & -\frac{G\,M_m\,M}{a_m} + \frac{G\,M_m\,(I_\parallel - I_\perp)}{2\,a_m^3}\Big[3\,\sin^2\theta\,\sin^2(\lambda_m - \phi) \\ & +3\,\iota_m\,\sin(2\,\theta)\,\sin(\lambda_m - \phi)\,\sin(\lambda_m - \varpi_n) - 1\Big] \end{aligned} \tag{12.125}$$

to first order in ι_m. Given that we are interested in the motion of the Earth's axis of rotation on time-scales that are much longer than a month, we can average the above expression over the Moon's orbit to give

$$U \simeq U_0' - \epsilon\,\alpha_m\,\cos(2\,\theta) + \epsilon\,\beta_m\,\sin(2\,\theta)\,\cos(\varpi_n - \phi), \tag{12.126}$$

[since the average of $\sin^2(\lambda_m-\phi)$ over a month is $1/2$, whereas that of $\sin(\lambda_m-\phi)\,\sin(\lambda_m-\varpi_m)$ is $(1/2)\,\cos(\varpi_m-\phi)$]. Here, U_0' is a constant,

$$\alpha_m = \frac{3}{8}\,I_\parallel\,\mu_m\,n_m^2, \tag{12.127}$$

$$\beta_m = \frac{3}{4}\,I_\parallel\,\iota_m\,\mu_m\,n_m^2, \tag{12.128}$$

and

$$\mu_m = \frac{M_m}{M} = 0.0123 \tag{12.129}$$

is the ratio of the lunar to the terrestrial mass. Now, gravity is a superposable force, so the total potential energy of the Earth-Moon-Sun system is the sum of Equations (12.103) and (12.126). In other words,

$$U = U_0'' - \epsilon\,\alpha\,\cos(2\,\theta) + \epsilon\,\beta_m\,\sin(2\,\theta)\,\cos(\varpi_n-\phi), \tag{12.130}$$

where U_0'' is a constant, and

$$\alpha = \alpha_s + \alpha_m. \tag{12.131}$$

Finally, making use of (12.107), the Lagrangian of the Earth is written

$$\mathcal{L} = \frac{1}{2}\left(I_\perp\,\dot\theta^2 + I_\perp\,\sin^2\theta\,\dot\phi^2 + I_\parallel\,\omega^2\right) + \epsilon\,\alpha\,\cos(2\,\theta) - \epsilon\,\beta_m\,\sin(2\,\theta)\,\cos(\varpi_n-\phi), \tag{12.132}$$

where any constant terms have been neglected. Recall that ω is given by (12.108), and is a constant of the motion.

Two equations of motion that can immediately be derived from the above Lagrangian are

$$\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot\theta}\right) - \frac{\partial\mathcal{L}}{\partial\theta} = 0, \tag{12.133}$$

$$\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot\phi}\right) - \frac{\partial\mathcal{L}}{\partial\phi} = 0. \tag{12.134}$$

(The third equation, involving ψ, merely confirms that ω is a constant of the motion.) The above two equations yield

$$\begin{aligned} 0 &= I_\perp\,\ddot\theta - \frac{1}{2}\,I_\perp\,\sin(2\,\theta)\,\dot\phi^{\,2} + I_\parallel\,\sin\theta\,\omega\,\dot\phi + 2\,\epsilon\,\alpha\,\sin(2\,\theta) \\ &\quad + 2\,\epsilon\,\beta_m\,\cos(2\,\theta)\,\cos(\varpi_n-\phi), \end{aligned} \tag{12.135}$$

$$0 = \frac{d}{dt}\left(I_\perp\,\sin^2\theta\,\dot\phi + I_\parallel\,\cos\theta\,\omega\right) + \epsilon\,\beta_m\,\sin(2\,\theta)\,\sin(\varpi_n-\phi), \tag{12.136}$$

respectively. Let

$$\theta(t) = \theta_0 + \epsilon\,\theta_1(t), \tag{12.137}$$

$$\phi(t) = \epsilon\,\phi_1(t), \tag{12.138}$$

where $\theta_0 = 23.44^\circ$ is the mean inclination of the ecliptic to the Earth's equatorial plane. To first order in ϵ, Equations (12.135) and (12.136) reduce to

$$0 \simeq I_\perp \ddot{\theta}_1 + I_\parallel \sin\theta_0\, \omega\, \dot{\phi}_1 + 2\,\alpha\, \sin(2\,\theta_0) + 2\,\beta_m\, \cos(2\,\theta_0)\, \cos(\Omega_n\, t), \quad (12.139)$$

$$0 \simeq I_\perp \sin^2\theta_0\, \ddot{\phi}_1 - I_\parallel \sin\theta_0\, \omega\, \dot{\theta}_1 - \beta_m\, \sin(2\,\theta_0)\, \sin(\Omega_n\, t), \quad (12.140)$$

respectively, where use has been made of Equation (12.123). However, as can easily be verified after the fact, $d/dt \ll \omega$, so we obtain

$$\dot{\phi}_1 \simeq -\frac{4\,\alpha\, \cos\theta_0}{I_\parallel\, \omega} - \frac{2\,\beta_m\, \cos(2\,\theta_0)}{I_\parallel\, \omega\, \sin\theta_0}\, \cos(\Omega_n\, t), \quad (12.141)$$

$$\dot{\theta}_1 \simeq -\frac{2\,\beta_m\, \cos\theta_0}{I_\parallel\, \omega}\, \sin(\Omega_n\, t). \quad (12.142)$$

The above equations can be integrated, and then combined with Equations (12.137) and (12.138), to give

$$\phi(t) = -\Omega_\phi\, t - \delta\phi\, \sin(\Omega_n\, t), \quad (12.143)$$

$$\theta(t) = \theta_0 + \delta\theta\, \cos(\Omega_n\, t), \quad (12.144)$$

where

$$\Omega_\phi = \frac{3}{2}\, \frac{\epsilon\, (n_s^2 + \mu_m\, n_m^2)}{\omega}\, \cos\theta_0, \quad (12.145)$$

$$\delta\phi = \frac{3}{2}\, \frac{\epsilon\, \iota_m\, \mu_m\, n_m^2}{\omega\, \Omega_n}\, \frac{\cos(2\,\theta_0)}{\sin\theta_0}, \quad (12.146)$$

$$\delta\theta = \frac{3}{2}\, \frac{\epsilon\, \iota_m\, \mu_m\, n_m^2}{\omega\, \Omega_n}\, \cos\theta_0. \quad (12.147)$$

Incidentally, in the above, we have assumed that the lunar ascending node coincides with the vernal equinox at time $t = 0$ (*i.e.*, $\varpi_m = 0$ at $t = 0$), in accordance with our previous assumption that $\phi = 0$ at $t = 0$.

According to Equation (12.143), the combined gravitational interaction of the Sun and the Moon with the quadrupole field generated by the Earth's slight oblateness causes the Earth's axis of rotation to precess steadily about the normal to the ecliptic plane at the rate $-\Omega_\phi$. As before, the negative sign indicates that the precession is in the opposite direction to the (apparent) orbital motion of the sun and moon. The period of the precession in years is given by

$$T_\phi(\mathrm{yr}) = \frac{n_s}{\Omega_\phi} = \frac{2\, T_s(\mathrm{day})}{\epsilon\, (1 + \mu_m/[T_m(\mathrm{yr})]^2)\, \cos\theta_0}, \quad (12.148)$$

where $T_m(\mathrm{yr}) = n_s/n_m = 0.081$ is the Moon's (synodic) orbital period in years. Given that $\epsilon = 0.00335$, $\theta_0 = 23.44^\circ$, $T_s(\mathrm{day}) = 365.24$, and $\mu_m = 0.0123$, we obtain

$$T_\phi \simeq 27,600\,\mathrm{years}. \quad (12.149)$$

This prediction is fairly close to the observed precession period of $25,800$ years. The main reason that our estimate is slightly inaccurate is because we have neglected to take into account the small eccentricities of the Earth's orbit around the Sun, and the Moon's orbit around the Earth.

The point in the sky toward which the Earth's axis of rotation points is known as the *north celestial pole*. Currently, this point lies within about a degree of the fairly bright star *Polaris*, which is consequently sometimes known as the *north star* or the *pole star*. It follows that Polaris appears to be almost stationary in the sky, always lying *due north*, and can thus be used for navigational purposes. Indeed, mariners have relied on the north star for many hundreds of years to determine direction at sea. Unfortunately, because of the precession of the Earth's axis of rotation, the north celestial pole is not a fixed point in the sky, but instead traces out a circle, of angular radius 23.44°, about the north ecliptic pole, with a period of 25,800 years. Hence, a few thousand years from now, the north celestial pole will no longer coincide with Polaris, and there will be no convenient way of telling direction from the stars.

The projection of the ecliptic plane onto the sky is called the *ecliptic*, and coincides with the apparent path of the Sun against the backdrop of the stars. Furthermore, the projection of the Earth's equator onto the sky is known as the *celestial equator*. As has been previously mentioned, the ecliptic is inclined at 23.44° to the celestial equator. The two points in the sky at which the ecliptic crosses the celestial equator are called the *equinoxes*, since night and day are equally long when the Sun lies at these points. Thus, the Sun reaches the *vernal equinox* on about March 21st, and this traditionally marks the beginning of spring. Likewise, the Sun reaches the *autumn equinox* on about September 22nd, and this traditionally marks the beginning of autumn. However, the precession of the Earth's axis of rotation causes the celestial equator (which is always normal to this axis) to precess in the sky, and thus also causes the equinoxes to precess along the ecliptic. This effect is known as the *precession of the equinoxes*. The precession is in the opposite direction to the Sun's apparent motion around the ecliptic, and is of magnitude 1.4° per century. Amazingly, this miniscule effect was discovered by the Ancient Greeks (with the help of ancient Babylonian observations). In about 2000 BC, when the science of astronomy originated in ancient Egypt and Babylonia, the vernal equinox lay in the constellation Aries. Indeed, the vernal equinox is still sometimes called the *first point of Aries* in astronomical texts. About 90 BC, the vernal equinox moved into the constellation Pisces, where it still remains. The equinox will move into the constellation Aquarius (marking the beginning of the much heralded "Age of Aquarius") in about 2600 AD. Incidentally, the position of the vernal equinox in the sky is of great significance in astronomy, since it is used as the zero of celestial longitude (much as Greenwich is used as the zero of terrestrial longitude).

Equations (12.143) and (12.144) indicate that the small inclination of the lunar orbit to the ecliptic plane, combined with the precession of the lunar ascending node, causes the Earth's axis of rotation to wobble sightly. This wobble is known as *nutation*, and is superimposed on the aforementioned precession. In the absence of precession, nutation would cause the north celestial pole to periodically trace out a small ellipse on the sky, the

sense of rotation being *counter-clockwise*. The nutation period is 18.6 years: *i.e.*, the same as the precession period of the lunar ascending node. The nutation amplitudes in the polar and azimuthal angles θ and ϕ are

$$\delta\theta = \frac{3}{2}\,\frac{\epsilon\,\iota_m\,\mu_m\,T_n(\mathrm{yr})}{T_s(\mathrm{day})\,[T_m(\mathrm{yr})]^2}\,\cos\theta_0, \tag{12.150}$$

$$\delta\phi = \frac{3}{2}\,\frac{\epsilon\,\iota_m\,\mu_m\,T_n(\mathrm{yr})}{T_s(\mathrm{day})\,[T_m(\mathrm{yr})]^2}\,\frac{\cos(2\,\theta_0)}{\sin\theta_0}, \tag{12.151}$$

respectively, where $T_n(\mathrm{yr}) = n_s/\Omega_n = 18.6$. Given that $\epsilon = 0.00335$, $\theta_0 = 23.44^\circ$, $\iota_m = 5.16^\circ$, $T_s(\mathrm{day}) = 365.24$, $T_m(\mathrm{yr}) = 0.081$, and $\mu_m = 0.0123$, we obtain

$$\delta\theta = 8.2'', \tag{12.152}$$

$$\delta\phi = 15.3''. \tag{12.153}$$

The observed nutation amplitudes are $9.2''$ and $17.2''$, respectively. Hence, our estimates are quite close to the mark. Any inaccuracy is mainly due to the fact that we have neglected to take into account the small eccentricities of the Earth's orbit around the Sun, and the Moon's orbit around the Earth. The nutation of the Earth was discovered in 1728 by the English astronomer James Bradley, and was explained theoretically about 20 years later by d'Alembert and L. Euler. Nutation is important because the corresponding gyration of the Earth's rotation axis appears to be transferred to celestial objects when they are viewed using terrestrial telescopes. This effect causes the celestial longitudes and latitudes of heavenly objects to oscillate sinusoidally by up to $20''$ (*i.e.*, about the maximum angular size of Saturn) with a period of 18.6 years. It is necessary to correct for this oscillation in order to accurately guide terrestrial telescopes to particular objects.

Note, finally, that the type of forced nutation discussed above, which is driven by an external torque, is quite distinct from the free nutation described in Section 8.9.

12.11 Potential Due to a Uniform Ring

Consider a uniform ring of mass M and radius r, centered on the origin, and lying in the x-y plane. Let us consider the gravitational potential $\Phi(r)$ generated by such a ring in the x-y plane (which corresponds to $\theta = 90^\circ$). It follows, from Section 12.3, that for $r > a$,

$$\Phi(r) = -\frac{G\,M}{a}\sum_{n=0,\infty}[P_n(0)]^2\left(\frac{a}{r}\right)^{n+1}. \tag{12.154}$$

However, $P_0(0) = 1$, $P_1(0) = 0$, $P_2(0) = -1/2$, $P_3(0) = 0$, and $P_4(0) = 3/8$. Hence,

$$\Phi(r) = -\frac{G\,M}{r}\left[1 + \frac{1}{4}\left(\frac{a}{r}\right)^2 + \frac{9}{64}\left(\frac{a}{r}\right)^4 + \cdots\right]. \tag{12.155}$$

Likewise, for $r < a$,

$$\Phi(r) = -\frac{G\,M}{a}\sum_{n=0,\infty}[P_n(0)]^2\left(\frac{r}{a}\right)^n, \tag{12.156}$$

giving

$$\Phi(r) = -\frac{G\,M}{a}\left[1+\frac{1}{4}\left(\frac{r}{a}\right)^2+\frac{9}{64}\left(\frac{r}{a}\right)^4+\cdots\right]. \tag{12.157}$$

12.12 Perihelion Precession of the Planets

The Solar System consists of eight major planets (Mercury to Neptune) moving around the Sun in slightly elliptical orbits which are approximately co-planar with one another. According to Chapter 5, if we neglect the relatively weak interplanetary gravitational interactions then the perihelia of the various planets (*i.e.*, the points on their orbits at which they are closest to the Sun) remain *fixed* in space. However, once these interactions are taken into account, it turns out that the planetary perihelia all slowly *precess* around the Sun. We can calculate the approximate rate of perihelion precession of a given planet by treating the other planets as *uniform concentric rings*, centered on the Sun, of mass equal to the planetary mass, and radius equal to the mean orbital radius.[1] This is equivalent to averaging the interplanetary gravitational interactions over the orbits of the other planets. It is reasonable to do this, since the precession period in question is very much longer than the orbital period of any planet in the Solar System. Thus, by treating the other planets as rings, we can calculate the mean gravitational perturbation due to these planets, and, thereby, determine the desired precession rate.

We can conveniently index the planets in the Solar System such that Mercury is planet 1, and Neptune planet 8. Let the M_i and the R_i, for $i = 1, 8$, be the planetary masses and orbital radii, respectively. Furthermore, let M_0 be the mass of the Sun. It follows, from the previous section, that the gravitational potential generated at the ith planet by the Sun and the other planets is

$$\begin{aligned}\Phi(R_i) &= -\frac{G\,M_0}{R_i} - G\sum_{j<i}\frac{M_j}{R_i}\left[1+\frac{1}{4}\left(\frac{R_j}{R_i}\right)^2+\frac{9}{64}\left(\frac{R_j}{R_i}\right)^4+\cdots\right]\\ &\quad -G\sum_{j>i}\frac{M_j}{R_j}\left[1+\frac{1}{4}\left(\frac{R_i}{R_j}\right)^2+\frac{9}{64}\left(\frac{R_i}{R_j}\right)^4+\cdots\right].\end{aligned} \tag{12.158}$$

Now, the radial force per unit mass acting on the ith planet is written $f(R_i) = -d\Phi(R_i)/dr$, giving

$$f(R_i) = -\frac{G\,M_0}{R_i^2} - \frac{G}{R_i^2}\sum_{j<i}M_j\left[1+\frac{3}{4}\left(\frac{R_j}{R_i}\right)^2+\frac{45}{64}\left(\frac{R_j}{R_i}\right)^4+\cdots\right]$$

[1] M.G. Stewart, American Jou. Physics **73**, 730 (2005).

$$+\frac{G}{R_i^2}\sum_{j>i} M_j \left(\frac{R_i}{R_j}\right)\left[\frac{1}{2}\left(\frac{R_i}{R_j}\right)^2+\frac{9}{16}\left(\frac{R_i}{R_j}\right)^4+\cdots\right]. \tag{12.159}$$

Hence, we obtain

$$\begin{aligned} R_i\, f'(R_i) &= \frac{2\,G\,M_0}{R_i^2}+\frac{G}{R_i^2}\sum_{j<i} M_j\left[2+3\left(\frac{R_j}{R_i}\right)^2+\frac{135}{32}\left(\frac{R_j}{R_i}\right)^4+\cdots\right] \\ &\quad +\frac{G}{R_i^2}\sum_{j>i} M_j\left(\frac{R_i}{R_j}\right)\left[\frac{1}{2}\left(\frac{R_i}{R_j}\right)^2+\frac{27}{16}\left(\frac{R_i}{R_j}\right)^4+\cdots\right], \end{aligned} \tag{12.160}$$

where $' \equiv d/dr$. It follows that

$$\begin{aligned} \left[3+\frac{R_i\, f'(R_i)}{f(R_i)}\right]^{-1/2} &= 1+\frac{3}{4}\sum_{j<i}\left(\frac{M_j}{M_0}\right)\left(\frac{R_j}{R_i}\right)^2\left[1+\frac{15}{8}\left(\frac{R_j}{R_i}\right)^2+\frac{175}{64}\left(\frac{R_j}{R_i}\right)^4+\cdots\right] \\ &\quad +\frac{3}{4}\sum_{j>i}\left(\frac{M_j}{M_0}\right)\left(\frac{R_i}{R_j}\right)^3\left[1+\frac{15}{8}\left(\frac{R_i}{R_j}\right)^2+\frac{175}{64}\left(\frac{R_i}{R_j}\right)^4+\cdots\right]. \end{aligned} \tag{12.161}$$

Thus, according to Equation (5.108), the apsidal angle for the ith planet is

$$\begin{aligned} \psi_i &= \pi\left\{1+\frac{3}{4}\sum_{j<i}\left(\frac{M_j}{M_0}\right)\left(\frac{R_j}{R_i}\right)^2\left[1+\frac{15}{8}\left(\frac{R_j}{R_i}\right)^2+\frac{175}{64}\left(\frac{R_j}{R_i}\right)^4+\cdots\right]\right. \\ &\quad \left.+\frac{3}{4}\sum_{j>i}\left(\frac{M_j}{M_0}\right)\left(\frac{R_i}{R_j}\right)^3\left[1+\frac{15}{8}\left(\frac{R_i}{R_j}\right)^2+\frac{175}{64}\left(\frac{R_i}{R_j}\right)^4+\cdots\right]\right\}. \end{aligned} \tag{12.162}$$

Hence, the perihelion of the ith planet advances by

$$\begin{aligned} \delta\psi_i &= \frac{3\pi}{2}\sum_{j<i}\left(\frac{M_j}{M_0}\right)\left(\frac{R_j}{R_i}\right)^2\left[1+\frac{15}{8}\left(\frac{R_j}{R_i}\right)^2+\frac{175}{64}\left(\frac{R_j}{R_i}\right)^4+\cdots\right] \\ &\quad +\frac{3\pi}{2}\sum_{j>i}\left(\frac{M_j}{M_0}\right)\left(\frac{R_i}{R_j}\right)^3\left[1+\frac{15}{8}\left(\frac{R_i}{R_j}\right)^2+\frac{175}{64}\left(\frac{R_i}{R_j}\right)^4+\cdots\right] \end{aligned} \tag{12.163}$$

radians per revolution around the Sun. Now, the time for one revolution is $T_i = 2\pi/\omega_i$, where $\omega_i^2 = G\,M_0/R_i^3$. Thus, the rate of perihelion precession, in *arc seconds per year*, is given by

$$\begin{aligned} \delta\dot{\psi}_i &= \frac{75}{T_i(\mathrm{yr})}\left\{\sum_{j<i}\left(\frac{M_j}{M_0}\right)\left(\frac{R_j}{R_i}\right)^2\left[1+\frac{15}{8}\left(\frac{R_j}{R_i}\right)^2+\frac{175}{64}\left(\frac{R_j}{R_i}\right)^4+\cdots\right]\right. \\ &\quad \left.+\sum_{j>i}\left(\frac{M_j}{M_0}\right)\left(\frac{R_i}{R_j}\right)^3\left[1+\frac{15}{8}\left(\frac{R_i}{R_j}\right)^2+\frac{175}{64}\left(\frac{R_i}{R_j}\right)^4+\cdots\right]\right\}. \end{aligned} \tag{12.164}$$

Planet	M/M_0	T(yr)	R(au)
Mercury	1.66×10^{-7}	0.241	0.387
Venus	2.45×10^{-6}	0.615	0.723
Earth	3.04×10^{-6}	1.000	1.00
Mars	3.23×10^{-7}	1.881	1.52
Jupiter	9.55×10^{-4}	11.86	5.20
Saturn	2.86×10^{-4}	29.46	9.54
Uranus	4.36×10^{-5}	84.01	19.19
Neptune	5.18×10^{-5}	164.8	30.07

Table 12.1: *Data for the major planets in the Solar System, giving the planetary mass relative to that of the Sun, the orbital period in years, and the mean orbital radius relative to that of the Earth.*

Planet	$(\delta\dot{\Psi})_{obs}$	$(\delta\dot{\Psi})_{th}$
Mercury	5.75	5.50
Venus	2.04	10.75
Earth	11.45	11.87
Mars	16.28	17.60
Jupiter	6.55	7.42
Saturn	19.50	18.36
Uranus	3.34	2.72
Neptune	0.36	0.65

Table 12.2: *The observed perihelion precession rates of the planets compared with the theoretical precession rates calculated from Equation (12.164) and Table 12.1. The precession rates are in arc seconds per year.*

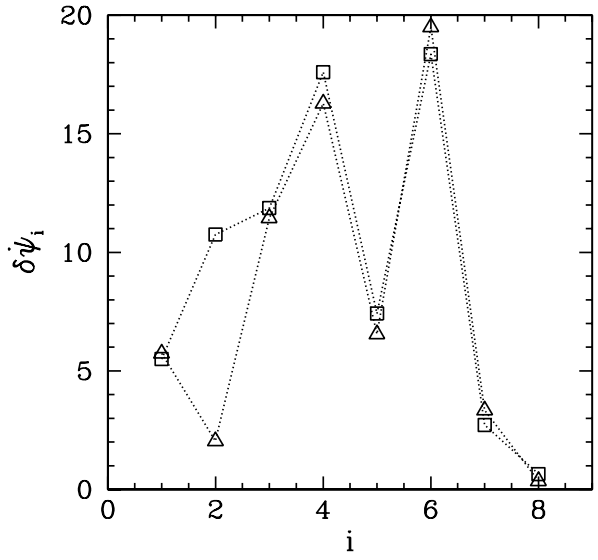


Figure 12.8: *The triangular points show the observed perihelion precession rates of the major planets in the Solar System, whereas the square points show the theoretical rates calculated from Equation (12.164) and Table 12.1. The precession rates are in arc seconds per year.*

Table 12.2 and Figure 12.8 compare the observed perihelion precession rates with the theoretical rates calculated from Equation (12.164) and the planetary data given in Table 12.1. It can be seen that there is excellent agreement between the two, except for the planet Venus. The main reason for this is that Venus has an unusually low eccentricity ($e = 0.0068$), which renders its perihelion point extremely sensitive to small perturbations.

12.13 Perihelion Precession of Mercury

If the calculation described in the previous section is carried out more accurately, taking into account the slight eccentricities of the planetary orbits, as well as their small mutual inclinations, and retaining many more terms in the expansions (12.155) and (12.157), then the perihelion precession rate of the planet Mercury is found to be 5.32 arc seconds per year. However, the observed precession rate is 5.75 arc seconds per year. It turns out that the cause of this discrepancy is the general relativistic correction to Newtonian gravity.

General relativity gives rise to a small correction to the force per unit mass exerted by the Sun (mass M_0) on a planet in a circular orbit of radius r, and angular momentum per unit mass h. In fact, the modified formula for f is

$$f \simeq -\frac{G\,M_0}{r^2} - \frac{3\,G\,M_0\,h^2}{c^2\,r^4}, \tag{12.165}$$

where c is the velocity of light in vacuum. It follows that

$$\frac{r\,f'}{f} = -2\left(1 + \frac{3\,h^2}{c^2\,r^2} + \cdots\right). \tag{12.166}$$

Hence, from Equation (5.108), the apsidal angle is

$$\psi \simeq \pi\left(1 + \frac{3\,h^2}{c^2\,r^2}\right). \tag{12.167}$$

Thus, the perihelion advances by

$$\delta\psi \simeq \frac{6\pi\,G\,M_0}{c^2\,r} \tag{12.168}$$

radians per revolution due to the general relativistic correction to Newtonian gravity. Here, use has been made of $h^2 = G\,M_0\,r$. It follows that the rate of perihelion precession due to the general relativistic correction is

$$\delta\dot{\psi} \simeq \frac{0.0383}{R\,T} \tag{12.169}$$

arc seconds per year, where R is the mean orbital radius in mean Earth orbital radii, and T is the orbital period in years. Hence, from Table 12.1, the general relativistic contribution to $\delta\dot{\psi}$ for Mercury is 0.41 arc seconds per year. It is easily demonstrated that the corresponding contribution is negligible for the other planets in the Solar System. If the above calculation is carried out sightly more accurately, taking the eccentricity of Mercury's orbit into account, then the general relativistic contribution to $\delta\dot{\psi}$ becomes 0.43 arc seconds per year. It follows that the total perihelion precession rate for Mercury is $5.32 + 0.43 = 5.75$ arc seconds per year. This is in exact agreement with the observed precession rate. Indeed, the ability of general relativity to explain the discrepancy between the observed perihelion precession rate of Mercury, and that calculated from Newtonian dynamics, was one of the first major successes of this theory.

12.14 Exercises

12.1. Show that

$$\epsilon = \frac{5-\alpha}{4}\,\frac{\Omega^2\,a^3}{G\,M}.$$

for a self-gravitating, rotating spheroid of ellipticity $\epsilon \ll 1$, mass M, mean radius a, and angular velocity Ω whose mass density varies as $r^{-\alpha}$ (where $\alpha < 3$). Demonstrate that the above formula matches the observed rotational flattening of the Earth when $\alpha = 1.09$ and of Jupiter when $\alpha = 1.79$.

12.2. The Moon's orbital period about the Earth is approximately 27.3 days, and is in the same direction as the Earth's axial rotation (whose period is 24 hours). Use this data to show that high tides at a given point on the Earth occur every 12 hours and 26 minutes.

12.3. Estimate the tidal elongation of the Moon due to the Earth.

12.4. Consider an artificial satellite in a circular orbit of radius L about the Earth. Suppose that the normal to the plane of the orbit subtends an angle θ with the Earth's axis of rotation. By approximating the orbiting satellite as a uniform ring, demonstrate that the Earth's oblateness causes the plane of the satellite's orbit to precess about the Earth's rotational axis at the rate

$$\frac{\dot{\phi}}{\omega} \simeq -\frac{1}{2}\,\epsilon\left(\frac{R}{L}\right)^2 \cos\theta.$$

Here, ω is the satellite's orbital angular velocity, $\epsilon = 0.00335$ the Earth's ellipticity, and R the Earth's radius. Note that the Earth's axial moment of inertial is $I_{\parallel} \simeq (1/3)\,M\,R^2$, where M is the mass of the Earth.

12.5. A *sun-synchronous* satellite is one which always passes over a given point on the Earth at the same local solar time. This is achieved by fixing the precession rate of the satellite's orbital plane such that it matches the rate at which the Sun appears to move against the background of the stars. What orbital altitude above the surface of the Earth would such a satellite need to have in order to fly over all latitudes between 50° N and 50° S? Is the direction of the satellite orbit in the same sense as the Earth's rotation (prograde), or the opposite sense (retrograde)?

13 The Three-Body Problem

13.1 Introduction

We saw earlier, in Section 6.2, that an isolated dynamical system consisting of two freely moving point masses exerting forces on one another—which is usually referred to as a *two-body problem*—can always be converted into an equivalent one-body problem. In particular, this implies that we can *exactly solve* a dynamical system containing *two* gravitationally interacting point masses, since the equivalent one-body problem is exactly soluble—see Chapter 5 and Section 6.3. What about a system containing *three* gravitationally interacting point masses? Despite hundreds of years of research, no exact solution of this famous problem—which is generally known as the *three-body problem*—has ever been found. It is, however, possible to make some progress by severely restricting the problem's scope.

13.2 Circular Restricted Three-Body Problem

Consider an isolated dynamical system consisting of three gravitationally interacting point masses, m_1, m_2, and m_3. Suppose, however, that the third mass, m_3, is so much smaller than the other two that it has a negligible effect on their motion. Suppose, further, that the first two masses, m_1 and m_2, execute circular orbits about their common center of mass. In the following, we shall investigate this simplified problem, which is generally known as the *circular restricted three-body problem*.

Let us define a Cartesian coordinate system (ξ, η, ζ) in an inertial reference frame whose origin coincides with the center of mass, C, of the two orbiting masses. Furthermore, let the orbital plane of these masses coincide with the ξ-η plane, and let them both lie on the ξ-axis at time $t = 0$—see Figure 13.1. Suppose that R is the constant distance between the two orbiting masses, r_1 the constant distance between mass m_1 and the origin, and r_2 the constant distance between mass m_2 and the origin. Moreover, let ω be the constant orbital angular velocity. It follows, from Section 6.3, that

$$\omega^2 = \frac{G M}{R^3}, \tag{13.1}$$

$$\frac{r_1}{r_2} = \frac{m_2}{m_1}, \tag{13.2}$$

where $M = m_1 + m_2$.

It is convenient to choose our unit of length such that $R = 1$, and our unit of mass such that $G M = 1$. It follows, from Equation (13.1), that $\omega = 1$. However, we shall continue to retain ω in our equations, for the sake of clarity. Let $\mu_1 = G m_1$, and $\mu_2 = G m_2 = 1 - \mu_1$. It is easily demonstrated that $r_1 = \mu_2$, and $r_2 = 1 - r_1 = \mu_1$. Hence, the two orbiting masses,

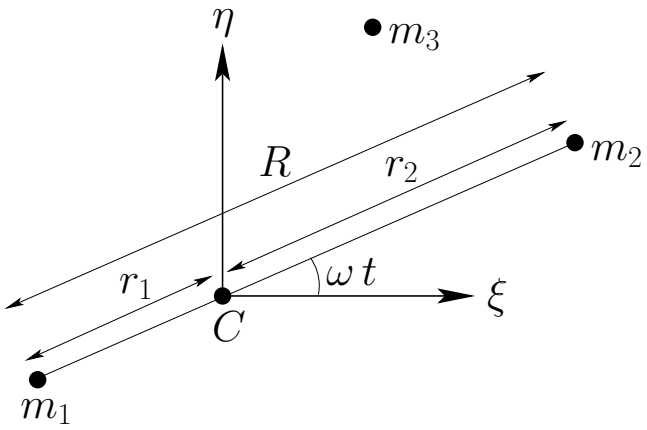


Figure 13.1: *The circular restricted three-body problem.*

m_1 and m_2, have position vectors $\mathbf{r}_1 = (\xi_1, \eta_1, 0)$ and $\mathbf{r}_2 = (\xi_2, \eta_2, 0)$, respectively, where (see Figure 13.1)

$$\mathbf{r}_1 = \mu_2\,(-\cos\omega t, -\sin\omega t, 0), \tag{13.3}$$

$$\mathbf{r}_2 = \mu_1\,(\cos\omega t, \sin\omega t, 0). \tag{13.4}$$

Let the third mass have position vector $\mathbf{r} = (\xi, \eta, \zeta)$. The Cartesian components of the equation of motion of this mass are thus

$$\ddot{\xi} = -\mu_1\,\frac{(\xi-\xi_1)}{\rho_1^3} - \mu_2\,\frac{(\xi-\xi_2)}{\rho_2^3}, \tag{13.5}$$

$$\ddot{\eta} = -\mu_1\,\frac{(\eta-\eta_1)}{\rho_1^3} - \mu_2\,\frac{(\eta-\eta_2)}{\rho_2^3}, \tag{13.6}$$

$$\ddot{\zeta} = -\mu_1\,\frac{\zeta}{\rho_1^3} - \mu_2\,\frac{\zeta}{\rho_2^3}, \tag{13.7}$$

where

$$\rho_1^2 = (\xi-\xi_1)^2 + (\eta-\eta_1)^2 + \zeta^2, \tag{13.8}$$

$$\rho_2^2 = (\xi-\xi_2)^2 + (\eta-\eta_2)^2 + \zeta^2. \tag{13.9}$$

13.3 Jacobi Integral

Consider the function

$$C = 2\left(\frac{\mu_1}{\rho_1} + \frac{\mu_2}{\rho_2}\right) + 2\,\omega\,(\xi\,\dot{\eta} - \eta\,\dot{\xi}) - \dot{\xi}^2 - \dot{\eta}^2 - \dot{\zeta}^2. \tag{13.10}$$

The time derivative of this function is written

$$\dot{C} = -\frac{2\,\mu_1\,\dot{\rho}_1}{\rho_1^2} - \frac{2\,\mu_2\,\dot{\rho}_2}{\rho_2^2} + 2\,\omega\,(\xi\,\ddot{\eta} - \eta\,\ddot{\xi}) - 2\,\dot{\xi}\,\ddot{\xi} - 2\,\dot{\eta}\,\ddot{\eta} - 2\,\dot{\zeta}\,\ddot{\zeta}. \tag{13.11}$$

Moreover, it follows, from Equations (13.3)–(13.4) and (13.8)–(13.9), that

$$\rho_1\,\dot{\rho}_1 = -(\xi_1\,\dot{\xi} + \eta_1\,\dot{\eta}) + \omega\,(\xi\,\eta_1 - \eta\,\xi_1) + \xi\,\dot{\xi} + \eta\,\dot{\eta} + \zeta\,\dot{\zeta}, \tag{13.12}$$

$$\rho_2\,\dot{\rho}_2 = -(\xi_2\,\dot{\xi} + \eta_2\,\dot{\eta}) + \omega\,(\xi\,\eta_2 - \eta\,\xi_2) + \xi\,\dot{\xi} + \eta\,\dot{\eta} + \zeta\,\dot{\zeta}. \tag{13.13}$$

Combining Equations (13.5)–(13.7) with the above three expressions, we obtain (after considerable algebra)

$$\dot{C} = 0. \tag{13.14}$$

In other words, the function C—which is usually referred to as the *Jacobi integral*—is a *constant of the motion*.

Now, we can rearrange Equation (13.10) to give

$$\mathcal{E} \equiv \frac{1}{2}\,(\dot{\xi}^2 + \dot{\eta}^2 + \dot{\zeta}^2) - \left(\frac{\mu_1}{\rho_1} + \frac{\mu_2}{\rho_2}\right) = \boldsymbol{\omega}\cdot\mathbf{h} - \frac{C}{2}, \tag{13.15}$$

where $\mathcal{E}$ is the energy (per unit mass) of mass m_3, $\mathbf{h} = \mathbf{r}\times\dot{\mathbf{r}}$ its angular momentum (per unit mass), and $\boldsymbol{\omega} = (0,\,0,\,\omega)$ the orbital angular velocity of the other two masses. Note, however, that $\mathbf{h}$ is *not* a constant of the motion. Hence, $\mathcal{E}$ is not a constant of the motion either. In fact, the Jacobi integral is the *only* constant of the motion in the circular restricted three-body problem. Incidentally, the energy of mass m_3 is not a conserved quantity because the other two masses in the system are *moving*.

13.4 Tisserand Criterion

Consider a dynamical system consisting of three gravitationally interacting point masses, m_1, m_2, and m_3. Let mass m_1 represent the Sun, mass m_2 the planet Jupiter, and mass m_3 a comet. Since the mass of a comet is very much less than that of the Sun or Jupiter, and the Sun and Jupiter are in (almost) circular orbits about their common center of mass, the dynamical system in question satisfies all of the necessary criteria to be considered an example of a restricted three-body problem.

Now, the mass of the Sun is much greater than that of Jupiter. It follows that the gravitational effect of Jupiter on the cometary orbit is *negligible* unless the comet makes a very *close approach* to Jupiter. Hence, as described in Chapter 5, before and after such an approach, the comet executes a standard elliptical orbit about the Sun with fixed orbital parameters: *i.e.*, fixed major radius, eccentricity, and inclination to the ecliptic plane. However, in general, the orbital parameters before and after the close approach will *not* be the same as one another. Let us investigate further.

Now, since $m_1 \gg m_2$, we have $\mu_1 = G\,m_1 \simeq G\,(m_1 + m_2) = 1$, and $\rho_1 \simeq r$. Hence, according to Equations (5.51) and (5.60), the (approximately) conserved energy (per unit mass) of the comet before and after its close approach to Jupiter is

$$\mathcal{E} \equiv \frac{1}{2}\left(\dot{\xi}^{\,2} + \dot{\eta}^{\,2} + \dot{\zeta}^{\,2}\right) - \frac{1}{r} = -\frac{1}{2\,a}. \tag{13.16}$$

Note that the comet's orbital energy is entirely determined by its major radius, a. (Incidentally, we are working in units such that the major radius of Jupiter's orbit is unity.) Furthermore, the (approximately) conserved angular momentum (per unit mass) of the comet before and after its approach to Jupiter is written $\mathbf{h}$, where $\mathbf{h}$ is directed *normal* to the comet's orbital plane, and, from Equations (5.27) and (5.47),

$$h^2 = a\,(1 - e^2). \tag{13.17}$$

Here, e is the comet's orbital eccentricity. It follows that

$$\boldsymbol{\omega}\cdot\mathbf{h} = \omega\,h\,\cos I = \sqrt{a\,(1-e^2)}\,\cos I, \tag{13.18}$$

since $\omega = 1$ in our adopted system of units. Here, I is the angle of inclination of the normal to the comet's orbital plane to that of Jupiter's orbital plane.

Let a, e, and I be the major radius, eccentricity, and inclination angle of the cometary orbit before the close encounter with Jupiter, and let a', e', and I' be the corresponding parameters after the encounter. It follows from Equations (13.15), (13.16), and (13.18), and the fact that C is conserved during the encounter, whereas $\mathcal{E}$ and h are not, that

$$\frac{1}{2\,a} + \sqrt{a\,(1-e^2)}\,\cos I = \frac{1}{2\,a'} + \sqrt{a'\,(1-e'^{\,2})}\,\cos I'. \tag{13.19}$$

This result is known as the *Tisserand criterion*, and restricts the possible changes in the orbital parameters of a comet due to a close encounter with Jupiter (or any other massive planet).

The Tisserand criterion is very useful. For instance, whenever a new comet is discovered, astronomers immediately calculate its *Tisserand parameter*,

$$T_J = \frac{1}{a} + 2\,\sqrt{a\,(1-e^2)}\,\cos I. \tag{13.20}$$

If this parameter has the same value as that of a previously observed comet then it is quite likely that the new comet is, in fact, the same comet, but that its orbital parameters have changed since it was last observed, due to a close encounter with Jupiter. Incidentally, the subscript J in the above formula is to remind us that we are dealing with the Tisserand parameter for close encounters with *Jupiter*. (The parameter is, thus, evaluated in a system of units in which the major radius of Jupiter's orbit is unity). Obviously, it is also possible to calculate Tisserand parameters for close encounters with other planets.

The Tisserand criterion is also applicable to so-called *gravity assists*, in which a spacecraft gains energy due to a close encounter with a moving planet. Such assists are often

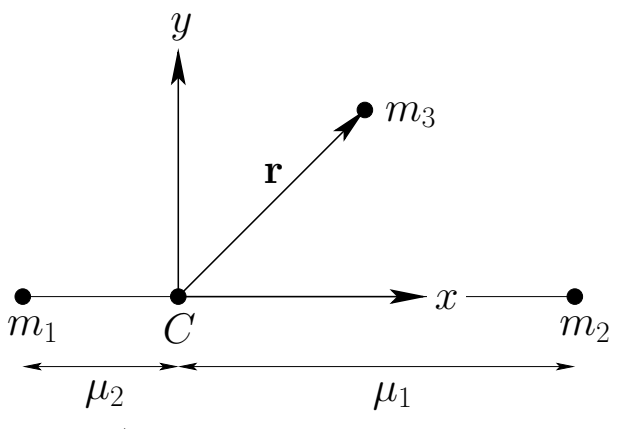


Figure 13.2: *The co-rotating frame.*

employed in missions to the outer planets to reduce the amount of fuel which the space-craft must carry in order to reach its destination. In fact, it is clear, from Equations (13.16) and (13.19), that a space-craft can make use of a close encounter with a moving planet to increase (or decrease) its orbital major radius a, and, hence, to increase (or decrease) its total orbital energy.

13.5 Co-Rotating Frame

Let us transform to a non-inertial frame of reference rotating with angular velocity ω about an axis normal to the orbital plane of masses m_1 and m_2, and passing through their center of mass. It follows that masses m_1 and m_2 appear *stationary* in this new reference frame. Let us define a Cartesian coordinate system (x, y, z) in the rotating frame of reference which is such that masses m_1 and m_2 always lie on the x-axis, and the z-axis is parallel to the previously defined ζ-axis. It follows that masses m_1 and m_2 have the fixed position vectors $\mathbf{r}_1 = \mu_2\,(-1, 0, 0)$ and $\mathbf{r}_2 = \mu_1\,(1, 0, 0)$ in our new coordinate system. Finally, let the position vector of mass m_3 be $\mathbf{r} = (x, y, z)$—see Figure 13.2.

According to Chapter 7, the equation of motion of mass m_3 in the rotating reference frame takes the form

$$\ddot{\mathbf{r}} + 2\,\boldsymbol{\omega}\times\dot{\mathbf{r}} = -\mu_1\,\frac{(\mathbf{r}-\mathbf{r}_1)}{\rho_1^3} - \mu_2\,\frac{(\mathbf{r}-\mathbf{r}_2)}{\rho_2^3} - \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{r}), \tag{13.21}$$

where $\boldsymbol{\omega} = (0, 0, \omega)$, and

$$\rho_1^2 = (x+\mu_2)^2 + y^2 + z^2, \tag{13.22}$$

$$\rho_2^2 = (x-\mu_1)^2 + y^2 + z^2. \tag{13.23}$$

Here, the second term on the left-hand side of Equation (13.21) is the *Coriolis* acceleration, whereas the final term on the right-hand side is the *centrifugal* acceleration. The

components of Equation (13.21) reduce to

$$\ddot{x} - 2\,\omega\,\dot{y} = -\frac{\mu_1\,(x+\mu_2)}{\rho_1^3} - \frac{\mu_2\,(x-\mu_1)}{\rho_2^3} + \omega^2\,x, \tag{13.24}$$

$$\ddot{y} + 2\,\omega\,\dot{x} = -\frac{\mu_1\,y}{\rho_1^3} - \frac{\mu_2\,y}{\rho_2^3} + \omega^2\,y, \tag{13.25}$$

$$\ddot{z} = -\frac{\mu_1\,z}{\rho_1^3} - \frac{\mu_2\,z}{\rho_2^3}, \tag{13.26}$$

which yield

$$\ddot{x} - 2\,\omega\,\dot{y} = -\frac{\partial U}{\partial x}, \tag{13.27}$$

$$\ddot{y} + 2\,\omega\,\dot{x} = -\frac{\partial U}{\partial y}, \tag{13.28}$$

$$\ddot{z} = -\frac{\partial U}{\partial z}, \tag{13.29}$$

where

$$U = -\frac{\mu_1}{\rho_1} - \frac{\mu_2}{\rho_2} - \frac{\omega^2}{2}\,(x^2+y^2) \tag{13.30}$$

is the sum of the gravitational and centrifugal potentials.

Now, it follows from Equations (13.27)–(13.29) that

$$\ddot{x}\,\dot{x} - 2\,\omega\,\dot{x}\,\dot{y} = -\dot{x}\,\frac{\partial U}{\partial x}, \tag{13.31}$$

$$\ddot{y}\,\dot{y} + 2\,\omega\,\dot{x}\,\dot{y} = -\dot{y}\,\frac{\partial U}{\partial y}, \tag{13.32}$$

$$\ddot{z}\,\dot{z} = -\dot{z}\,\frac{\partial U}{\partial z}. \tag{13.33}$$

Summing the above three equations, we obtain

$$\frac{d}{dt}\left[\frac{1}{2}\left(\dot{x}^2+\dot{y}^2+\dot{z}^2\right) + U\right] = 0. \tag{13.34}$$

In other words,

$$C = -2\,U - v^2 \tag{13.35}$$

is a *constant of the motion*, where $v^2 = \dot{x}^2+\dot{y}^2+\dot{z}^2$. In fact, C is the *Jacobi integral* introduced in Section 13.3 [it is easily demonstrated that Equations (13.10) and (13.35) are identical]. Note, finally, that the mass m_3 is restricted to regions in which

$$-2\,U \geq C, \tag{13.36}$$

since v^2 is a positive definite quantity.

13.6 Lagrange Points

Let us search for possible *equilibrium points* of the mass m_3 in the rotating reference frame. Such points are termed *Lagrange points*. Thus, in the rotating frame, the mass m_3 would remain at rest if placed at one of the Lagrange points. It is, thus, clear that these points are *fixed* in the rotating frame. Conversely, in the inertial reference frame, the Lagrange points *rotate* about the center of mass with angular velocity ω, and the mass m_3 would consequently also rotate about the center of mass with angular velocity ω if placed at one of these points (with the appropriate velocity). In the following, we shall assumed, without loss of generality, that $m_1 \geq m_2$.

The Lagrange points satisfy $\dot{\mathbf{r}} = \ddot{\mathbf{r}} = \mathbf{0}$ in the rotating frame. It thus follows, from Equations (13.27)–(13.29), that the Lagrange points are the solutions of

$$\frac{\partial U}{\partial x} = \frac{\partial U}{\partial y} = \frac{\partial U}{\partial z} = 0. \tag{13.37}$$

Now, it is easily seen that

$$\frac{\partial U}{\partial z} = \left(\frac{\mu_1}{\rho_1^3} + \frac{\mu_2}{\rho_2^3}\right) z. \tag{13.38}$$

Since the term in brackets is positive definite, we conclude that the only solution to the above equation is $z = 0$. Hence, all of the Lagrange points lie in the x-y plane.

If $z = 0$ then it is readily demonstrated that

$$\mu_1 \rho_1^2 + \mu_2 \rho_2^2 = x^2 + y^2 + \mu_1 \mu_2, \tag{13.39}$$

where use has been made of the fact that $\mu_1 + \mu_2 = 1$. Hence, Equation (13.30) can also be written

$$U = -\mu_1 \left(\frac{1}{\rho_1} + \frac{\rho_1^2}{2}\right) - \mu_2 \left(\frac{1}{\rho_2} + \frac{\rho_2^2}{2}\right) + \frac{\mu_1 \mu_2}{2}. \tag{13.40}$$

The Lagrange points thus satisfy

$$\frac{\partial U}{\partial x} = \frac{\partial U}{\partial \rho_1}\frac{\partial \rho_1}{\partial x} + \frac{\partial U}{\partial \rho_2}\frac{\partial \rho_2}{\partial x} = 0, \tag{13.41}$$

$$\frac{\partial U}{\partial y} = \frac{\partial U}{\partial \rho_1}\frac{\partial \rho_1}{\partial y} + \frac{\partial U}{\partial \rho_2}\frac{\partial \rho_2}{\partial y} = 0, \tag{13.42}$$

which reduce to

$$\mu_1 \left(\frac{1-\rho_1^3}{\rho_1^2}\right)\left(\frac{x+\mu_2}{\rho_1}\right) + \mu_2 \left(\frac{1-\rho_2^3}{\rho_2^2}\right)\left(\frac{x-\mu_1}{\rho_2}\right) = 0, \tag{13.43}$$

$$\mu_1 \left(\frac{1-\rho_1^3}{\rho_1^2}\right)\left(\frac{y}{\rho_1}\right) + \mu_2 \left(\frac{1-\rho_2^3}{\rho_2^2}\right)\left(\frac{y}{\rho_2}\right) = 0. \tag{13.44}$$

Now, one obvious solution of Equation (13.44) is $y = 0$, corresponding to a Lagrange point which lies on the x-axis. It turns out that there are *three* such points. L_1 lies between

masses m_1 and m_2, L_2 lies to the right of mass m_2, and L_3 lies to the left of mass m_1—see Figure 13.2. At the L_1 point, we have $x = -\mu_2 + \rho_1 = \mu_1 - \rho_2$ and $\rho_1 = 1 - \rho_2$. Hence, from Equation (13.43),

$$\frac{\mu_2}{3\,\mu_1} = \frac{\rho_2^3\,(1 - \rho_2 + \rho_2^2/3)}{(1 + \rho_2 + \rho_2^2)\,(1 - \rho_2)^3}. \tag{13.45}$$

Assuming that $\rho_2 \ll 1$, we can find an approximate solution of Equation (13.45) by expanding in powers of ρ_2:

$$\alpha = \rho_2 + \frac{\rho_2^2}{3} + \frac{\rho_2^3}{3} + \frac{51\,\rho_2^4}{81} + \mathcal{O}(\rho_2^5). \tag{13.46}$$

This equation can be inverted to give

$$\rho_2 = \alpha - \frac{\alpha^2}{3} - \frac{\alpha^3}{9} - \frac{23\,\alpha^4}{81} + \mathcal{O}(\alpha^5), \tag{13.47}$$

where

$$\alpha = \left(\frac{\mu_2}{3\,\mu_1}\right)^{1/3}. \tag{13.48}$$

is assumed to be a small parameter.

At the L_2 point, we have $x = -\mu_2 + \rho_1 = \mu_1 + \rho_2$ and $\rho_1 = 1 + \rho_2$. Hence, from Equation (13.43),

$$\frac{\mu_2}{3\,\mu_1} = \frac{\rho_2^3\,(1 + \rho_2 + \rho_2^2/3)}{(1 + \rho_2)^2\,(1 - \rho_2^3)}. \tag{13.49}$$

Again, expanding in powers of ρ_2, we obtain

$$\alpha = \rho_2 - \frac{\rho_2^2}{3} + \frac{\rho_2^3}{3} + \frac{\rho_2^4}{81} + \mathcal{O}(\rho_2^5), \tag{13.50}$$

$$\rho_2 = \alpha + \frac{\alpha^2}{3} - \frac{\alpha^3}{9} - \frac{31\,\alpha^4}{81} + \mathcal{O}(\alpha^5). \tag{13.51}$$

Finally, at the L_3 point, we have $x = -\mu_2 - \rho_1 = \mu_1 - \rho_2$ and $\rho_2 = 1 + \rho_1$. Hence, from Equation (13.43),

$$\frac{\mu_2}{\mu_1} = \frac{(1 - \rho_1^3)\,(1 + \rho_1)^2}{\rho_1^3\,(\rho_1^2 + 3\,\rho_1 + 3)}. \tag{13.52}$$

Let $\rho_1 = 1 - \beta$. Expanding in powers of β, we obtain

$$\frac{\mu_2}{\mu_1} = \frac{12\,\beta}{7} + \frac{144\,\beta^2}{49} + \frac{1567\,\beta^3}{343} + \mathcal{O}(\beta^4), \tag{13.53}$$

$$\beta = \frac{7}{12}\left(\frac{\mu_2}{\mu_1}\right) - \frac{7}{12}\left(\frac{\mu_2}{\mu_1}\right)^2 + \frac{13223}{20736}\left(\frac{\mu_2}{\mu_1}\right)^3 + \mathcal{O}\left(\frac{\mu_2}{\mu_1}\right)^4, \tag{13.54}$$

where μ_2/μ_1 is assumed to be a small parameter.

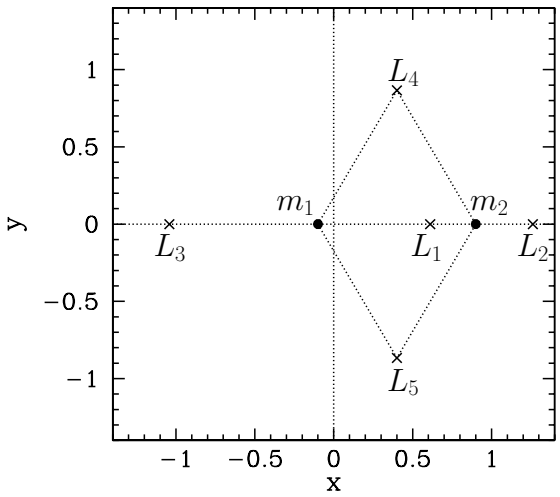


Figure 13.3: *The masses* m_1 *and* m_2*, and the five Lagrange points,* L_1 *to* L_5*, calculated for* $\mu_2 = 0.1$.

Let us now search for Lagrange points which *do not lie* on the x-axis. One obvious solution of Equations (13.41) and (13.42) is

$$\frac{\partial U}{\partial \rho_1} = \frac{\partial U}{\partial \rho_2} = 0, \tag{13.55}$$

giving, from Equation (13.40),

$$\rho_1 = \rho_2 = 1, \tag{13.56}$$

or

$$(x + \mu_2)^2 + y^2 = (x - 1 + \mu_2)^2 + y^2 = 1, \tag{13.57}$$

since $\mu_1 = 1 - \mu_2$. The two solutions of the above equation are

$$x = \frac{1}{2} - \mu_2, \tag{13.58}$$

$$y = \pm\frac{\sqrt{3}}{2}, \tag{13.59}$$

and specify the positions of the Lagrange points designated L_4 and L_5. Note that point L_4 and the masses m_1 and m_2 lie at the apexes of an *equilateral triangle*. The same is true for point L_5. We have now found all of the possible Lagrange points.

Figure 13.3 shows the positions of the two masses, m_1 and m_2, and the five Lagrange points, L_1 to L_5, calculated for the case where $\mu_2 = 0.1$.

13.7 Zero-Velocity Surfaces

Consider the surface

$$V(x, y, z) = C, \tag{13.60}$$

where

$$V = -2\,U = \frac{2\,\mu_1}{\rho_1} + \frac{2\,\mu_2}{\rho_2} + x^2 + y^2. \tag{13.61}$$

Note that $V \geq 0$. It follows, from Equation (13.35), that if the mass m_3 has the Jacobi integral C, and lies on the surface specified in Equation (13.60), then it must have *zero* velocity. Hence, such a surface is termed a *zero-velocity surface*. The zero-velocity surfaces are important because they form the boundary of regions from which the mass m_3 is dynamically *excluded*: *i.e.*, regions in which $V < C$. Generally speaking, the regions from which m_3 is excluded grow in area as C increases, and *vice versa*.

Let C_i be the value of V at the L_i Lagrange point, for $i = 1, 5$. When $\mu_2 \ll 1$, it is easily demonstrated that

$$C_1 \simeq 3 + 3^{4/3}\,\mu_2^{2/3} - 10\,\mu_2/3, \tag{13.62}$$

$$C_2 \simeq 3 + 3^{4/3}\,\mu_2^{2/3} - 14\,\mu_2/3, \tag{13.63}$$

$$C_3 \simeq 3 + \mu_2, \tag{13.64}$$

$$C_4 \simeq 3 - \mu_2, \tag{13.65}$$

$$C_4 \simeq 3 - \mu_2. \tag{13.66}$$

Note that $C_1 > C_2 > C_3 > C_4 = C_5$.

Figures 13.4–13.8 show the intersection of the zero-velocity surface $V = C$ with the x-y plane for various different values of C, and illustrate how the region from which m_3 is dynamically excluded—which we shall term the *excluded region*—evolves as the value of C is varied. Of course, any point not in the excluded region is in the so-called *allowed region*. For $C > C_1$, the allowed region consists of two separate oval regions centered on m_1 and m_2, respectively, plus an outer region which lies beyond a large circle centered on the origin. All three allowed regions are separated from one another by an excluded region—see Figure 13.4. When $C = C_1$, the two inner allowed regions merge at the L_1 point—see Figure 13.5. When $C = C_2$, the inner and outer allowed regions merge at the L_2 point, forming a horseshoe-like excluded region—see Figure 13.6. When $C = C_3$, the excluded region splits in two at the L_3 point—see Figure 13.7. For $C_4 < C < C_3$, the two excluded regions are localized about the L_4 and L_5 points—see Figure 13.8. Finally, for $C < C_4$, there is no excluded region.

Figure 13.9 shows the zero-velocity surfaces and Lagrange points calculated for the case $\mu_2 = 0.01$. It can be seen that, at very small values of μ_2, the L_1 and L_2 Lagrange points are almost *equidistant* from mass m_2. Furthermore, mass m_2, and the L_3, L_4, and L_5 Lagrange points all lie approximately on a *unit circle*, centered on mass m_1. It follows that, when μ_2 is small, the Lagrange points L_3, L_4 and L_5 all share the orbit of mass m_2 about m_1

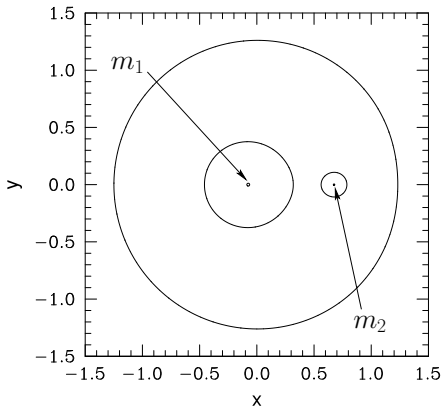


Figure 13.4: *The zero-velocity surface* $V = C$*, where* $C > C_1$*, calculated for* $\mu_2 = 0.1$*. The mass* m_3 *is excluded from the region lying between the two inner curves and the outer curve.*

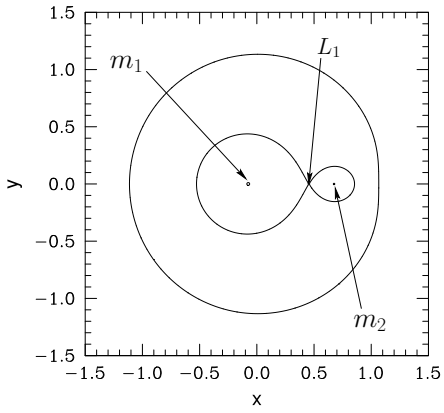


Figure 13.5: *The zero-velocity surface* $V = C$*, where* $C = C_1$*, calculated for* $\mu_2 = 0.1$*. The mass* m_3 *is excluded from the region lying between the inner and outer curves.*

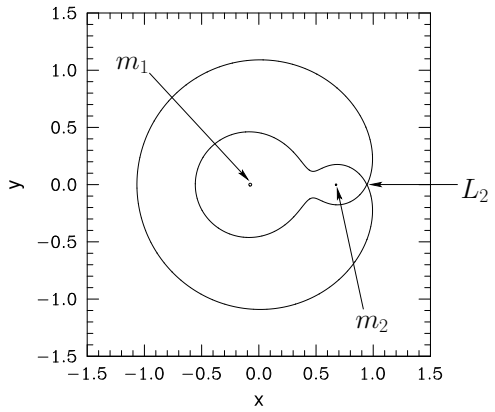


Figure 13.6: *The zero-velocity surface* $V = C$*, where* $C = C_2$*, calculated for* $\mu_2 = 0.1$*. The mass* m_3 *is excluded from the region lying between the inner and outer curve.*

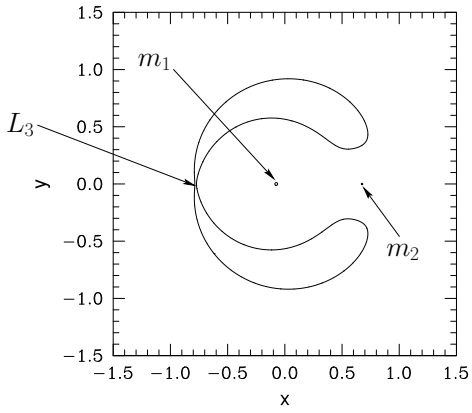


Figure 13.7: *The zero-velocity surface* $V = C$*, where* $C = C_3$*, calculated for* $\mu_2 = 0.1$*. The mass* m_3 *is excluded from the regions lying inside the curve.*

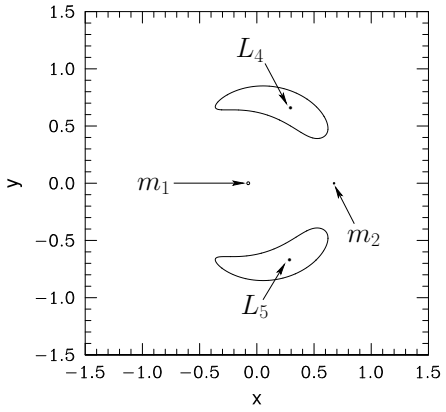


Figure 13.8: *The zero-velocity surface* $V = C$*, where* $C_4 < C < C_3$*, calculated for* $\mu_2 = 0.1$*. The mass* m_3 *is excluded from the regions lying inside the two curves.*

(in the inertial frame) with C_3 being directly opposite m_2, L_4 (by convention) 60° ahead of m_2, and L_5 60° behind.

13.8 Stability of Lagrange Points

We have seen that the five Lagrange points, L_1 to L_5, are the equilibrium points of mass m_3 in the co-rotating frame. Let us now determine whether or not these equilibrium points are *stable* to small displacements.

Now, the equations of motion of mass m_3 in the co-rotating frame are specified in Equations (13.27)–(13.29). Note that the motion in the x-y plane is complicated by presence of the Coriolis acceleration. However, the motion parallel to the z-axis simply corresponds to motion in the potential U. Hence, the condition for the stability of the Lagrange points (which all lie at $z = 0$) to small displacements parallel to the z-axis is simply (see Section 3.2)

$$\left(\frac{\partial^2 U}{\partial z^2}\right)_{z=0} = \frac{\mu_1}{\rho_1^3} + \frac{\mu_2}{\rho_2^3} > 0. \tag{13.67}$$

This condition is satisfied everywhere in the x-y plane. Hence, the Lagrange points are all *stable* to small displacements parallel to the z-axis. It, thus, remains to investigate their stability to small displacements lying within the x-y plane.

Suppose that a Lagrange point is situated in the x-y plane at coordinates $(x_0, y_0, 0)$. Let us consider small amplitude x-y motion in the vicinity of this point by writing

$$x = x_0 + \delta x, \tag{13.68}$$

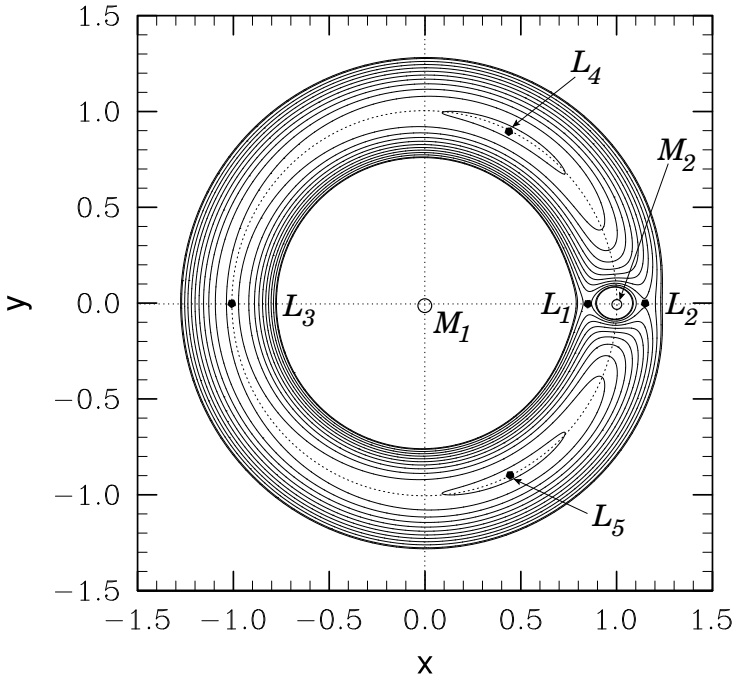


Figure 13.9: *The zero-velocity surfaces and Lagrange points calculated for* $\mu_2 = 0.01$.

$$y = y_0 + \delta y, \tag{13.69}$$

$$z = 0, \tag{13.70}$$

where δx and δy are infinitesimal. Expanding $U(x, y, 0)$ about the Lagrange point as a Taylor series, and retaining terms up to second-order in small quantities, we obtain

$$U = U_0 + U_x\,\delta x + U_y\,\delta y + \frac{1}{2}\,U_{xx}\,(\delta x)^2 + U_{xy}\,\delta x\,\delta y + \frac{1}{2}\,U_{yy}\,(\delta y)^2, \tag{13.71}$$

where $U_0 = U(x_0, y_0, 0)$, $U_x = \partial U(x_0, y_0, 0)/\partial x$, $U_{xx} = \partial^2 U(x_0, y_0, 0)/\partial x^2$, *etc.* However, by definition, $U_x = U_y = 0$ at a Lagrange point, so the expansion simplifies to

$$U = U_0 + \frac{1}{2}\,U_{xx}\,(\delta x)^2 + U_{xy}\,\delta x\,\delta y + \frac{1}{2}\,U_{yy}\,(\delta y)^2. \tag{13.72}$$

Finally, substitution of Equations (13.68)–(13.70), and (13.72) into the equations of x-y motion, (13.27) and (13.28), yields

$$\delta\ddot{x} - 2\,\delta\dot{y} = -U_{xx}\,\delta x - U_{xy}\,\delta y, \tag{13.73}$$

$$\delta\ddot{y} + 2\,\delta\dot{x} = -U_{xy}\,\delta x - U_{yy}\,\delta y, \tag{13.74}$$

since $\omega = 1$.

Let us search for a solution of the above pair of equations of the form $\delta x(t) = \delta x_0\,\exp(\gamma\,t)$ and $\delta y(t) = \delta y_0\,\exp(\gamma\,t)$. We obtain

$$\begin{pmatrix} \gamma^2 + U_{xx} & -2\,\gamma + U_{xy} \\ 2\,\gamma + U_{xy} & \gamma^2 + U_{yy} \end{pmatrix} \begin{pmatrix} \delta x_0 \\ \delta y_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{13.75}$$

This equation only has a non-trivial solution if the determinant of the matrix is zero. Hence, we get

$$\gamma^4 + (4 + U_{xx} + U_{yy})\,\gamma^2 + (U_{xx}\,U_{yy} - U_{xy}^2) = 0. \tag{13.76}$$

Now, it is convenient to define

$$A = \frac{\mu_1}{\rho_1^3} + \frac{\mu_2}{\rho_2^3}, \tag{13.77}$$

$$B = 3\left[\frac{\mu_1}{\rho_1^5} + \frac{\mu_2}{\rho_2^5}\right] y^2, \tag{13.78}$$

$$C = 3\left[\frac{\mu_1\,(x + \mu_2)}{\rho_1^5} + \frac{\mu_2\,(x - \mu_1)}{\rho_2^5}\right] y, \tag{13.79}$$

$$D = 3\left[\frac{\mu_1\,(x + \mu_2)^2}{\rho_1^3} + \frac{\mu_2\,(x - \mu_1)^2}{\rho_2^3}\right], \tag{13.80}$$

where all terms are evaluated at the point $(x_0, y_0, 0)$. It thus follows that

$$U_{xx} = A - D - 1, \tag{13.81}$$

$$U_{yy} = A - B - 1, \tag{13.82}$$

$$U_{xy} = -C. \tag{13.83}$$

Consider the co-linear Lagrange points, L_1, L_2, and L_3. These all lie on the x-axis, and are thus characterized by $y = 0$, $\rho_1^2 = (x + \mu_2)^2$, and $\rho_2^2 = (x - \mu_1)^2$. It follows, from the above equations, that $B = C = 0$ and $D = 3\,A$. Hence, $U_{xx} = -1 - 2\,A$, $U_{yy} = A - 1$, and $U_{xy} = 0$. Equation (13.76) thus yields

$$\Gamma^2 + (2 - A)\,\Gamma + (1 - A)\,(1 + 2\,A) = 0, \tag{13.84}$$

where $\Gamma = \gamma^2$. Now, in order for a Lagrange point to be stable to small displacements, all four of the roots, γ, of Equation (13.76) must be *purely imaginary*. This, in turn, implies that the two roots of the above equation,

$$\Gamma = \frac{A - 2 \pm \sqrt{A\,(9\,A - 8)}}{2}, \tag{13.85}$$

must both be *real and negative*. Thus, the stability criterion is

$$\frac{8}{9} \leq A \leq 1. \tag{13.86}$$

Figure 13.10 shows A calculated at the three co-linear Lagrange points as a function of μ_2, for all allowed values of this parameter (*i.e.*, $0 < \mu_2 \leq 0.5$). It can be seen that A is always greater than unity for all three points. Hence, we conclude that the co-linear Lagrange points, L_1, L_2, and L_3, are intrinsically *unstable* equilibrium points in the co-rotating frame.

Let us now consider the triangular Lagrange points, L_4 and L_5. These points are characterized by $\rho_1 = \rho_2 = 1$. It follows that $A = 1$, $B = 9/4$, $C = \pm\sqrt{27/16}\,(1 - 2\,\mu_2)$, and $D = 3/4$. Hence, $U_{xx} = -3/4$, $U_{yy} = -9/4$, and $U_{xy} = \mp\sqrt{27/16}\,(1 - 2\,\mu_2)$, where the upper/lower signs corresponds to L_4 and L_5, respectively. Equation (13.76) thus yields

$$\Gamma^2 + \Gamma + \frac{27}{4}\,\mu_2\,(1 - \mu_2) = 0 \tag{13.87}$$

for both points, where $\Gamma = \gamma^2$. As before, the stability criterion is that the two roots of the above equation must both be real and negative. This is the case provided that $1 > 27\,\mu_2\,(1 - \mu_2)$, which yields the stability criterion

$$\mu_2 < \frac{1}{2}\left(1 - \sqrt{\frac{23}{27}}\right) = 0.0385. \tag{13.88}$$

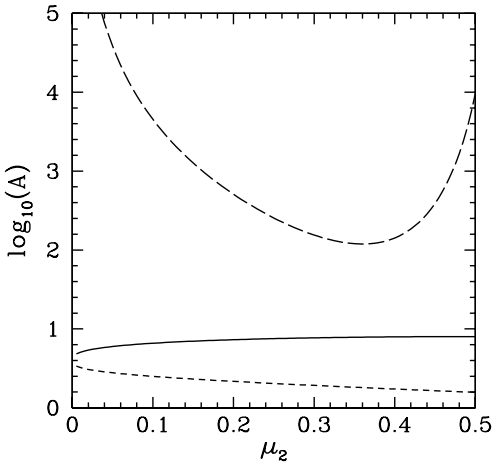


Figure 13.10: *The solid, short-dashed, and long-dashed curves show* A *as a function of* μ_2 *at the* L_1*,* L_2*, and* L_3 *Lagrange points.*

In unnormalized units, this criterion becomes

$$\frac{m_2}{m_1 + m_2} < 0.0385. \tag{13.89}$$

We thus conclude that the L_4 and L_5 Lagrange points are *stable* equilibrium points, in the co-rotating frame, provided that mass m_2 is less than about 4% of mass m_1. If this is the case then mass m_3 can orbit around these points indefinitely. In the inertial frame, the mass will share the orbit of mass m_2 about mass m_1, but will stay approximately 60° *ahead of* mass m_2, if it is orbiting the L_4 point, or 60° *behind*, if it is orbiting the L_5 point—see Figure 13.9. This type of behavior has been observed in the Solar System. For instance, there is a sub-class of asteroids, known as the *Trojan asteroids*, which are trapped in the vicinity of the L_4 and L_5 points of the Sun-Jupiter system (which easily satisfies the above stability criterion), and consequently share Jupiter's orbit around the Sun, staying approximately 60° *ahead of*, and 60° *behind*, Jupiter, respectively. Furthermore, the L_4 and L_5 points of the Sun-Earth system are occupied by clouds of dust.

14 Lunar Motion

14.1 Historical Background

The motion of the Planets around the Sun is fairly accurately described by *Kepler's laws* (see Chapter 5). Likewise, to a first approximation, the motion of the Moon around the Earth can also be accounted for via these laws. However, unlike the planetary orbits, the deviations of the lunar orbit from a Keplerian ellipse are sufficiently large that they are easily apparent to the naked eye. Indeed, the largest of these deviations, which is generally known as *evection*, was discovered in ancient times by the Alexandrian astronomer Claudius Ptolemy (90–168 AD). Moreover, the next largest deviation, which is called *variation*, was first observed by Tycho Brahe (1546–1601) without the aid of a telescope. Another non-Keplerian feature of the lunar orbit, which is sufficient obvious that it was known to the ancient Greeks, is the fact that the lunar *perigee* (*i.e.*, the point of closest approach to the Earth) *precesses* (*i.e.*, orbits about the Earth in the same direction as the Moon) at such a rate that it completes a full circuit every 8.85 years.[1] The ancient Greeks also noticed that the lunar *ascending node* (*i.e.*, the point at which the Moon passes through the fixed plane of the Earth's orbit around the Sun from south to north) *regresses* (*i.e.*, orbits about the Earth in the opposite direction to the Moon) at such a rate that it completes a full circuit every 18.6 years. Of course, according to standard two-body orbit theory, both the lunar perigee and ascending node should be *stationary* (see Chapter 5).

Newton demonstrated, in Book III of his Principia, that the deviations of the lunar orbit from a Keplerian ellipse are due to the gravitational influence of the Sun, which is sufficiently large that it is not completely negligible compared with the mutual gravitational attraction of the Earth and the Moon. However, Newton was not able to give a complete account of these deviations, due to the complexity of the equations of motion which arise in a system of three mutually gravitating bodies (see Chapter 13). In fact, Clairaut (1713–1765) is generally credited with the first reasonably accurate and complete theoretical explanation of the Moon's orbit. His method of calculation makes use of an expansion of the lunar equations of motion in terms of various small parameters. Clairaut, however, initially experienced difficulty in accounting for the precession of the lunar perigee. Indeed, his first calculation overestimated the period of this precession by a factor of about two, leading him to question Newton's inverse-square law of gravitation. Later on, he realized that he could indeed account for the precession, in terms of standard Newtonian dynamics, by continuing his expansion in small parameters to higher order. After Clairaut, the theory of lunar motion was further elaborated in major works by D'Alembert (1717–1783), Euler (1707–1783), Laplace (1749–1827), Damoiseau (1768–1846), Plana (1781–1864), Poisson (1781–1840), Hansen (1795–1874), De Pontécoulant (1795–1824),

[1] Note that this precession rate is about 10^4 times greater than any of the planetary perihelion precession rates discussed in Section 12.12.

J. Herschel (1792–1871), Airy (1801–1892), Delaunay (1816–1872), G.W. Hill (1836–1914), and E.W. Brown (1836–1938). The fact that so many celebrated mathematicians and astronomers devoted so much time and effort to lunar theory is a tribute to its inherent difficulty, as well as its great theoretical and practical interest. Indeed, for a period of about one hundred years (between 1767 and about 1850) the so-called *method of lunar distance* was the principal means used by mariners to determine longitude at sea. This method depends crucially on a precise knowledge of the position of the Moon in the sky as a function of time. Consequently, astronomers and mathematicians during the period in question were spurred to make ever more accurate observations of the Moon's orbit, and to develop lunar theory to greater and greater precision. An important outcome of these activities were various tables of lunar motion (*e.g.*, those of Mayer, Damoiseau, Plana, Hansen, and Brown), most of which were published at public expense. Finally, it is worth noting that the development of lunar theory gave rise to many significant advances in mathematics and dynamics, and was, in many ways, a precursor to the discovery of chaotic motion in the 20th century (see Chapter 15).

This chapter contains an introduction to lunar theory in which approximate expressions for evection, variation, the precession of the perigee, and the regression of the ascending node are derived from the laws of Newtonian dynamics.[2]

14.2 Preliminary Analysis

Let $\mathbf{r}_E$ and $\mathbf{r}_M$ be the position vectors of the Earth and Moon, respectively, in a *non-rotating* reference frame in which the Sun is at rest at the origin. Treating this reference frame as *inertial* (which is an excellent approximation, given that the mass of the Sun is very much greater than that of the Earth or the Moon—see Chapter 6), the Earth's equation of motion becomes (see Chapter 5)

$$\ddot{\mathbf{r}}_E = -n'^2\, a'^3\, \frac{\mathbf{r}_E}{|\mathbf{r}_E|^3}, \tag{14.1}$$

where $n' = 0.98560025^\circ$ per day and $a' = 149,598,261$ km are the mean angular velocity and major radius, respectively, of the terrestrial orbit about the Sun. Here, $\dot{} \equiv d/dt^2$. On the other hand, the Moon's equation of motion takes the form

$$\ddot{\mathbf{r}}_M = -n^2\, a^3\, \frac{(\mathbf{r}_M - \mathbf{r}_E)}{|\mathbf{r}_M - \mathbf{r}_E|^3} - n'^2\, a'^3\, \frac{\mathbf{r}_M}{|\mathbf{r}_M|^3}, \tag{14.2}$$

where $n = 13.17639646^\circ$ per day and $a = 384,399$ km are the mean angular velocity and major radius, respectively, of the lunar orbit about the Earth. Note that we have retained the perturbing influence (*i.e.*, acceleration) of the Sun in the lunar equation of motion, (14.2), whilst neglecting the perturbing influence of the Moon in the terrestrial

[2]For more information on lunar theory, see *An Elementary Treatise on the Lunar Theory*, H. Godfray (Macmillan & Co., 1853); *An Introductory Treatise on the Lunar Theory*, E.W. Brown (Cambridge University Press, 1896); *Lectures on the Lunar Theory*, J.C. Adams (Cambridge University Press, 1900).

equation of motion, (14.1), since the former influence is significantly greater (by a factor $M_E/M_M \simeq 81$, where M_E is the mass of the Earth, and M_M the mass of the Moon) than the latter.

Let

$$\mathbf{r} = \mathbf{r}_M - \mathbf{r}_E, \tag{14.3}$$

$$\mathbf{r}' = -\mathbf{r}_E, \tag{14.4}$$

be the position vectors of the Moon and Sun, respectively, relative to the Earth. It follows, from Equations (14.1)–(14.4), that in a non-inertial reference frame, S (say), in which the Earth is at rest at the origin, but the coordinate axes point in *fixed* directions, the lunar and solar equations of motion take the form

$$\ddot{\mathbf{r}} = -n^2\, a^3 \frac{\mathbf{r}}{|\mathbf{r}|^3} + n'^2\, a'^3 \left[\frac{(\mathbf{r}'-\mathbf{r})}{|\mathbf{r}'-\mathbf{r}|^3} - \frac{\mathbf{r}'}{|\mathbf{r}'|^3}\right], \tag{14.5}$$

$$\ddot{\mathbf{r}}' = -n'^2\, a'^3 \frac{\mathbf{r}'}{|\mathbf{r}'|^3}, \tag{14.6}$$

respectively.

Let us set up a conventional Cartesian coordinate system in S which is such that the (apparent) orbit of the Sun about the Earth lies in the x-y plane. This implies that the x-y plane corresponds to the so-called *ecliptic plane*. Accordingly, in S, the Sun appears to orbit the Earth at the mean angular velocity $\boldsymbol{\omega}' = n'\,\mathbf{e}_z$ (assuming that the z-axis points toward the so-called north ecliptic pole), whereas the projection of the Moon onto the ecliptic plane orbits the Earth at the mean angular velocity $\boldsymbol{\omega} = n\,\mathbf{e}_z$.

In the following, for the sake of simplicity, we shall neglect the small eccentricity, $e' = 0.016711$, of the Sun's apparent orbit about the Earth (which is actually the eccentricity of the Earth's orbit about the Sun), and approximate the solar orbit as a *circle*, centered on the Earth. Thus, if x', y', z' are the Cartesian coordinates of the Sun in S then an appropriate solution of the solar equation of motion, (14.6), is

$$x' = a'\cos(n't), \tag{14.7}$$

$$y' = a'\sin(n't), \tag{14.8}$$

$$z' = 0. \tag{14.9}$$

14.3 Lunar Equations of Motion

It is convenient to solve the lunar equation of motion, (14.5), in a *geocentric* frame of reference, S_1 (say), which *rotates* with respect to S at the fixed angular velocity $\boldsymbol{\omega}$. Thus, if the lunar orbit were a circle, centered on the Earth, and lying in the ecliptic plane, then the Moon would appear *stationary* in S_1. In fact, the small eccentricity of the lunar orbit,

$e = 0.05488$, combined with its slight inclination to the ecliptic plane, $\iota = 5.161^\circ$, causes the Moon to execute a small periodic orbit about the stationary point.

Let x, y, z and x_1, y_1, z_1 be the Cartesian coordinates of the Moon in S and S_1, respectively. It is easily demonstrated that (see Section A.16)

$$x = x_1 \cos(n\,t) - y_1 \sin(n\,t), \tag{14.10}$$

$$y = x_1 \sin(n\,t) + y_1 \cos(n\,t), \tag{14.11}$$

$$z = z_1. \tag{14.12}$$

Moreover, if x_1', y_1', z_1' are the Cartesian components of the Sun in S_1 then (see Section A.5)

$$x_1' = x' \cos(n\,t) + y' \sin(n\,t), \tag{14.13}$$

$$y_1' = -x' \sin(n\,t) + y' \cos(n\,t), \tag{14.14}$$

$$z_1' = z', \tag{14.15}$$

giving

$$x_1' = a' \cos[(n - n')\,t], \tag{14.16}$$

$$y_1' = -a' \sin[(n - n')\,t], \tag{14.17}$$

$$z_1' = 0, \tag{14.18}$$

where use has been made of Equations (14.7)–(14.9).

Now, in the rotating frame S_1, the lunar equation of motion (14.5) transforms to (see Chapter 7)

$$\ddot{\mathbf{r}} + 2\,\boldsymbol{\omega} \times \dot{\mathbf{r}} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) = -n^2\,a^3\,\frac{\mathbf{r}}{|\mathbf{r}|^3} + n'^2\,a'^3\left[\frac{(\mathbf{r}' - \mathbf{r})}{|\mathbf{r}' - \mathbf{r}|^3} - \frac{\mathbf{r}'}{|\mathbf{r}'|^3}\right], \tag{14.19}$$

where $\dot{} \equiv d/dt$. Furthermore, expanding the final term on the right-hand side of (14.19) to lowest order in the small parameter $a/a' = 0.00257$, we obtain

$$\ddot{\mathbf{r}} + 2\,\boldsymbol{\omega} \times \dot{\mathbf{r}} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \simeq -n^2\,a^3\,\frac{\mathbf{r}}{|\mathbf{r}|^3} + \frac{n'^2\,a'^3}{|\mathbf{r}'|^3}\left[\frac{(3\,\mathbf{r}\cdot\mathbf{r}')\,\mathbf{r}'}{|\mathbf{r}'|^2} - \mathbf{r}\right]. \tag{14.20}$$

When written in terms of Cartesian coordinates, the above equation yields

$$\ddot{x}_1 - 2\,n\,\dot{y}_1 - \left(n^2 + n'^2/2\right) x_1 \simeq -n^2\,a^3\,\frac{x_1}{r^3} + \frac{3}{2}\,n'^2\cos[2\,(n - n')\,t]\,x_1 - \frac{3}{2}\,n'^2\sin[2\,(n - n')\,t]\,y_1, \tag{14.21}$$

$$\ddot{y}_1 + 2\,n\,\dot{x}_1 - \left(n^2 + n'^2/2\right) y_1 \simeq -n^2\,a^3\,\frac{y_1}{r^3} - \frac{3}{2}\,n'^2\sin[2\,(n - n')\,t]\,x_1 - \frac{3}{2}\,n'^2\cos[2\,(n - n')\,t]\,y_1, \tag{14.22}$$

$$\ddot{z}_1 + n'^2\,z_1 \simeq -n^2\,a^3\,\frac{z_1}{r^3}, \tag{14.23}$$

where $r = (x_1^2 + y_1^2 + z_1^2)^{1/2}$, and use has been made of Equations (14.16)–(14.18).

It is convenient, at this stage, to normalize all lengths to a, and all times to n^{-1}. Accordingly, let

$$X = x_1/a, \tag{14.24}$$

$$Y = y_1/a, \tag{14.25}$$

$$Z = z_1/a, \tag{14.26}$$

and $r/a = R = (X^2 + Y^2 + Z^2)^{1/2}$, and $T = n\,t$. In normalized form, Equations (14.21)–(14.23) become

$$\ddot{X} - 2\,\dot{Y} - (1 + m^2/2)\,X \simeq -\frac{X}{R^3} + \frac{3}{2}\,m^2\,\cos[2\,(1-m)\,T]\,X - \frac{3}{2}\,m^2\,\sin[2\,(1-m)\,T]\,Y, \tag{14.27}$$

$$\ddot{Y} + 2\,\dot{X} - (1 + m^2/2)\,Y \simeq -\frac{Y}{R^3} - \frac{3}{2}\,m^2\,\sin[2\,(1-m)\,T]\,X - \frac{3}{2}\,m^2\,\cos[2\,(1-m)\,T]\,Y, \tag{14.28}$$

$$\ddot{Z} + m^2\,Z \simeq -\frac{Z}{R^3}, \tag{14.29}$$

respectively, where $m = n'/n = 0.07480$ is a measure of the perturbing influence of the Sun on the lunar orbit. Here, $\ddot{} \equiv d^2/dT^2$ and $\dot{} \equiv d/dT$.

Finally, let us write

$$X = X_0 + \delta X, \tag{14.30}$$

$$Y = \delta Y, \tag{14.31}$$

$$Z = \delta Z, \tag{14.32}$$

where $X_0 = (1 + m^2/2)^{-1/3}$, and $|\delta X|$, $|\delta Y|$, $|\delta Z| \ll X_0$. Thus, if the lunar orbit were a circle, centered on the Earth, and lying in the ecliptic plane, then, in the rotating frame S_1, the Moon would appear stationary at the point $(X_0, 0, 0)$. Expanding Equations (14.27)–(14.29) to *second-order* in δX, δY, δZ, and neglecting terms of order m^4 and $m^2\,\delta X^2$, *etc.*, we obtain

$$\delta\ddot{X} - 2\,\delta\dot{Y} - 3\,(1 + m^2/2)\,\delta X \simeq \frac{3}{2}\,m^2\,\cos[2\,(1-m)\,T] + \frac{3}{2}\,m^2\,\cos[2\,(1-m)\,T]\,\delta X - \frac{3}{2}\,m^2\,\sin[2\,(1-m)\,T]\,\delta Y - 3\,\delta X^2 + \frac{3}{2}\,(\delta Y^2 + \delta Z^2), \tag{14.33}$$

$$\delta\ddot{Y} + 2\,\delta\dot{X} \simeq -\frac{3}{2}\,m^2\,\sin[2\,(1-m)\,T] - \frac{3}{2}\,m^2\,\sin[2\,(1-m)\,T]\,\delta X$$

$$-\frac{3}{2}\,m^2\,\cos[2\,(1-m)\,T]\,\delta Y + 3\,\delta X\,\delta Y, \tag{14.34}$$

$$\delta\ddot{Z} + (1+3\,m^2/2)\,\delta Z \simeq 3\,\delta X\,\delta Z. \tag{14.35}$$

Now, once the above three equations have been solved for δX, δY, and δZ, the Cartesian coordinates, x, y, z, of the Moon in the non-rotating geocentric frame S are obtained from Equations (14.10)–(14.12), (14.24)–(14.26), and (14.30)–(14.32). However, it is more convenient to write $x = r\cos\theta$, $y = r\sin\theta$, and $z = r\sin\beta$, where r is the *radial distance* between the Earth and Moon, and θ and β are termed the Moon's *ecliptic longitude* and *ecliptic latitude*, respectively. Moreover, it is easily seen that, to second-order in δX, δY, δZ, and neglecting terms of order m^4,

$$\frac{r}{a} - 1 + \frac{m^2}{6} \simeq \delta X + \frac{1}{2}\,\delta Y^2 + \frac{1}{2}\,\delta Z^2, \tag{14.36}$$

$$\theta - n\,t \simeq \delta Y - \delta X\,\delta Y, \tag{14.37}$$

$$\beta \simeq \delta Z - \delta X\,\delta Z. \tag{14.38}$$

14.4 Unperturbed Lunar Motion

Let us, first of all, neglect the perturbing influence of the Sun on the Moon's orbit by setting $m = 0$ in the lunar equations of motion (14.33)–(14.35). For the sake of simplicity, let us also neglect nonlinear effects in these equations by setting $\delta X^2 = \delta Y^2 = \delta Z^2 = \delta X\,\delta Y = \delta X\,\delta Z = 0$. In this case, the equations reduce to

$$\delta\ddot{X} - 2\,\delta\dot{Y} - 3\,\delta X \simeq 0, \tag{14.39}$$

$$\delta\ddot{Y} + 2\,\delta\dot{X} \simeq 0, \tag{14.40}$$

$$\delta\ddot{Z} + \delta Z \simeq 0. \tag{14.41}$$

By inspection, appropriate solutions are

$$\delta X \simeq -e\,\cos(T-\alpha_0), \tag{14.42}$$

$$\delta Y \simeq 2\,e\,\sin(T-\alpha_0), \tag{14.43}$$

$$\delta Z \simeq i\,\sin(T-\gamma_0), \tag{14.44}$$

where e, α_0, i, and γ_0 are arbitrary constants. Recalling that $T = n\,t$, it follows from (14.36)–(14.38) that

$$r \simeq a\,[1 - e\,\cos(n\,t-\alpha_0)], \tag{14.45}$$

$$\theta \simeq n\,t + 2\,e\,\sin(n\,t-\alpha_0), \tag{14.46}$$

$$\beta \simeq \iota\,\sin(n\,t-\gamma_0). \tag{14.47}$$

However, Equations (14.45) and (14.46) are simply first-order (in e) approximations to the familiar Keplerian laws (see Chapter 5)

$$r = \frac{a\,(1-e^2)}{1+e\,\cos(\theta-\alpha_0)}, \tag{14.48}$$

$$r^2\,\dot{\theta} = (1-e^2)^{1/2}\,n\,a^2, \tag{14.49}$$

where $\dot{} \equiv d/dt$. Of course, the above two laws describe a body which executes an *elliptical* orbit, confocal with the Earth, of major radius a, mean angular velocity n, and eccentricity e, such that the radius vector connecting the body to the Earth sweeps out equal areas in equal time intervals. We conclude, unsurprisingly, that the unperturbed lunar orbit is a Keplerian ellipse. Note that the lunar *perigee* lies at the fixed ecliptic longitude $\theta = \alpha_0$. Equation (14.47) is the first-order approximation to

$$\beta = \iota\,\sin(\theta-\gamma_0). \tag{14.50}$$

This expression implies that the unperturbed lunar orbit is *co-planar*, but is *inclined* at an angle ι to the ecliptic plane. Moreover, the *ascending node* lies at the fixed ecliptic longitude $\theta = \gamma_0$. Incidentally, the neglect of nonlinear terms in Equations (14.39)–(14.41) is only valid as long as e, $\iota \ll 1$: *i.e.*, provided that the unperturbed lunar orbit is only slightly elliptical, and slightly inclined to the ecliptic plane. In fact, the observed values of e and ι are 0.05488 and 0.09008 radians, respectively, so this is indeed the case.

14.5 Perturbed Lunar Motion

The perturbed nonlinear lunar equations of motion, (14.33)–(14.35), take the general form

$$\delta\ddot{X} - 2\,\delta\dot{Y} - 3\,(1+m^2/2)\,\delta X \simeq R_X, \tag{14.51}$$

$$\delta\ddot{Y} + 2\,\delta\dot{X} \simeq R_Y, \tag{14.52}$$

$$\delta\ddot{Z} + (1+3\,m^2/2)\,\delta Z \simeq R_Z, \tag{14.53}$$

where

$$R_X = a_0 + \sum_{j>0} a_j\,\cos(\omega_j\,T - \alpha_j), \tag{14.54}$$

$$R_Y = \sum_{j>0} b_j\,\sin(\omega_j\,T - \alpha_j), \tag{14.55}$$

$$R_Z = \sum_{j>0} c_j\,\sin(\Omega_j\,T - \gamma_j). \tag{14.56}$$

j	ω_j	α_j	Ω_j	γ_j
1	$1+c\,m^2$	α_0	$1+g\,m^2$	γ_0
2	$2\,(1+c\,m^2)$	$2\,\alpha_0$	$(c-g)\,m^2$	$\alpha_0-\gamma_0$
3	$2\,(1+g\,m^2)$	$2\,\gamma_0$	$2+(c+g)\,m^2$	$\alpha_0+\gamma_0$
4	$2-2\,m$	0		
5	$1-2\,m-c\,m^2$	$-\alpha_0$	$1-2\,m-g\,m^2$	$-\gamma_0$

Table 14.1: *Angular frequencies and phase-shifts associated with the principal periodic driving terms appearing in the perturbed nonlinear lunar equations of motion.*

Let us search for solutions of the general form

$$\delta X = x_0 + \sum_{j>0} x_j \cos(\omega_j T - \alpha_j), \tag{14.57}$$

$$\delta Y = \sum_{j>0} y_j \sin(\omega_j T - \alpha_j), \tag{14.58}$$

$$\delta Z = \sum_{j>0} z_j \sin(\Omega_j T - \gamma_j). \tag{14.59}$$

Substituting expressions (14.54)–(14.59) into Equations (14.51)–(14.53), it is easily demonstrated that

$$x_0 = -\frac{a_0}{3\,(1+m^2/2)}, \tag{14.60}$$

$$x_j = \frac{\omega_j\,a_j - 2\,b_j}{\omega_j\,(1-3\,m^2/2-\omega_j^2)}, \tag{14.61}$$

$$y_j = \frac{(\omega_j^2+3+3\,m^2/2)\,b_j - 2\,\omega_j\,a_j}{\omega_j^2\,(1-3\,m^2/2-\omega_j^2)}, \tag{14.62}$$

$$z_j = \frac{c_j}{1+3\,m^2/2-\Omega_j^2}, \tag{14.63}$$

where $j>0$.

The angular frequencies, ω_j, Ω_j, and phase shifts, α_j, γ_j, of the principal periodic driving terms that appear on the right-hand sides of the perturbed nonlinear lunar equations of motion, (14.51)–(14.53), are specified in Table 14.1. Here, c and g are, as yet, unspecified $\mathcal{O}(1)$ constants associated with the precession of the lunar perigee, and the regression of the ascending node, respectively. Note that ω_1 and Ω_1 are the frequencies of the Moon's unforced motion in ecliptic longitude and latitude, respectively. Moreover, ω_4 is the forcing frequency associated with the perturbing influence of the Sun. All other frequencies appearing in Table 14.1 are combinations of these three fundamental frequencies. In fact, $\omega_2=2\,\omega_1$, $\omega_3=2\,\Omega_1$, $\omega_5=\omega_4-\omega_1$, $\Omega_2=\omega_1-\Omega_1$, $\Omega_3=\omega_1+\Omega_1$, and $\Omega_5=\omega_4-\Omega_1$. Note that there is no Ω_4.

j	a_j	b_j	c_j
0	$\frac{3}{2}\,e^2+\frac{3}{4}\,\iota^2$		
1	$\frac{3}{4}\,m^2\,x_5-\frac{3}{4}\,m^2\,y_5-3\,x_4\,x_5+\frac{3}{2}\,y_4\,y_5$	$-\frac{3}{4}\,m^2\,x_5+\frac{3}{4}\,m^2\,y_5+\frac{3}{2}\,y_4\,x_5-\frac{3}{2}\,y_5\,x_4$	$-\frac{3}{2}\,x_4\,z_5$
2	$-\frac{9}{2}\,e^2$	$-3\,e^2$	$\frac{3}{2}\,e\,\iota$
3	$-\frac{3}{4}\,\iota^2$	0	$-\frac{3}{2}\,e\,\iota$
4	$\frac{3}{2}\,m^2$	$-\frac{3}{2}\,m^2$	0
5	$-\frac{9}{4}\,m^2\,e+3\,e\,x_4+3\,e\,y_4$	$\frac{9}{4}\,m^2\,e-3\,e\,x_4-\frac{3}{2}\,e\,y_4$	$-\frac{3}{2}\,\iota\,x_4$

Table 14.2: *Amplitudes of the periodic driving terms appearing in the perturbed nonlinear lunar equations of motion.*

Now, a comparison of Equations (14.33)–(14.35), (14.51)–(14.53), and Table 14.1 reveals that

$$\begin{aligned} R_X &= \frac{3}{2}\,m^2\,\cos(\omega_4\,T-\alpha_4)+\frac{3}{2}\,m^2\,\cos(\omega_4\,T-\alpha_4)\,\delta X \\ &\quad -\frac{3}{2}\,m^2\,\sin(\omega_4\,T-\alpha_4)\,\delta Y-3\,\delta X^2+\frac{3}{2}\,(\delta Y^2+\delta Z^2), \qquad (14.64)\\ R_Y &= -\frac{3}{2}\,m^2\,\sin(\omega_4\,T-\alpha_4)-\frac{3}{2}\,m^2\,\sin(\omega_4\,T-\alpha_4)\,\delta X \\ &\quad -\frac{3}{2}\,m^2\,\cos(\omega_4\,T-\alpha_4)\,\delta Y+3\,\delta X\,\delta Y, \qquad (14.65)\\ R_Z &= 3\,\delta X\,\delta Z. \qquad (14.66) \end{aligned}$$

Substitution of the solutions (14.57)–(14.59) into the above equations, followed by a comparison with expressions (14.54)–(14.56), yields the amplitudes a_j, b_j, and c_j specified in Table 14.2. Note that, in calculating these amplitudes, we have neglected all contributions to the periodic driving terms, appearing in Equations (14.51)–(14.53) which involve cubic, or higher order, combinations of e, ι, m^2, x_j, y_j, and z_j, since we only expanded Equations (14.33)–(14.35) to *second-order* in δX, δY, and δZ.

For $j=0$, it follows from Equation (14.60) and Table 14.2 that

$$x_0=-\frac{1}{2}\,e^2-\frac{1}{4}\,\iota^2. \qquad (14.67)$$

For $j=2$, making the approximation $\omega_2\simeq 2$ (see Table 14.1), it follows from Equations (14.61), (14.62) and Table 14.2 that

$$x_2 \simeq \frac{1}{2}\,e^2, \qquad (14.68)$$

$$y_2 \simeq \frac{1}{4}\,e^2. \qquad (14.69)$$

Likewise, making the approximation $\Omega_2 \simeq 0$ (see Table 14.1), it follows from Equation (14.63) and Table 14.2 that

$$z_2 \simeq \frac{3}{2}\, e\, \iota. \tag{14.70}$$

For $j = 3$, making the approximation $\omega_3 \simeq 2$ (see Table 14.1), it follows from Equations (14.61), (14.62) and Table 14.2 that

$$x_3 \simeq \frac{1}{4}\,\iota^2, \tag{14.71}$$

$$y_3 \simeq -\frac{1}{4}\,\iota^2. \tag{14.72}$$

Likewise, making the approximation $\Omega_3 \simeq 2$ (see Table 14.1), it follows from Equation (14.63) and Table 14.2 that

$$z_3 \simeq \frac{1}{2}\, e\, \iota. \tag{14.73}$$

For $j = 4$, making the approximation $\omega_4 \simeq 2$ (see Table 14.1), it follows from Equations (14.61), (14.62) and Table 14.2 that

$$x_4 \simeq -m^2, \tag{14.74}$$

$$y_4 \simeq \frac{11}{8}\, m^2. \tag{14.75}$$

Thus, according to Table 14.2,

$$a_5 \simeq -\frac{9}{8}\, m^2\, e, \tag{14.76}$$

$$b_5 \simeq \frac{51}{16}\, m^2\, e, \tag{14.77}$$

$$c_5 \simeq \frac{3}{2}\, m^2\, \iota. \tag{14.78}$$

For $j = 5$, making the approximation $\omega_5 \simeq 1 - 2\,m$ (see Table 14.1), it follows from Equations (14.61), (14.62), (14.76), and (14.77) that

$$x_5 \simeq -\frac{15}{8}\, m\, e, \tag{14.79}$$

$$y_5 \simeq \frac{15}{4}\, m\, e. \tag{14.80}$$

Likewise, making the approximation $\Omega_5 \simeq 1 - 2\,m$ (see Table 14.1), it follows from Equations (14.63) and (14.78) that

$$z_5 \simeq \frac{3}{8}\, m\, \iota. \tag{14.81}$$

Thus, according to Table 14.2,

$$a_1 = -\frac{135}{64}\,m^3\,e, \tag{14.82}$$

$$b_1 = \frac{765}{128}\,m^3\,e, \tag{14.83}$$

$$c_1 = \frac{9}{16}\,m^3\,\iota. \tag{14.84}$$

Finally, for $j = 1$, by analogy with Equations (14.42)–(14.44), we expect

$$a_1 = -e, \tag{14.85}$$

$$b_1 = 2\,e, \tag{14.86}$$

$$c_1 = \iota. \tag{14.87}$$

Thus, since $\omega_1 = 1 + c\,m^2$ (see Table 14.1), it follows from Equations (14.61), (14.82), (14.83), and (14.85) that

$$-e \simeq \frac{-(225/16)\,m^3\,e}{-(3/2)\,m^2 - 2\,c\,m^2}, \tag{14.88}$$

which yields

$$c \simeq -\frac{3}{4} - \frac{225}{32}\,m + \mathcal{O}(m^2). \tag{14.89}$$

Likewise, since $\Omega_1 = 1 + g\,m^2$ (see Table 14.1), it follows from Equations (14.63), (14.84), and (14.87) that

$$\iota \simeq \frac{(9/16)\,m^3\,\iota}{(3/2)\,m^2 - 2\,g\,m^2}, \tag{14.90}$$

which yields

$$g \simeq \frac{3}{4} + \frac{9}{32}\,m + \mathcal{O}(m^2). \tag{14.91}$$

According to the above analysis, our final expressions for δX, δY, and δZ are

$$\begin{aligned}\delta X &= -\frac{1}{2}\,e^2 - \frac{1}{4}\,\iota^2 - e\,\cos[(1 + c\,m^2)\,T - \alpha_0] + \frac{1}{2}\,e^2\,\cos[2\,(1 + c\,m^2)\,T - 2\,\alpha_0]\\ &\quad + \frac{1}{4}\,\iota^2\,\cos[2\,(1 + g\,m^2)\,T - 2\,\gamma_0] - m^2\,\cos[2\,(1 - m)\,T]\\ &\quad - \frac{15}{8}\,m\,e\,\cos[(1 - 2\,m - c\,m^2)\,T + \alpha_0],\end{aligned} \tag{14.92}$$

$$\begin{aligned}\delta Y &= 2\,e\,\sin[(1 + c\,m^2)\,T - \alpha_0] + \frac{1}{4}\,e^2\,\sin[2\,(1 + c\,m^2)\,T - 2\,\alpha_0]\\ &\quad - \frac{1}{4}\,\iota^2\,\sin[2\,(1 + g\,m^2)\,T - 2\,\gamma_0] + \frac{11}{8}\,m^2\,\cos[2\,(1 - m)\,T]\end{aligned}$$

$$+\frac{15}{4}\,m\,e\,\sin[(1-2\,m-c\,m^2)\,T+\alpha_0], \tag{14.93}$$

$$\begin{aligned}\delta Z &= \iota\,\sin[(1+g\,m^2)\,T-\gamma_0]+\frac{3}{2}\,e\,\iota\,\sin[(c-g)\,m^2\,T-\alpha_0+\gamma_0]\\ &\quad+\frac{1}{2}\,e\,\iota\,\sin[(2+c\,m^2+g\,m^2)\,T-\alpha_0-\gamma_0]\\ &\quad+\frac{3}{8}\,m\,\iota\,\sin[(1-2\,m-g\,m^2)\,T+\gamma_0].\end{aligned} \tag{14.94}$$

Thus, making use of Equations (14.36)–(14.38), we find that

$$\begin{aligned}\frac{r}{a} &= 1-e\,\cos[(1+c\,m^2)\,T-\alpha_0]+\frac{1}{2}\,e^2-\frac{1}{6}\,m^2-\frac{1}{2}\,e^2\,\cos[2\,(1+c\,m^2)\,T-2\,\alpha_0]\\ &\quad-m^2\,\cos[2\,(1-m)\,T]-\frac{15}{8}\,m\,e\,\cos[(1-2\,m-c\,m^2)\,T+\alpha_0],\end{aligned} \tag{14.95}$$

$$\begin{aligned}\theta &= T+2\,e\,\sin[(1+c\,m^2)\,T-\alpha_0]+\frac{5}{4}\,e^2\,\sin[2\,(1+c\,m^2)\,T-2\,\alpha_0]\\ &\quad-\frac{1}{4}\,\iota^2\,\sin[2\,(1+g\,m^2)\,T-2\,\gamma_0]+\frac{11}{8}\,m^2\,\sin[2\,(1-m)\,T]\\ &\quad+\frac{15}{4}\,m\,e\,\sin[(1-2\,m-c\,m^2)\,T+\alpha_0],\end{aligned} \tag{14.96}$$

$$\begin{aligned}\beta &= \iota\,\sin[(1+g\,m^2)\,T-\gamma_0]+e\,\iota\,\sin[(c-g)\,m^2\,T-\alpha_0+\gamma_0]\\ &\quad+e\,\iota\,\sin[(2+c\,m^2+g\,m^2)\,T-\alpha_0-\gamma_0]\\ &\quad+\frac{3}{8}\,m\,\iota\,\sin[(1-2\,m-g\,m^2)\,T+\gamma_0].\end{aligned} \tag{14.97}$$

The above expressions are accurate up to *second-order* in the small parameters e, ι, and m.

14.6 Description of Lunar Motion

In order to better understand the perturbed lunar motion derived in the previous section, it is helpful to introduce the concept of the *mean moon*. This is an imaginary body which orbits the Earth, in the ecliptic plane, at a *steady* angular velocity that is equal to the Moon's mean orbital angular velocity, n. Likewise, the *mean sun* is a second imaginary body which orbits the Earth, in the ecliptic plane, at a *steady* angular velocity that is equal to the Sun's mean orbital angular velocity, n'. Thus, the ecliptic longitudes of the mean moon and the mean sun are

$$\bar{\theta} = n\,t, \tag{14.98}$$

$$\bar{\theta}' = n'\,t, \tag{14.99}$$

respectively. Here, for the sake of simplicity, and also for the sake of consistency with our previous analysis, we have assumed that both objects are located at ecliptic longitude 0° at time $t = 0$.

Now, from Equation (14.95), to first-order in small parameters, the lunar perigee corresponds to $(1 + c\,m^2)\,n\,t - \alpha_0 = j\,2\pi$, where j is an integer. However, this condition can also be written $\bar{\theta} = \alpha$, where

$$\alpha = \alpha_0 + \alpha_1\,n'\,t, \tag{14.100}$$

and, making use of Equation (14.89), together with the definition $m = n'/n$,

$$\alpha_1 = \frac{3}{4}\,m + \frac{255}{32}\,m^2 + \mathcal{O}(m^3). \tag{14.101}$$

Thus, we can identify α as the mean ecliptic longitude of the perigee. Moreover, according to Equation (14.100), the perigee *precesses* (*i.e.*, its longitude increases in time) at the mean rate of $360\,\alpha_1$ degrees per year. (Of course, a year corresponds to $\Delta t = 2\pi/n'$.) Furthermore, it is clear that this precession is entirely due to the perturbing influence of the Sun, since it only depends on the parameter m, which is a measure of this influence. Given that $m = 0.07480$, we find that the perigee advances by 34.36° degrees per year. Hence, we predict that the perigee completes a full circuit about the Earth every $1/\alpha_1 = 10.5$ years. In fact, the lunar perigee completes a full circuit every 8.85 years. Our prediction is somewhat inaccurate because our previous analysis neglected $\mathcal{O}(m^2)$, and smaller, contributions to the parameter c [see Equation (14.89)], and these turn out to be significant.

From Equation (14.97), to first-order in small parameters, the Moon passes through its ascending node when $(1 + g\,m^2)\,n\,t - \gamma_0 = \gamma_0 + j\,2\pi$, where j is an integer. However, this condition can also be written $\bar{\theta} = \gamma$, where

$$\gamma = \gamma_0 - \gamma_1\,n'\,t, \tag{14.102}$$

and, making use of Equation (14.91),

$$\gamma_1 = \frac{3}{4}\,m - \frac{9}{32}\,m^2 + \mathcal{O}(m^3). \tag{14.103}$$

Thus, we can identify γ as the mean ecliptic longitude of the ascending node. Moreover, according to Equation (14.102), the ascending node *regresses* (*i.e.*, its longitude decreases in time) at the mean rate of $360\,\gamma_1$ degrees per year. As before, it is clear that this regression is entirely due to the perturbing influence of the Sun. Moreover, we find that the ascending node retreats by 19.63° degrees per year. Hence, we predict that the ascending node completes a full circuit about the Earth every $1/\gamma_1 = 18.3$ years. In fact, the lunar ascending node completes a full circuit every 18.6 years, so our prediction is fairly accurate.

It is helpful to introduce the lunar *mean anomaly*,

$$M = \bar{\theta} - \alpha, \tag{14.104}$$

which is defined as the angular distance (in longitude) between the mean Moon and the perigee. It is also helpful to introduce the lunar *mean argument of latitude*,

$$F = \bar{\theta} - \gamma, \tag{14.105}$$

which is defined as the angular distance (in longitude) between the mean Moon and the ascending node. Finally, it is helpful to introduce the *mean elongation* of the Moon,

$$D = \bar{\theta} - \bar{\theta}', \tag{14.106}$$

which is defined as the difference between the longitudes of the mean Moon and the mean Sun.

When expressed in terms of M, F, and D, our previous expression (14.96) for the true ecliptic longitude of the Moon becomes

$$\theta = \bar{\theta} + \lambda, \tag{14.107}$$

where

$$\lambda = 2\,e\,\sin M + \frac{5}{4}\,e^2\,\sin 2M - \frac{1}{4}\,\iota^2\,\sin 2F + \frac{11}{8}\,m^2\,\sin 2D + \frac{15}{4}\,m\,e\,\sin(2D - M) \tag{14.108}$$

is the angular distance (in longitude) between the Moon and the mean Moon. The first three terms on the right-hand side of the above expression are Keplerian (*i.e.*, they are independent of the perturbing action of the Sun). In fact, the first is due to the eccentricity of the lunar orbit (*i.e.*, the fact that the geometric center of the orbit is slightly shifted from the Earth), the second is due to the ellipticity of the orbit (*i.e.*, the fact that the orbit is slightly non-circular), and the third is due to the slight inclination of the orbit to the ecliptic plane. However, the final two terms are caused by the perturbing action of the Sun. In fact, the fourth term corresponds to *variation*, whilst the fifth corresponds to *evection*. Note that variation attains its maximal amplitude around the so-called *octant points*, at which the Moon's disk is either one-quarter or three-quarters illuminated (*i.e.*, when $D = 45^\circ$, 135°, 225°, or 315°). Conversely, the amplitude of variation is zero around the so-called *quadrant points*, at which the Moon's disk is either fully illuminated, half illuminated, or not illuminated at all (*i.e.*, when $D = 0^\circ$, 90°, 180°, or 270°). Evection can be thought of as causing a slight reduction in the eccentricity of the lunar orbit around the times of the new moon and the full moon (*i.e.*, $D = 0^\circ$ and $D = 180^\circ$), and causing a corresponding slight increase in the eccentricity around the times of the first and last quarter moons (*i.e.*, $D = 90^\circ$ and $D = 270^\circ$). This follows because the evection term in Equation (14.108) augments the eccentricity term, $2\,e\,\sin M$, when $\cos 2D = -1$, and reduces the term when $\cos 2D = +1$. The variation and evection terms appearing in expression (14.108) oscillate sinusoidally with periods of half a synodic month,[3] or 14.8

[3]A synodic month, which is 29.53 days, is the mean period between successive new moons.

days, and 31.8 days, respectively. These periods are in good agreement with observations. Finally, the amplitudes of the variation and evection terms (calculated using $m = 0.07480$ and $e = 0.05488$) are 1630 and 3218 arc seconds, respectively. However, the observed amplitudes are 2370 and 4586 arc seconds, respectively. It turns out that our expressions for these amplitudes are somewhat inaccurate because, for the sake of simplicity, we have only calculated the lowest order (in m) contributions to them. Recall that we also neglected the slight eccentricity, $e' = 0.016711$, of the Sun's apparent orbit about the Earth in our calculation. In fact, the eccentricity of the solar orbit gives rise to a small addition term $-3\,m\,e'\,\sin M'$ on the right-hand side of (14.108), where M' is the Sun's mean anomaly. This term, which is known as the *annual equation*, oscillates with a period of a solar year, and has an amplitude of 772 arc seconds.

When expressed in terms of D and F our previous expression (14.97) for the ecliptic latitude of the Moon becomes

$$\beta = \iota\,\sin(F+\lambda) + \frac{3}{8}\,m\,\iota\,\sin(2\,D - F). \tag{14.109}$$

The first term on the right-hand side of this expression is Keplerian (*i.e.*, it is independent of the perturbing influence of the Sun). However, the final term, which is known as *evection in latitude*, is due to the Sun's action. Evection in latitude can be thought of as causing a slight increase in the inclination of the lunar orbit to the ecliptic at the times of the first and last quarter moons, and a slight decrease at the times of the new moon and the full moon. The evection in latitude term oscillates sinusoidally with a period of 32.3 days, and has an amplitude of 521 arc seconds. This period is in good agreement with observations. However, the observed amplitude of the evection in latitude term is 624 arc seconds. As before, our expression for the amplitude is somewhat inaccurate because we have only calculated the lowest order (in m) contribution.

15 The Chaotic Pendulum

15.1 Introduction

Up until now, we have only dealt with dynamical problems which are capable of analytic solution. Let us now investigate a problem which is quite intractable analytically, and in which meaningful progress can only be made via numerical methods.

15.2 Basic Problem

Consider a simple pendulum consisting of a point mass m, at the end of a light rigid rod of length l, attached to a fixed frictionless pivot which allows the rod (and the mass) to move freely under gravity in the vertical plane. Such a pendulum is sketched in Figure 15.1. Let us parameterize the instantaneous position of the pendulum via the angle θ the rod makes with the downward vertical. It is assumed that the pendulum is free to swing through a full circle. Hence, θ and $\theta + 2\pi$ both correspond to the same pendulum position.

The angular equation of motion of the pendulum is simply

$$m\,l\,\frac{d^2\theta}{dt^2} + m\,g\,\sin\theta = 0, \tag{15.1}$$

where g is the downward acceleration due to gravity—see Section 3.10. Suppose that the pendulum is embedded in a viscous medium (*e.g.*, air). Let us assume that the viscous drag torque acting on the pendulum is directly proportional to the pendulum's instantaneous velocity. It follows that, in the presence of viscous drag, the above equation generalizes to

$$m\,l\,\frac{d^2\theta}{dt^2} + \nu\,\frac{d\theta}{dt} + m\,g\,\sin\theta = 0, \tag{15.2}$$

where ν is a positive constant parameterizing the viscosity of the medium in question—see Section 3.5. Of course, viscous damping will eventually drain all energy from the pendulum, leaving it in a stationary state. In order to maintain the motion against viscosity, it is necessary to add some external driving. For the sake of simplicity, we choose a fixed amplitude periodic drive (which could arise, for instance, via periodic oscillations of the pendulum's pivot point). Thus, the final equation of motion of the pendulum is written

$$m\,l\,\frac{d^2\theta}{dt^2} + \nu\,\frac{d\theta}{dt} + m\,g\,\sin\theta = A\,\cos(\omega\,t), \tag{15.3}$$

where A and ω are constants parameterizing the amplitude and frequency of the external driving torque, respectively.

Let

$$\omega_0 = \sqrt{\frac{g}{l}}. \tag{15.4}$$

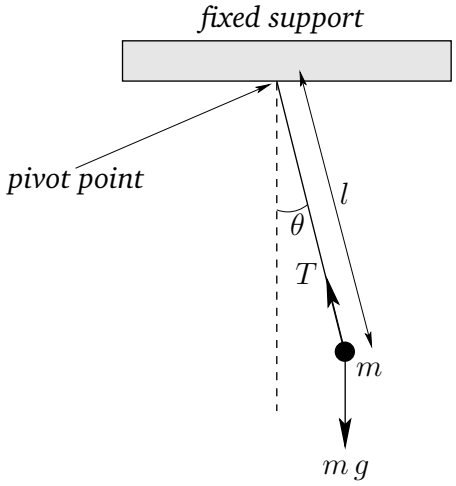


Figure 15.1: *A simple pendulum.*

Of course, we recognize ω_0 as the natural frequency of small amplitude oscillations of the pendulum. We can conveniently normalize the pendulum's equation of motion by writing,

$$\hat{t} = \omega_0\, t, \tag{15.5}$$

$$\hat{\omega} = \frac{\omega}{\omega_0}, \tag{15.6}$$

$$Q = \frac{m\, g}{\omega_0\, \nu}, \tag{15.7}$$

$$\hat{A} = \frac{A}{m\, g}, \tag{15.8}$$

in which case Equation (15.3) becomes

$$\frac{d^2\theta}{d\hat{t}^2} + \frac{1}{Q}\,\frac{d\theta}{d\hat{t}} + \sin\theta = \hat{A}\,\cos(\hat{\omega}\,\hat{t}). \tag{15.9}$$

From now on, the hats on normalized quantities will be omitted, for ease of notation. Note that, in normalized units, the natural frequency of small amplitude oscillations is *unity*. Moreover, Q is the familiar *quality-factor*, which is, roughly speaking, the number of oscillations of the undriven system which must elapse before its energy is significantly reduced via the action of viscosity—see Section 3.6. The quantity A is the amplitude of the external torque measured in units of the maximum possible gravitational torque. Finally, ω is the oscillation frequency of the external torque measured in units of the pendulum's natural frequency.

Equation (15.9) is clearly a second-order ordinary differential equation. It can, therefore, also be written as two coupled first-order ordinary differential equations:

$$\frac{d\theta}{dt} = \nu, \tag{15.10}$$

$$\frac{d\nu}{dt} = -\frac{\nu}{Q} - \sin\theta + A\,\cos(\omega\,t). \tag{15.11}$$

15.3 Analytic Solution

Before attempting a numerical solution of the equations of motion of any dynamical system, it is a good idea to, first, investigate the equations as thoroughly as possible via standard analytic techniques. Unfortunately, Equations (15.10) and (15.11) constitute a *non-linear* dynamical system—because of the presence of the $\sin\theta$ term on the right-hand side of Equation (15.11). This system, like most non-linear systems, does not possess a simple analytic solution. Fortunately, however, if we restrict our attention to *small amplitude* oscillations, such that the approximation

$$\sin\theta \simeq \theta \tag{15.12}$$

is valid, then the system becomes *linear*, and can easily be solved analytically.

The linearized equations of motion of the pendulum take the form:

$$\frac{d\theta}{dt} = \nu, \tag{15.13}$$

$$\frac{d\nu}{dt} = -\frac{\nu}{Q} - \theta + A\,\cos(\omega\,t). \tag{15.14}$$

Suppose that the pendulum's position, $\theta(0)$, and velocity, $\nu(0)$, are specified at time $t = 0$. As is well-known, in this case, the above equations of motion can be solved analytically to give:

$$\begin{aligned}\theta(t) &= \left\{\theta(0) - \frac{A\,(1-\omega^2)}{[(1-\omega^2)^2+\omega^2/Q^2]}\right\} e^{-t/2Q}\,\cos(\omega_*\,t)\\ &+\frac{1}{\omega_*}\left\{\nu(0) + \frac{\theta(0)}{2Q} - \frac{A\,(1-3\,\omega^2)/2Q}{[(1-\omega^2)^2+\omega^2/Q^2]}\right\} e^{-t/2Q}\,\sin(\omega_*\,t)\\ &+\frac{A\left[(1-\omega^2)\,\cos(\omega\,t) + (\omega/Q)\,\sin(\omega\,t)\right]}{[(1-\omega^2)^2+\omega^2/Q^2]},\end{aligned} \tag{15.15}$$

$$\begin{aligned}\nu(t) &= \left\{\nu(0) - \frac{A\,\omega^2/Q}{[(1-\omega^2)^2+\omega^2/Q^2]}\right\} e^{-t/2Q}\,\cos\omega_* t\\ &-\frac{1}{\omega_*}\left\{\theta(0) + \frac{\nu(0)}{2Q} - \frac{A\,[(1-\omega^2)-\omega^2/2Q^2]}{[(1-\omega^2)^2+\omega^2/Q^2]}\right\} e^{-t/2Q}\,\sin(\omega_*\,t)\\ &+\frac{\omega\,A\left[-(1-\omega^2)\,\sin(\omega\,t) + (\omega/Q)\,\cos(\omega\,t)\right]}{[(1-\omega^2)^2+\omega^2/Q^2]}.\end{aligned} \tag{15.16}$$

Here,

$$\omega_* = \sqrt{1 - \frac{1}{4Q^2}}, \tag{15.17}$$

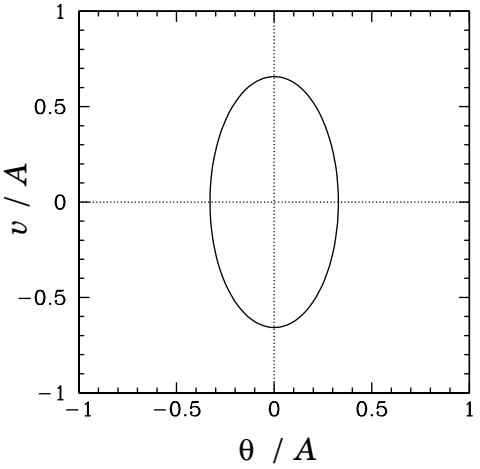


Figure 15.2: *A phase-space plot of the periodic attractor for a linear damped periodically driven pendulum. Data calculated analytically for* $Q = 4$ *and* $\omega = 2$.

and it is assumed that $Q > 1/2$. It can be seen that the above expressions for θ and v both consist of *three* terms. The first two terms clearly represent *transients*, since they depend on the initial conditions, and *decay* exponentially in time—see Section 3.9. In fact, the e-folding time for the decay of these terms is $2\,Q$ (in normalized time units). The final term represents the *time-asymptotic motion* of the pendulum, and is manifestly *independent* of the initial conditions—see Section 3.5.

It is often convenient to *visualize* the motion of a dynamical system as an orbit, or trajectory, in *phase-space*, which is defined as the space of all of the dynamical variables required to specify the instantaneous state of the system. For the case in hand, there are two dynamical variables, v and θ, and so phase-space corresponds to the θ-v plane. Note that each different point in this plane corresponds to a unique instantaneous state of the pendulum. [Strictly speaking, we should also consider t to be a dynamical variable, since it appears explicitly on the right-hand side of Equation (15.11).]

It is clear, from Equations (15.15) and (15.16), that if we wait long enough for all of the transients to decay away then the motion of the pendulum settles down to the following simple orbit in phase-space:

$$\theta(t) = \frac{A\left[(1-\omega^2)\,\cos(\omega\,t) + (\omega/Q)\,\sin(\omega\,t)\right]}{[(1-\omega^2)^2 + \omega^2/Q^2]}, \tag{15.18}$$

$$v(t) = \frac{\omega\,A\left[-(1-\omega^2)\,\sin(\omega\,t) + (\omega/Q)\,\cos(\omega\,t)\right]}{[(1-\omega^2)^2 + \omega^2/Q^2]}. \tag{15.19}$$

This orbit traces out the closed curve

$$\left(\frac{\theta}{\tilde{A}}\right)^2 + \left(\frac{v}{\omega\,\tilde{A}}\right)^2 = 1, \tag{15.20}$$

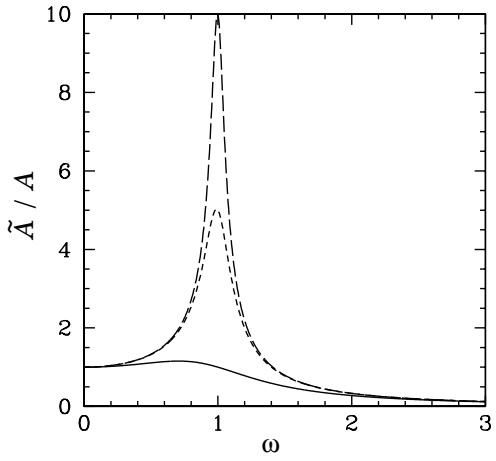


Figure 15.3: *The maximum angular displacement of a linear damped periodically driven pendulum as a function of driving frequency. The solid curve corresponds to* $Q = 1$. *The short-dashed curve corresponds to* $Q = 5$. *The long-dashed curve corresponds to* $Q = 10$. *Analytic data.*

in phase-space, where

$$\tilde{A} = \frac{A}{\sqrt{(1-\omega^2)^2 + \omega^2/Q^2}}. \tag{15.21}$$

As illustrated in Figure 15.2, this curve is an *ellipse* whose principal axes are aligned with the v and θ coordinate axes. Observe that the curve is *closed*, which suggests that the associated motion is *periodic* in time. In fact, the motion repeats itself exactly every

$$\tau = \frac{2\pi}{\omega} \tag{15.22}$$

normalized time units. The maximum angular displacement of the pendulum from its undriven rest position ($\theta = 0$) is $\tilde{A}$. As illustrated in Figure 15.3, the variation of $\tilde{A}$ with driving frequency ω [see Equation (15.21)] displays all of the features of a classic resonance curve. The maximum amplitude of the driven oscillation is proportional to the quality-factor, Q, and is achieved when the driving frequency matches the natural frequency of the pendulum (*i.e.*, when $|\omega| = 1$). Moreover, the width of the resonance in ω-space is proportional to $1/Q$—see Section 3.7.

The phase-space curve shown in Figure 15.2 is called a *periodic attractor*. It is termed an "attractor" because, irrespective of the initial conditions, the trajectory of the system in phase-space tends asymptotically to—in other words, is attracted to—this curve as $t \rightarrow \infty$. This gravitation of phase-space trajectories towards the attractor is illustrated in Figures 15.4 and 15.5. Of course, the attractor is termed "periodic" because it corresponds to motion which is periodic in time.

Let us summarize our findings so far. We have discovered that if a damped pendulum is subject to a low amplitude periodic drive then its *time-asymptotic* response (*i.e.*, its

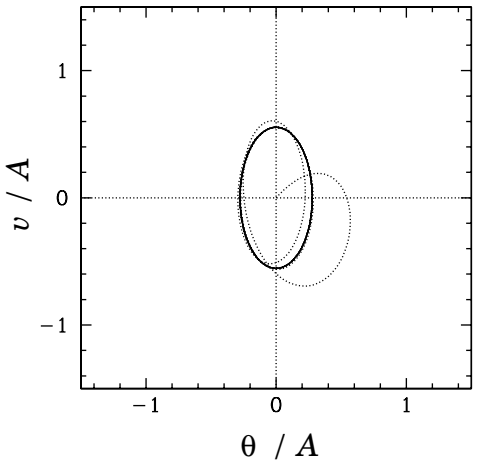


Figure 15.4: *The phase-space trajectory of a linear damped periodically driven pendulum. Data calculated analytically for* $Q = 1$ *and* $\omega = 2$*. Here,* $v(0)/A = 0$ *and* $\theta(0)/A = 0$*.*

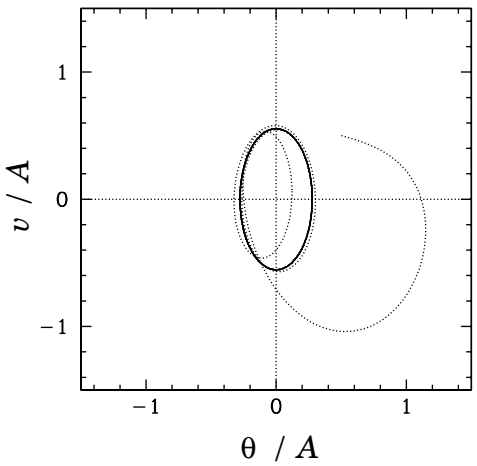


Figure 15.5: *The phase-space trajectory of a linear damped periodically driven pendulum. Data calculated analytically for* $Q = 1$ *and* $\omega = 2$*. Here,* $v(0)/A = 0.5$ *and* $\theta(0)/A = 0.5$*.*

response after any transients have died away) is *periodic*, with the same period as the driving torque. Moreover, the response exhibits *resonant* behaviour as the driving frequency approaches the natural frequency of oscillation of the pendulum. The amplitude of the resonant response, as well as the width of the resonant window, is governed by the amount of damping in the system. After a little reflection, we can easily appreciate that all of these results are a direct consequence of the *linearity* of the pendulum's equations of motion in the low amplitude limit. In fact, it is easily demonstrated that the time-asymptotic response of *any* intrinsically stable linear system (with a discrete spectrum of normal modes) to a periodic drive is periodic, with the same period as the drive. Moreover, if the driving frequency approaches one of the natural frequencies of oscillation of the system then the response exhibits resonant behaviour. But, is this the only allowable time-asymptotic response of a dynamical system to a periodic drive? It turns out that it is not. Indeed, the response of a *non-linear* system to a periodic drive is generally much richer and far more diverse than simple periodic motion. Since the majority of naturally occurring dynamical systems are non-linear, it is clearly important that we gain a basic understanding of this phenomenon. Unfortunately, we cannot achieve this goal via a standard analytic approach, since non-linear equations of motion generally do not possess simple analytic solutions. Instead, we must employ *numerical methods*. As an example, let us investigate the dynamics of a damped pendulum, subject to a periodic drive, with *no restrictions* on the amplitude of the pendulum's motion.

15.4 Numerical Solution

In the following, we present numerical solutions of Equations (15.10) and (15.11) obtained using a fairly standard fixed step-length fourth-order Runge-Kutta integration scheme.[1]

15.5 Poincaré Section

For the sake of definiteness, let us fix the normalized amplitude and frequency of the external drive to be $A = 1.5$ and $\omega = 2/3$, respectively.[2] Furthermore, let us investigate any changes which may develop in the nature of the pendulum's time-asymptotic motion as the quality-factor Q is varied. Of course, if Q is made sufficiently small (*i.e.*, if the pendulum is embedded in a sufficiently viscous medium) then we expect the amplitude of the pendulum's time-asymptotic motion to become low enough that the linear analysis outlined in Section 15.3 is valid. Indeed, we expect non-linear effects to manifest themselves as Q is gradually made larger, and the amplitude of the pendulum's motion consequently increases to such an extent that the small angle approximation breaks down.

[1] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical recipes in C: The art of scientific computing*, 2nd Edition (Cambridge University Press, Cambridge UK, 1992), Section 16.1.

[2] G.L. Baker, *Control of the chaotic driven pendulum*, Am. J. Phys. **63**, 832 (1995).

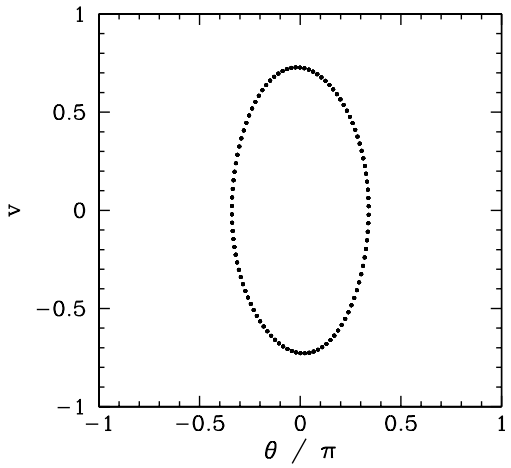


Figure 15.6: *Equally spaced (in time) points on a time-asymptotic orbit in phase-space. Data calculated numerically for* $Q = 0.5$*,* $A = 1.5$*,* $\omega = 2/3$*,* $\theta(0) = 0$*, and* $\nu(0) = 0$*.*

Figure 15.6 shows a time-asymptotic orbit in phase-space calculated numerically for a case where Q is sufficiently small (*i.e.*, $Q = 1/2$) that the small angle approximation holds reasonably well. Not surprisingly, the orbit is very similar to the analytic orbits described in Section 15.3. The fact that the orbit consists of a *single* loop, and forms a *closed* curve in phase-space, strongly suggests that the corresponding motion is periodic with the same period as the external drive—we term this type of motion *period-1* motion. More generally, period-n motion consists of motion which repeats itself exactly every n periods of the external drive (and, obviously, does not repeat itself on any time-scale less than n periods). Of course, period-1 motion is the only allowed time-asymptotic motion in the small angle limit.

It would certainly be helpful to possess a graphical test for period-n motion. In fact, such a test was developed more than a hundred years ago by the French mathematician Henry Poincaré. Nowadays, it is called a *Poincaré section* in his honour. The idea of a Poincaré section, as applied to a periodically driven pendulum, is very simple. As before, we calculate the time-asymptotic motion of the pendulum, and visualize it as a series of points in θ-ν phase-space. However, we only plot *one point per period* of the external drive. To be more exact, we only plot a point when

$$\omega\, t = \phi + k\, 2\pi \tag{15.23}$$

where k is any integer, and ϕ is referred to as the *Poincaré phase*. For period-1 motion, in which the motion repeats itself exactly every period of the external drive, we expect the Poincaré section to consist of only *one* point in phase-space (*i.e.*, we expect all of the points to plot on top of one another). Likewise, for period-2 motion, in which the motion repeats itself exactly every two periods of the external drive, we expect the Poincaré section to consist of *two* points in phase-space (*i.e.*, we expect alternating points to plot on top of one another). Finally, for period-n motion we expect the Poincaré section to consist of n points

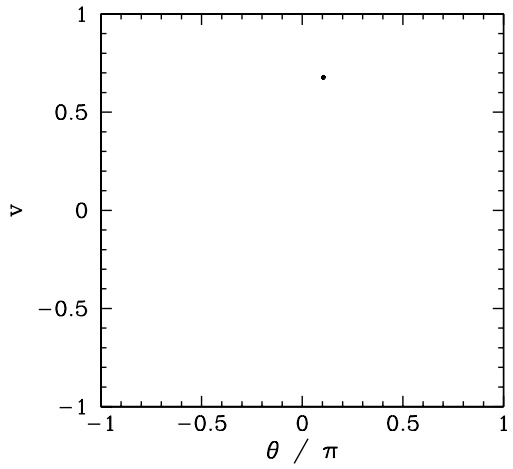


Figure 15.7: *The Poincaré section of a time-asymptotic orbit. Data calculated numerically for* $Q = 0.5$*,* $A = 1.5$*,* $\omega = 2/3$*,* $\theta(0) = 0$*,* $\nu(0) = 0$*, and* $\phi = 0$*.*

in phase-space.

Figure 15.7 displays the Poincaré section of the orbit shown in Figure 15.6. The fact that the section consists of a single point confirms that the motion displayed in Figure 15.6 is indeed period-1 motion.

15.6 Spatial Symmetry Breaking

Suppose that we now gradually increase the quality-factor Q. What happens to the simple orbit shown in Figure 15.6? It turns out that, at first, nothing particularly exciting happens. The size of the orbit gradually increases, indicating a corresponding increase in the amplitude of the pendulum's motion, but the general nature of the motion remains unchanged. However, something interesting does occur when Q is increased beyond about 1.2. Figure 15.8 shows the ν-coordinate of the orbit's Poincaré section plotted against Q in the range 1.2 and 1.3. Note the sharp downturn in the curve at $Q \simeq 1.245$. What does this signify? Well, Figure 15.9 shows the time-asymptotic phase-space orbit just before the downturn (*i.e.*, at $Q = 1.24$), and Figure 15.10 shows the orbit somewhat after the downturn (*i.e.*, at $Q = 1.30$). It is clear that the downturn is associated with a sudden change in the nature of the pendulum's time-asymptotic phase-space orbit. Prior to the downturn, the orbit spends as much time in the region $\theta < 0$ as in the region $\theta > 0$. However, after the downturn the orbit spends the majority of its time in the region $\theta < 0$. In other words, after the downturn, the pendulum bob favours the region to the *left* of the pendulum's vertical. This is somewhat surprising, since there is nothing in the pendulum's equations of motion which differentiates between the regions to the left and to the right of the vertical. We refer to a solution of this type—*i.e.*, one which fails to realize the full symmetry of the dynamical system in question—as a *symmetry breaking* solution. In this case, because the

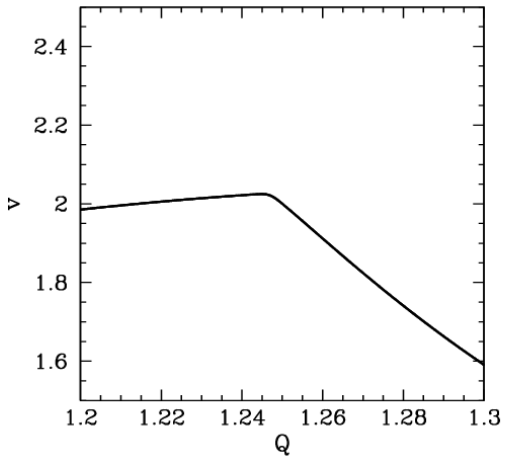


Figure 15.8: *The ν-coordinate of the Poincaré section of a time-asymptotic orbit plotted against the quality-factor* Q. *Data calculated numerically for* $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, $\nu(0) = 0$, *and* $\phi = 0$.

particular symmetry which is broken is a *spatial* symmetry, we refer to the process by which the symmetry breaking solution suddenly appears, as the control parameter Q is adjusted, as *spatial symmetry breaking*. Needless to say, spatial symmetry breaking is an intrinsically *non-linear* process—it cannot take place in dynamical systems possessing linear equations of motion.

It stands to reason that since the pendulum's equations of motion favour neither the left nor the right then the left-favouring orbit pictured in Figure 15.10 must be accompanied by a mirror image right-favouring orbit. How do we obtain this mirror image orbit? It turns out that all we have to do is keep the physical parameters Q, A, and ω fixed, but *change the initial conditions* $\theta(0)$ and $\nu(0)$. Figure 15.11 shows a time-asymptotic phase-space orbit calculated with the same physical parameters used in Figure 15.10, but with the initial conditions $\theta(0) = 0$ and $\nu(0) = -3$, instead of $\theta(0) = 0$ and $\nu(0) = 0$. It can be seen that the orbit is indeed the mirror image of that pictured in Figure 15.10.

Figure 15.12 shows the ν-coordinate of the Poincaré section of a time-asymptotic orbit, calculated with the same physical parameters used in Figure 15.8, versus Q in the range 1.2 and 1.3. Data is shown for the two sets of initial conditions discussed above. The figure is interpreted as follows. When Q is less than a critical value, which is about 1.245, then the two sets of initial conditions lead to motions which converge on the *same* left-right symmetric period-1 attractor. However, once Q exceeds the critical value then the attractor *bifurcates* into two asymmetric mirror image period-1 attractors. Obviously, the bifurcation is indicated by the forking of the curve shown in Figure 15.12. The lower and upper branches correspond to the left- and right-favouring attractors, respectively.

Spontaneous symmetry breaking, which is the fundamental non-linear process illustrated in the above discussion, plays an important role in many areas of physics. For instance, symmetry breaking gives mass to elementary particles in the unified theory of

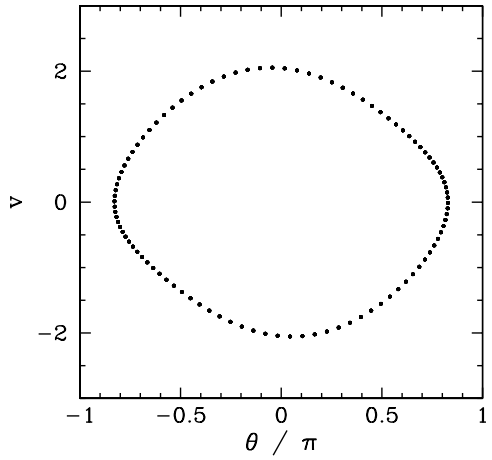


Figure 15.9: *Equally spaced (in time) points on a time-asymptotic orbit in phase-space. Data calculated numerically for* $Q = 1.24$*,* $A = 1.5$*,* $\omega = 2/3$*,* $\theta(0) = 0$*, and* $v(0) = 0$*.*

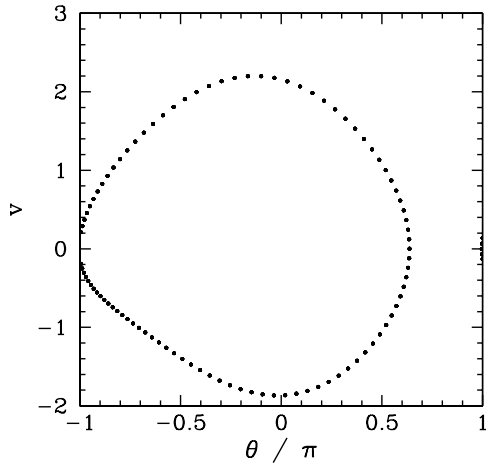


Figure 15.10: *Equally spaced (in time) points on a time-asymptotic orbit in phase-space. Data calculated numerically for* $Q = 1.30$*,* $A = 1.5$*,* $\omega = 2/3$*,* $\theta(0) = 0$*, and* $v(0) = 0$*.*

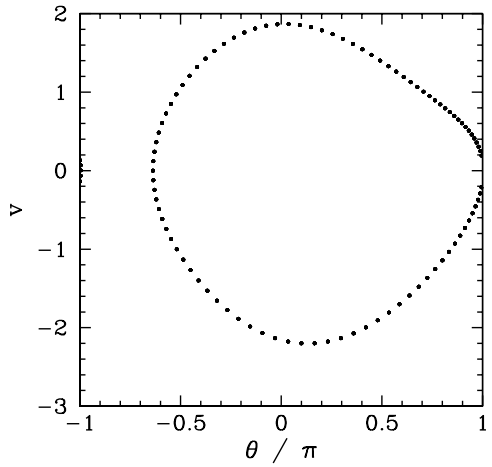


Figure 15.11: *Equally spaced (in time) points on a time-asymptotic orbit in phase-space. Data calculated numerically for* $Q = 1.30$, $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, *and* $\nu(0) = -3$.

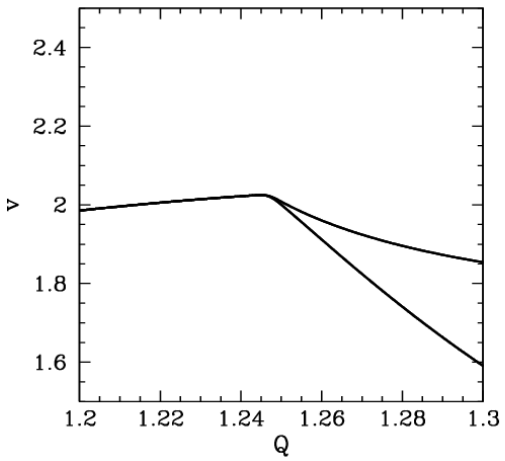


Figure 15.12: *The* ν*-coordinate of the Poincaré section of a time-asymptotic orbit plotted against the quality-factor* Q. *Data calculated numerically for* $A = 1.5$, *and* $\omega = 2/3$. *Data is shown for two sets of initial conditions:* $\theta(0) = 0$ *and* $\nu(0) = 0$ *(lower branch); and* $\theta(0) = 0$ *and* $\nu(0) = -3$ *(upper branch).*

electromagnetic and weak interactions.[3] Symmetry breaking also plays a pivotal role in the so-called "inflation" theory of the expansion of the early universe.[4]

15.7 Basins of Attraction

We have seen that when $Q = 1.3$, $A = 1.5$, and $\omega = 2/3$ there are two co-existing period-1 attractors in θ–ν phase-space. The time-asymptotic trajectory of the pendulum through phase-space converges on one or other of these attractors depending on the initial conditions: *i.e.*, depending on the values of $\theta(0)$ and $\nu(0)$. Let us define the *basin of attraction* of a given attractor as the locus of all points in the $\theta(0)$–$\nu(0)$ plane which lead to motion which ultimately converges on that attractor. We have seen that in the low-amplitude (*i.e.*, linear) limit (see Section 15.3) there is only a single period-1 attractor in phase-space, and all possible initial conditions lead to motion which converges on this attractor. In other words, the basin of attraction for the low-amplitude attractor constitutes the entire $\theta(0)$–$\nu(0)$ plane. The present case, in which there are *two* co-existing attractors in phase-space, is somewhat more complicated.

Figure 15.13 shows the basins of attraction, in $\theta(0)$–$\nu(0)$ space, of the asymmetric mirror image attractors pictured in Figures 15.10 and 15.11. The basin of attraction of the left-favoring attractor shown in Figure 15.10 is coloured black, whereas the basin of attraction of the right-favoring attractor shown in Figure 15.11 is coloured white. It can be seen that the two basins form a complicated interlocking pattern. Since we can identify the angles π and $-\pi$, the right-hand edge of the pattern connects smoothly with its left-hand edge. In fact, we can think of the pattern as existing on the surface of a *cylinder*.

Suppose that we take a diagonal from the bottom left-hand corner of Figure 15.13 to its top right-hand corner. This diagonal is intersected by a number of black bands of varying thickness. Observe that the two narrowest bands (*i.e.*, the fourth band from the bottom left-hand corner and the second band from the upper right-hand corner) both exhibit structure which is not very well resolved in the present picture.

Figure 15.14 is a blow-up of a region close to the lower left-hand corner of Figure 15.13. It can be seen that the unresolved band in the latter figure (*i.e.*, the second and third bands from the right-hand side in the former figure) actually consists of a closely spaced *pair* of bands. Note, however, that the narrower of these two bands exhibits structure which is not very well resolved in the present picture.

Figure 15.15 is a blow-up of a region of Figure 15.14. It can be seen that the unresolved band in the latter figure (*i.e.*, the first and second bands from the left-hand side in the former figure) actually consists of a closely spaced *pair* of bands. Note, however, that the broader of these two bands exhibits structure which is not very well resolved in the present picture.

[3] E.S. Albers and B.W. Lee, Phys. Rep. 9**C**, 1 (1973).

[4] P. Coles, and F. Lucchin, *Cosmology: The origin and evolution of cosmic structure*, (J. Wiley & Sons, Chichester UK, 1995).

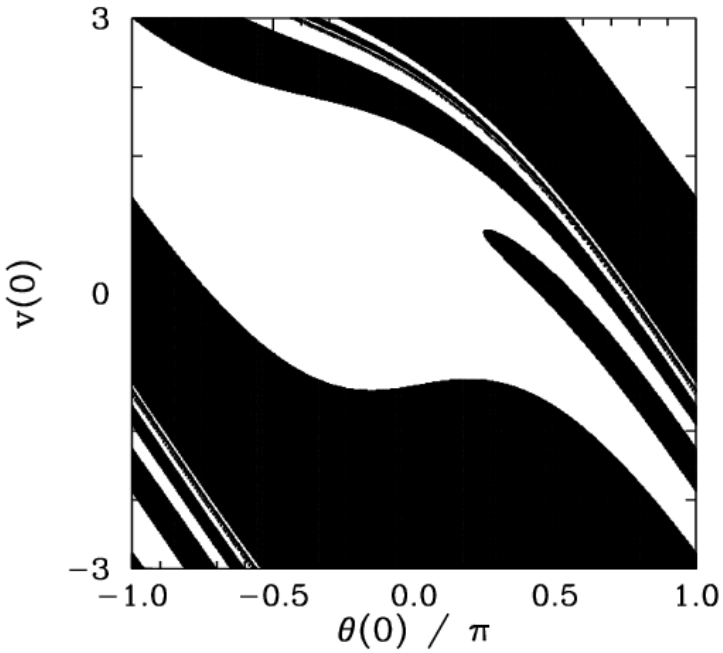


Figure 15.13: *The basins of attraction for the asymmetric, mirror image, attractors pictured in Figures 15.10 and 15.11. Regions of* $\theta(0)$–$v(0)$ *space which lead to motion converging on the left-favouring attractor shown in Figure 15.10 are coloured white: regions of* $\theta(0)$–$v(0)$ *space which lead to motion converging on the right-favouring attractor shown in Figure 15.11 are coloured black. Data calculated numerically for* $Q = 1.3$, $A = 1.5$, $\omega = 2/3$, *and* $\phi = 0$.

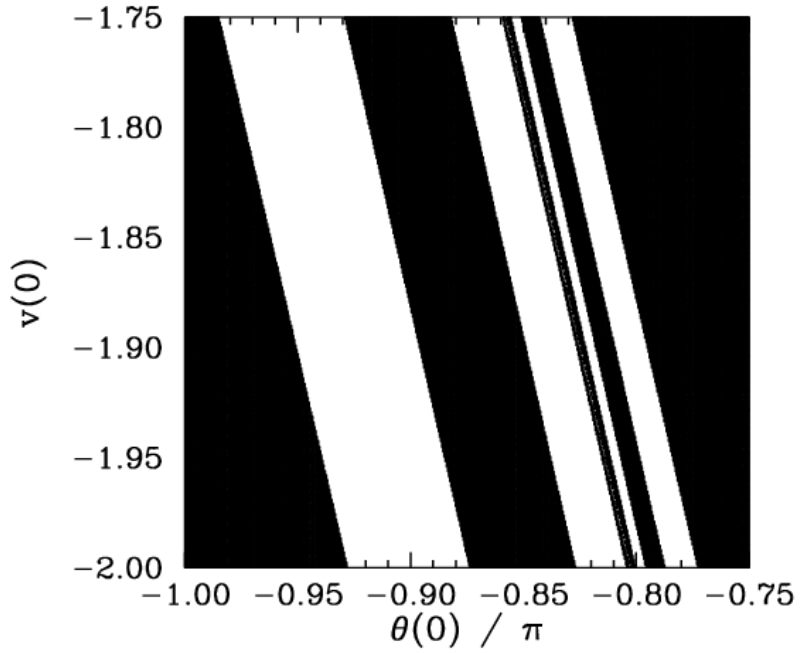


Figure 15.14: *Detail of the basins of attraction for the asymmetric, mirror image, attractors pictured in Figures 15.10 and 15.11. Regions of* $\theta(0)$–$v(0)$ *space which lead to motion converging on the left-favouring attractor shown in Figure 15.10 are coloured white: regions of* $\theta(0)$–$v(0)$ *space which lead to motion converging on the right-favouring attractor shown in Figure 15.11 are coloured black. Data calculated numerically for* $Q = 1.3$, $A = 1.5$, $\omega = 2/3$, *and* $\phi = 0$.

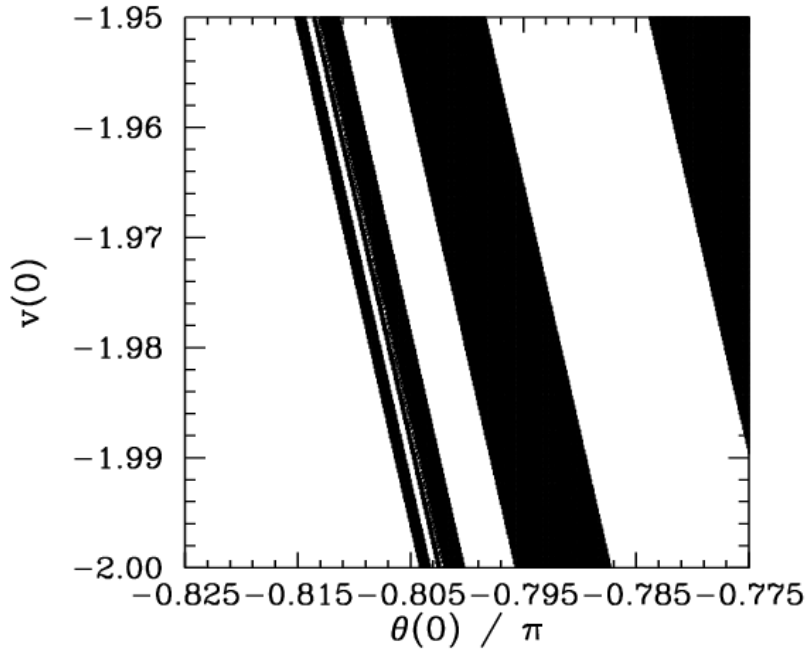


Figure 15.15: *Detail of the basins of attraction for the asymmetric, mirror image, attractors pictured in Figures 15.10 and 15.11. Regions of* $\theta(0)$–$v(0)$ *space which lead to motion converging on the left-favouring attractor shown in Figure 15.10 are coloured white: regions of* $\theta(0)$–$v(0)$ *space which lead to motion converging on the right-favouring attractor shown in Figure 15.11 are coloured black. Data calculated numerically for* $Q = 1.3$, $A = 1.5$, $\omega = 2/3$, *and* $\phi = 0$.

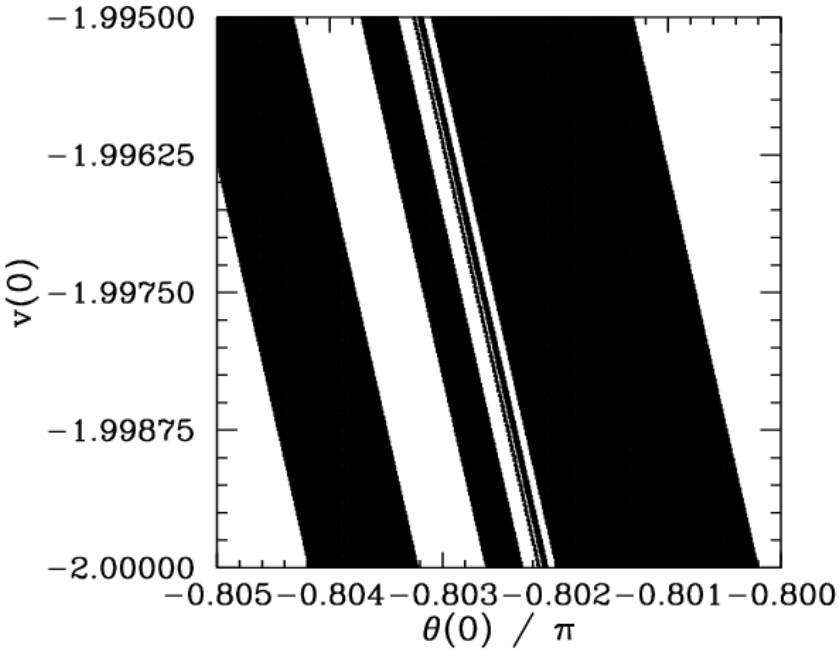


Figure 15.16: *Detail of the basins of attraction for the asymmetric, mirror image, attractors pictured in Figures 15.10 and 15.11. Regions of* $\theta(0)$–$v(0)$ *space which lead to motion converging on the left-favouring attractor shown in Figure 15.10 are coloured white: regions of* $\theta(0)$–$v(0)$ *space which lead to motion converging on the right-favouring attractor shown in Figure 15.11 are coloured black. Data calculated numerically for* $Q = 1.3$, $A = 1.5$, $\omega = 2/3$, *and* $\phi = 0$.

Figure 15.16 is a blow-up of a region of Figure 15.15. It can be seen that the unresolved band in the latter figure (*i.e.*, the first, second, and third bands from the right-hand side in the former figure) actually consists of a closely spaced *triplet* of bands. Note, however, that the narrowest of these bands exhibits structure which is not very well resolved in the present picture.

It should be clear, by this stage, that no matter how closely we look at Figure 15.13 we are going to find structure which we cannot resolve. In other words, the separatrix between the two basins of attraction shown in this figure is a curve which exhibits structure *at all scales*. Mathematicians have a special term for such a curve—they call it a *fractal*.[5]

Many people think of fractals as mathematical toys whose principal use is the generation of pretty pictures. However, it turns out that there is a close connection between fractals and the dynamics of non-linear systems—particularly systems which exhibit chaotic dynamics. We have just seen an example in which the boundary between the basins of attraction of two co-existing attractors in phase-space is a fractal curve. This turns out to be a fairly general result: *i.e.*, when multiple attractors exist in phase-space the separatrix between their various basins of attraction is invariably fractal. What is this telling us about the nature of non-linear dynamics? Well, returning to Figure 15.13, we can see that in the region of phase-space in which the fractal behaviour of the separatrix manifests itself most strongly (*i.e.*, the region where the light and dark bands fragment) the system exhibits abnormal sensitivity to initial conditions. In other words, we only have to change the initial conditions slightly (*i.e.*, so as to move from a dark to a light band, or *vice versa*) in order to significantly alter the time-asymptotic motion of the pendulum (*i.e.*, to cause the system to converge to a left-favouring instead of a right-favouring attractor, or *vice versa*). Fractals and extreme sensitivity to initial conditions are themes which will reoccur in our investigation of non-linear dynamics.

15.8 Period-Doubling Bifurcations

Let us now return to Figure 15.8. Recall, that as the quality-factor Q is gradually increased, the time-asymptotic orbit of the pendulum through phase-space undergoes a sudden transition, at $Q \simeq 1.245$, from a left-right symmetric period-1 orbit to a left-favouring period-1 orbit. What happens if we continue to increase Q? Figure 15.17 is basically a continuation of Figure 15.8. It can be seen that as Q is increased the left-favouring period-1 orbit gradually evolves until a critical value of Q, which is about 1.348, is reached. When Q exceeds this critical value the nature of the orbit undergoes another sudden change: this time from a left-favouring period-1 orbit to a left-favouring *period-2* orbit. Obviously, the change is indicated by the forking of the curve in Figure 15.17. This type of transition is termed a *period-doubling bifurcation*, since it involves a sudden doubling of the repetition period of the pendulum's time-asymptotic motion.

We can represent period-1 motion schematically as $AAAAAA\cdots$, where A represents a pattern of motion which is repeated every period of the external drive. Likewise, we

[5]B.B. Mandelbrot, *The fractal geometry of nature*, (W.H. Freeman, New York NY, 1982).

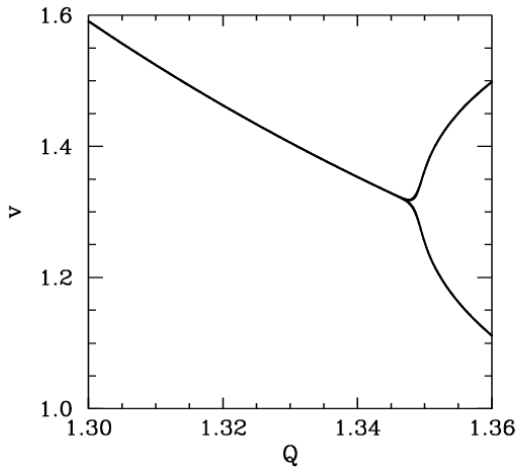


Figure 15.17: *The ν-coordinate of the Poincaré section of a time-asymptotic orbit plotted against the quality-factor* Q. *Data calculated numerically for* $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, $\nu(0) = 0$, *and* $\phi = 0$.

can represent period-2 motion as $ABABAB\cdots$, where A and B represent *distinguishable* patterns of motion which are repeated every alternate period of the external drive. A period-doubling bifurcation is represented: $AAAAAA\cdots \rightarrow ABABAB\cdots$. Clearly, all that happens in such a bifurcation is that the pendulum suddenly decides to do something slightly different in alternate periods of the external drive.

Figure 15.18 shows the time-asymptotic phase-space orbit of the pendulum calculated for a value of Q somewhat higher than that required to trigger the above mentioned period-doubling bifurcation. It can be seen that the orbit is left-favouring (*i.e.*, it spends the majority of its time on the left-hand side of the plot), and takes the form of a closed curve consisting of *two* interlocked loops in phase-space. Recall that for period-1 orbits there was only a single closed loop in phase-space. Figure 15.19 shows the Poincaré section of the orbit shown in Figure 15.18. The fact that the section consists of *two* points confirms that the orbit does indeed correspond to period-2 motion.

A period-doubling bifurcation is an example of *temporal symmetry breaking*. The equations of motion of the pendulum are invariant under the transformation $t \rightarrow t + \tau$, where τ is the period of the external drive. In the low amplitude (*i.e.*, linear) limit, the time-asymptotic motion of the pendulum always respects this symmetry. However, as we have just seen, in the non-linear regime it is possible to obtain solutions which spontaneously break this symmetry. Obviously, motion which repeats itself every two periods of the external drive is not as temporally symmetric as motion which repeats every period of the drive.

Figure 15.20 is essentially a continuation of Fig 15.12. Data is shown for two sets of initial conditions which lead to motions converging on left-favouring (lower branch) and right-favouring (upper branch) periodic attractors. We have already seen that the left-favouring periodic attractor undergoes a period-doubling bifurcation at $Q = 1.348$. It is

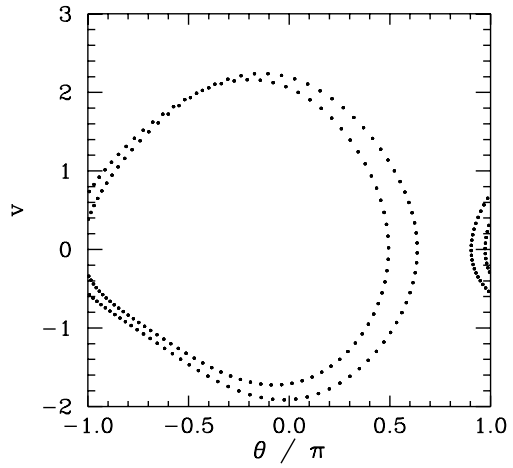


Figure 15.18: *Equally spaced (in time) points on a time-asymptotic orbit in phase-space. Data calculated numerically for* $Q = 1.36$*,* $A = 1.5$*,* $\omega = 2/3$*,* $\theta(0) = 0$*, and* $v(0) = -3$*.*

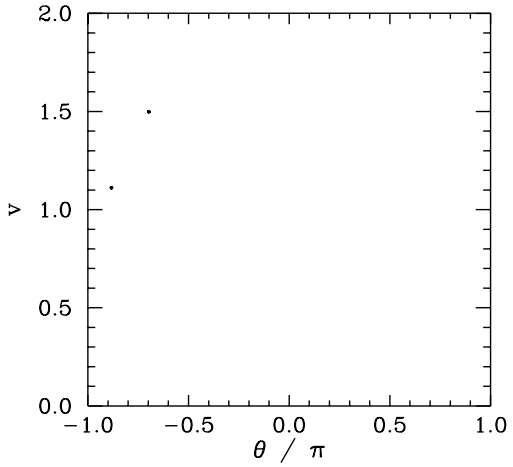


Figure 15.19: *The Poincaré section of a time-asymptotic orbit. Data calculated numerically for* $Q = 1.36$*,* $A = 1.5$*,* $\omega = 2/3$*,* $\theta(0) = 0$*,* $v(0) = 0$*, and* $\phi = 0$*.*

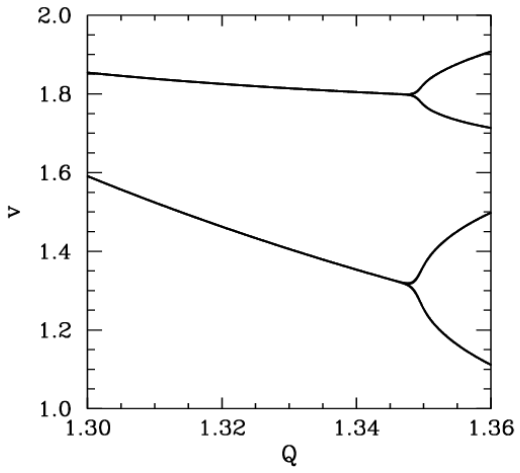


Figure 15.20: *The ν-coordinate of the Poincaré section of a time-asymptotic orbit plotted against the quality-factor* Q. *Data calculated numerically for* $A = 1.5$, $\omega = 2/3$, *and* $\phi = 0$. *Data is shown for two sets of initial conditions:* $\theta(0) = 0$ *and* $\nu(0) = 0$ *(lower branch); and* $\theta(0) = 0$ *and* $\nu(0) = -2$ *(upper branch).*

clear from Figure 15.20 that the right-favouring attractor undergoes a similar bifurcation at almost exactly the same Q-value. This is hardly surprising since, as has already been mentioned, for fixed physical parameters (*i.e.*, Q, A, ω), the left- and right-favouring attractors are mirror-images of one another.

15.9 Route to Chaos

Let us return to Figure 15.17, which tracks the evolution of a left-favouring periodic attractor as the quality-factor Q is gradually increased. Recall that when Q exceeds a critical value, which is about 1.348, then the attractor undergoes a period-doubling bifurcation which converts it from a period-1 to a period-2 attractor. This bifurcation is indicated by the forking of the curve in Figure 15.17. Let us now investigate what happens as we continue to increase Q. Figure 15.21 is basically a continuation of Figure 15.17. It can be seen that, as Q is gradually increased, the attractor undergoes a period-doubling bifurcation at $Q = 1.348$, as before, but then undergoes a *second* period-doubling bifurcation (indicated by the second forking of the curves) at $Q \simeq 1.370$, and a *third* bifurcation at $Q \simeq 1.375$. Obviously, the second bifurcation converts a period-2 attractor into a period-4 attractor (hence, two curves split apart to give four curves). Likewise, the third bifurcation converts a period-4 attractor into a period-8 attractor (hence, four curves split into eight curves). Shortly after the third bifurcation, the various curves in the figure seem to expand explosively and merge together to produce an area of almost solid black. As we shall see, this behaviour is indicative of the onset of *chaos*.

Figure 15.22 is a blow-up of Figure 15.21, showing more details of the onset of chaos.

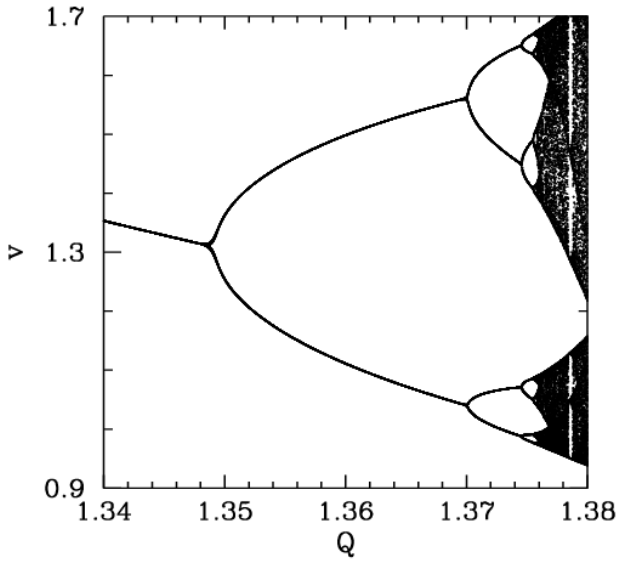


Figure 15.21: *The ν-coordinate of the Poincaré section of a time-asymptotic orbit plotted against the quality-factor* Q. *Data calculated numerically for* $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, $\nu(0) = 0$, *and* $\phi = 0$.

The period-4 to period-8 bifurcation can be seen quite clearly. However, we can also see a period-8 to period-16 bifurcation, at $Q \simeq 1.3755$. Finally, if we look carefully, we can see a hint of a period-16 to period-32 bifurcation, just before the start of the solid black region. Figures 15.21 and 15.22 seem to suggest that the onset of chaos is triggered by an *infinite series* of period-doubling bifurcations.

Table 15.1 gives some details of the sequence of period-doubling bifurcations shown in Figures 15.21 and 15.22. Let us introduce a bifurcation index n: the period-1 to period-2 bifurcation corresponds to $n = 1$; the period-2 to period-4 bifurcation corresponds to $n = 2$; and so on. Let Q_n be the critical value of the quality-factor Q at which the nth bifurcation is triggered. Table 15.1 shows the Q_n, determined from Figures 15.21 and 15.22, for $n = 1$ to 5. Also shown is the ratio

$$F_n = \frac{Q_{n-1} - Q_{n-2}}{Q_n - Q_{n-1}} \tag{15.24}$$

for $n = 3$ to 5. It can be seen that Table 15.1 offers reasonably convincing evidence that this ratio takes the *constant* value $F = 4.69$. It follows that we can estimate the critical Q-value required to trigger the nth bifurcation via the following formula:

$$Q_n = Q_1 + (Q_2 - Q_1) \sum_{j=0}^{n-2} \frac{1}{F^j}, \tag{15.25}$$

for $n > 1$. Note that the distance (in Q) between bifurcations decreases rapidly as n

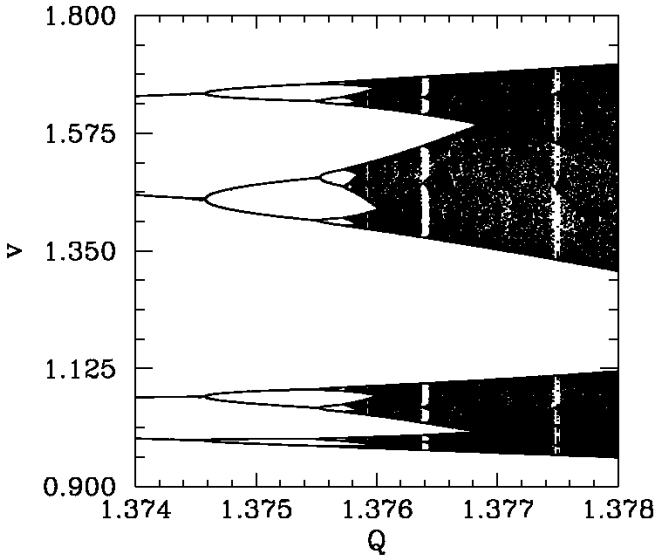


Figure 15.22: *The ν-coordinate of the Poincaré section of a time-asymptotic orbit plotted against the quality-factor* Q. *Data calculated numerically for* $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, $\nu(0) = 0$, *and* $\phi = 0$.

increases. In fact, the above formula predicts an *accumulation* of period-doubling bifurcations at $Q = Q_\infty$, where

$$Q_\infty = Q_1 + (Q_2 - Q_1) \sum_{j=0}^{\infty} \frac{1}{F^j} \equiv Q_1 + (Q_2 - Q_1) \frac{F}{F-1} = 1.3758. \tag{15.26}$$

Note that our calculated accumulation point corresponds almost exactly to the onset of the solid black region in Figure 15.22. By the time that Q exceeds Q_∞, we expect the attractor to have been converted into a *period-infinity* attractor via an infinite series of period-doubling bifurcations. A period-infinity attractor is one whose corresponding motion *never* repeats itself, no matter how long we wait. In dynamics, such bounded aperiodic motion is generally referred to as *chaos*. Hence, a period-infinity attractor is sometimes called a *chaotic attractor*. Now, period-n motion is represented by n separate curves in Figure 15.22. It is, therefore, not surprising that chaos (*i.e.*, period-infinity motion) is represented by an infinite number of curves which merge together to form a region of solid black.

Let us examine the onset of chaos in a little more detail. Figures 15.23–15.26 show details of the pendulum's time-asymptotic motion at various stages on the period-doubling cascade discussed above. Figure 15.23 shows period-4 motion: note that the Poincaré section consists of four points, and the associated sequence of net rotations per period of the pendulum repeats itself every four periods. Figure 15.24 shows period-8 motion: now the Poincaré section consists of eight points, and the rotation sequence repeats itself every eight periods. Figure 15.25 shows period-16 motion: as expected, the Poincaré

Bifurcation	n	Q_n	$Q_n - Q_{n-1}$	F_n
period-1→period-2	1	1.34870	-	-
period-2→period-4	2	1.37003	0.02133	-
period-4→period-8	3	1.37458	0.00455	4.69 ± 0.01
period-8→period-16	4	1.37555	0.00097	4.69 ± 0.04
period-16→period-32	5	1.37575	0.00020	4.9 ± 0.20

Table 15.1: *The period-doubling cascade.*

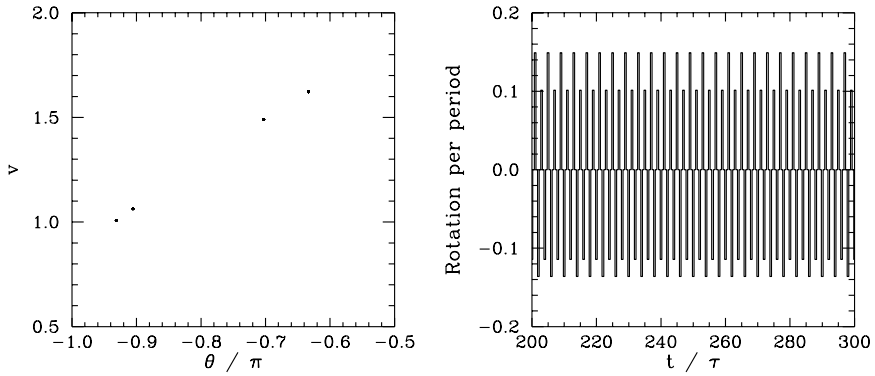


Figure 15.23: *The Poincaré section of a time-asymptotic orbit. Data calculated numerically for* $Q = 1.372$*,* $A = 1.5$*,* $\omega = 2/3$*,* $\theta(0) = 0$*,* $\nu(0) = 0$*, and* $\phi = 0$*. Also, shown is the net rotation per period,* $\Delta\theta/2\pi$*, calculated at the Poincaré phase* $\phi = 0$*.*

section consists of sixteen points, and the rotation sequence repeats itself every sixteen periods. Finally, Figure 15.26 shows chaotic motion. Note that the Poincaré section now consists of a set of four *continuous* line segments, which are, presumably, made up of an infinite number of points (corresponding to the infinite period of chaotic motion). Note, also, that the associated sequence of net rotations per period shows no obvious sign of ever repeating itself. In fact, this sequence looks rather like one of the previously shown periodic sequences with the addition of a small random component. The generation of *apparently random* motion from equations of motion, such as Equations (15.10) and (15.11), which contain no overtly random elements is one of the most surprising features of non-linear dynamics.

Many non-linear dynamical systems found in nature exhibit a transition from periodic to chaotic motion as some control parameter is varied. Moreover, there are various known mechanisms by which chaotic motion can arise from periodic motion. A transition to chaos via an infinite series of period-doubling bifurcations, as illustrated above, is one of the most commonly occurring mechanisms. Around 1975, the physicist Mitchell Feigenbaum was investigating a simple mathematical model, known as the *logistic map*, which exhibits a transition to chaos, via a sequence of period-doubling bifurcations, as a control parameter r is increased. Let r_n be the value of r at which the first 2^n-period cycle appears. Feigenbaum

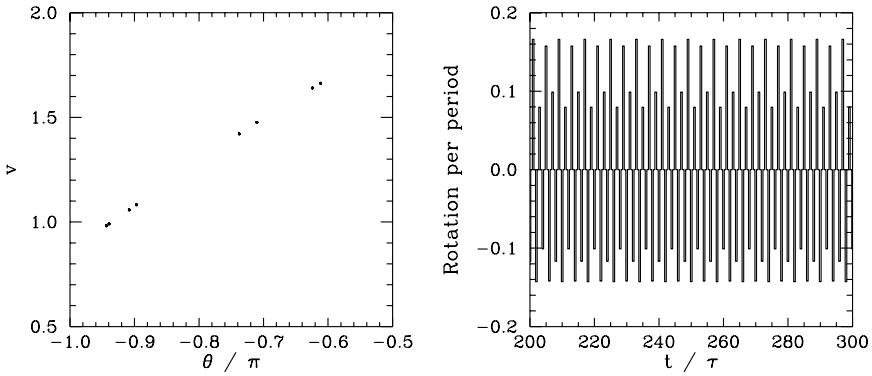


Figure 15.24: *The Poincaré section of a time-asymptotic orbit. Data calculated numerically for* $Q = 1.375$, $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, $\nu(0) = 0$, *and* $\phi = 0$. *Also, shown is the net rotation per period,* $\Delta\theta/2\pi$, *calculated at the Poincaré phase* $\phi = 0$.

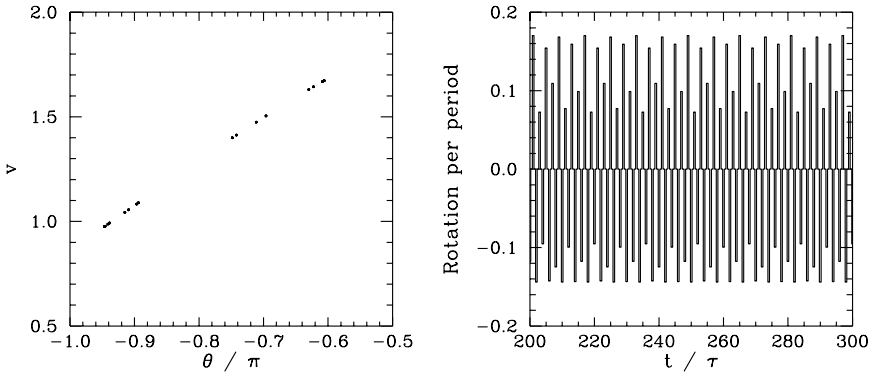


Figure 15.25: *The Poincaré section of a time-asymptotic orbit. Data calculated numerically for* $Q = 1.3757$, $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, $\nu(0) = 0$, *and* $\phi = 0$. *Also, shown is the net rotation per period,* $\Delta\theta/2\pi$, *calculated at the Poincaré phase* $\phi = 0$.

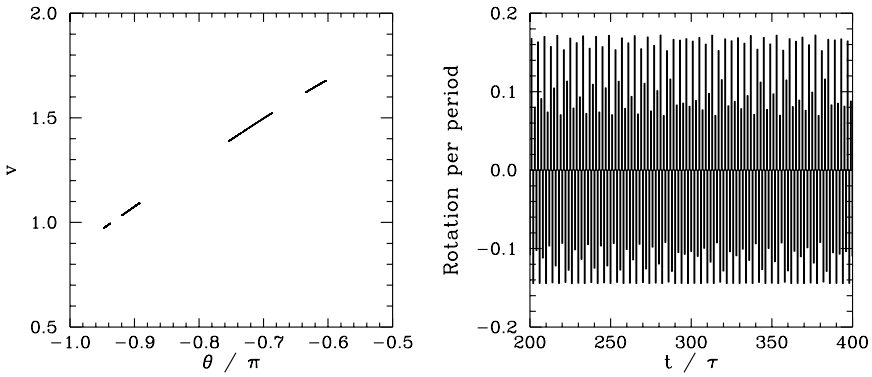


Figure 15.26: *The Poincaré section of a time-asymptotic orbit. Data calculated numerically for* $Q = 1.376$, $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, $\nu(0) = 0$, *and* $\phi = 0$. *Also, shown is the net rotation per period,* $\Delta\theta/2\pi$, *calculated at the Poincaré phase* $\phi = 0$.

noticed that the ratio

$$F_n = \frac{r_{n-1} - r_{n-2}}{r_n - r_{n-1}} \tag{15.27}$$

converges rapidly to a constant value, $F = 4.669$, as n increases. Feigenbaum was able to demonstrate that this value of F is common to a wide range of different mathematic models which exhibit transitions to chaos via period-doubling bifurcations.[6] Feigenbaum went on to argue that the *Feigenbaum ratio*, F_n, should converge to the value 4.669 in *any* dynamical system exhibiting a transition to chaos via period-doubling bifurcations.[7] This amazing prediction has been verified experimentally in a number of quite different physical systems.[8] Note that our best estimate of the Feigenbaum ratio (see Table 15.1) is 4.69 ± 0.01, in good agreement with Feigenbaum's prediction.

The existence of a universal ratio characterizing the transition to chaos via period-doubling bifurcations is one of many pieces of evidence indicating that chaos is a *universal* phenomenon (*i.e.*, the onset and nature of chaotic motion in different dynamical systems has many common features). This observation encourages us to believe that in studying the chaotic motion of a damped periodically driven pendulum we are learning lessons which can be applied to a wide range of non-linear dynamical systems.

15.10 Sensitivity to Initial Conditions

Suppose that we launch our pendulum and then wait until its motion has converged onto a particular attractor. The subsequent motion can be visualized as a trajectory $\theta_0(t)$, $\nu_0(t)$

[6]M.J. Feigenbaum, *Quantitative universality for a class of nonlinear transformations*, J. Stat. Phys. **19**, 25 (1978).

[7]M.J. Feigenbaum, *The universal metric properties of nonlinear transformations*, J. Stat. Phys. **21**, 69 (1979).

[8]P. Citanovic, *Universality in chaos*, (Adam Hilger, Bristol UK, 1989).

through phase-space. Suppose that we then somehow perturb the pendulum, at time $t = t_0$, such that its position in phase-space is instantaneously changed from $\theta_0(t_0)$, $\nu_0(t_0)$ to $\theta_0(t_0)+\delta\theta_0$, $\nu_0(t_0)+\delta\nu_0$. The subsequent motion can be visualized as a second trajectory $\theta_1(t)$, $\nu_1(t)$ through phase-space. What is the relationship between the original trajectory $\theta_0(t)$, $\nu_0(t)$ and the perturbed trajectory $\theta_1(t)$, $\nu_1(t)$? In particular, does the phase-space separation between the two trajectories, whose components are

$$\delta\theta(\Delta t) = \theta_1(t_0+\Delta t) - \theta_0(t_0+\Delta t), \tag{15.28}$$

$$\delta\nu(\Delta t) = \nu_1(t_0+\Delta t) - \nu_0(t_0+\Delta t), \tag{15.29}$$

grow in time, decay in time, or stay more or less the same? What we are really investigating is how sensitive the time-asymptotic motion of the pendulum is to initial conditions.

According to the linear analysis of Section 15.3,

$$\delta\theta(\Delta t) = \delta\theta_0\,\cos(\omega_*\,\Delta t)\,\mathrm{e}^{-\Delta t/2Q} + \frac{1}{\omega_*}\left\{\delta\nu_0 + \frac{\delta\theta_0}{2Q}\right\}\sin(\omega_*\,\Delta t)\,\mathrm{e}^{-\Delta t/2Q}, \tag{15.30}$$

$$\delta\nu(\Delta t) = \delta\nu_0\,\cos(\omega_*\,\Delta t)\,\mathrm{e}^{-\Delta t/2Q} - \frac{1}{\omega_*}\left\{\delta\theta_0 + \frac{\delta\nu_0}{2Q}\right\}\sin(\omega_*\,\Delta t)\,\mathrm{e}^{-\Delta t/2Q}, \tag{15.31}$$

assuming $\sin(\omega_*\,t_0) = 0$. It is clear that, in the linear regime, at least, the pendulum's time-asymptotic motion is not particularly sensitive to initial conditions. In fact, if we move the pendulum's phase-space trajectory slightly off the linear attractor, as described above, then the perturbed trajectory decays back to the attractor *exponentially* in time. In other words, if we wait long enough then the perturbed and unperturbed motions of the pendulum become effectively indistinguishable. Let us now investigate whether this insensitivity to initial conditions carries over into the non-linear regime.

Figures 15.27–15.30 show the results of the experiment described above, in which the pendulum's phase-space trajectory is moved slightly off an attractor and the phase-space separation between the perturbed and unperturbed trajectories is then monitored as a function of time, at various stages on the period-doubling cascade discussed in the previous section. To be more exact, the figures show the logarithm of the absolute magnitude of the ν-component of the phase-space separation between the perturbed and unperturbed trajectories as a function of normalized time.

Figure 15.27 shows the time evolution of the ν-component of the phase-space separation, $\delta\nu$, between two neighbouring trajectories, one of which is the period-4 attractor illustrated in Figure 15.23. It can be seen that $\delta\nu$ decays rapidly in time. In fact, the graph of $\log(|\delta\nu|)$ versus Δt can be plausibly represented as a *straight-line* of gradient λ. In other words,

$$|\delta\nu(\Delta t)| \simeq \delta\nu_0\,\mathrm{e}^{\,\lambda\,\Delta t}, \tag{15.32}$$

where the quantity λ is known as the *Liapunov exponent*, and is obviously negative. Clearly, in this case, λ measures the strength of the exponential convergence of the two trajectories

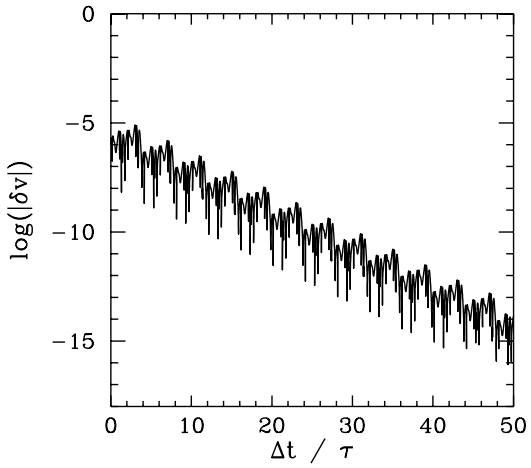


Figure 15.27: *The v-component of the separation between two neighbouring phase-space trajectories (one of which lies on an attractor) plotted against normalized time. Data calculated numerically for $Q = 1.372$, $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, and $v(0) = 0$. The separation between the two trajectories is initialized to $\delta\theta_0 = \delta v_0 = 10^{-6}$ at $\Delta t = 0$.*

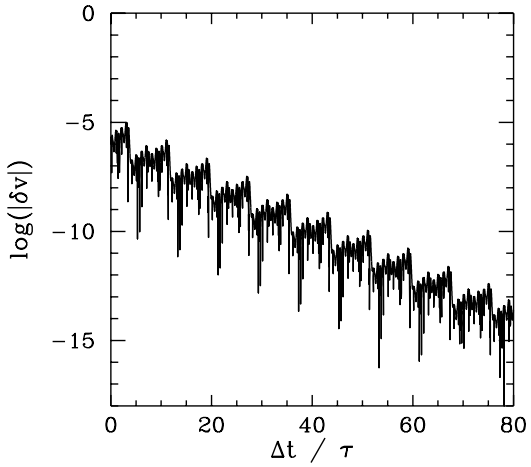


Figure 15.28: *The v-component of the separation between two neighbouring phase-space trajectories (one of which lies on an attractor) plotted against normalized time. Data calculated numerically for $Q = 1.375$, $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, and $v(0) = 0$. The separation between the two trajectories is initialized to $\delta\theta_0 = \delta v_0 = 10^{-6}$ at $\Delta t = 0$.*

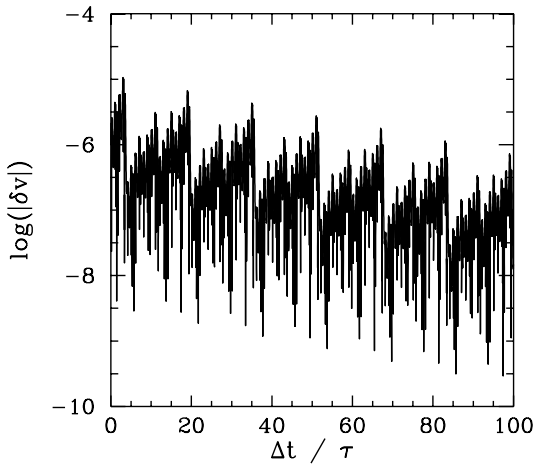


Figure 15.29: *The v-component of the separation between two neighbouring phase-space trajectories (one of which lies on an attractor) plotted against normalized time. Data calculated numerically for $Q = 1.3757$, $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, and $v(0) = 0$. The separation between the two trajectories is initialized to $\delta\theta_0 = \delta v_0 = 10^{-6}$ at $\Delta t = 0$.*

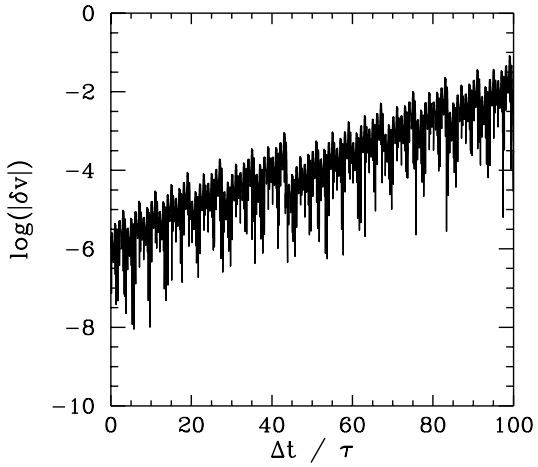


Figure 15.30: *The v-component of the separation between two neighbouring phase-space trajectories (one of which lies on an attractor) plotted against normalized time. Data calculated numerically for $Q = 1.376$, $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, and $v(0) = 0$. The separation between the two trajectories is initialized to $\delta\theta_0 = \delta v_0 = 10^{-6}$ at $\Delta t = 0$.*

in phase-space. Of course, the graph of $\log(|\delta\nu|)$ versus Δt is not exactly a straight-line. There are deviations due to the fact that $\delta\nu$ oscillates, as well as decays, in time. There are also deviations because the strength of the exponential convergence between the two trajectories varies along the attractor.

The above definition of the Liapunov exponent is rather inexact, for two main reasons. In the first place, the strength of the exponential convergence/divergence between two neighbouring trajectories in phase-space, one of which is an attractor, generally *varies* along the attractor. Hence, we should really take formula (15.32) and somehow *average* it over the attractor, in order to obtain a more unambiguous definition of λ. In the second place, since the dynamical system under investigation is a second-order system, it actually possesses *two* different Liapunov exponents. Consider the evolution of an infinitesimal circle of perturbed initial conditions, centred on a point in phase-space lying on an attractor. During its evolution, the circle will become distorted into an infinitesimal ellipse. Let δ_k, where $k = 1, 2$, denote the phase-space length of the kth principal axis of the ellipse. The two Liapunov exponents, λ_1 and λ_2, are defined via $\delta_k(\Delta t) \simeq \delta_k(0)\exp(\lambda_k \Delta t)$. However, for large Δt, the diameter of the ellipse is effectively controlled by the Liapunov exponent with the most positive real part. Hence, when we refer to *the* Liapunov exponent, λ, what we generally mean is the Liapunov exponent with the most positive real part.

Figure 15.28 shows the time evolution of the ν-component of the phase-space separation, $\delta\nu$, between two neighbouring trajectories, one of which is the period-8 attractor illustrated in Figure 15.24. It can be seen that $\delta\nu$ decays in time, though not as rapidly as in Figure 15.27. Another way of saying this is that the Liapunov exponent of the periodic attractor shown in Figure 15.24 is negative (*i.e.*, it has a negative real part), but not as negative as that of the periodic attractor shown in Figure 15.23.

Figure 15.29 shows the time evolution of the ν-component of the phase-space separation, $\delta\nu$, between two neighbouring trajectories, one of which is the period-16 attractor illustrated in Figure 15.25. It can be seen that $\delta\nu$ decays weakly in time. In other words, the Liapunov exponent of the periodic attractor shown in Figure 15.25 is small and negative.

Finally, Figure 15.30 shows the time evolution of the ν-component of the phase-space separation, $\delta\nu$, between two neighbouring trajectories, one of which is the chaotic attractor illustrated in Figure 15.26. It can be seen that $\delta\nu$ *increases* in time. In other words, the Liapunov exponent of the chaotic attractor shown in Figure 15.26 is *positive*. Further investigation reveals that, as the control parameter Q is gradually increased, the Liapunov exponent changes sign and becomes positive at exactly the same point that chaos ensues in Figure 15.22.

The above discussion strongly suggests that periodic attractors are characterized by negative Liapunov exponents, whereas chaotic attractors are characterized by positive exponents. But, how can an attractor have a positive Liapunov exponent? Surely, a positive exponent necessarily implies that neighbouring phase-space trajectories *diverge* from the attractor (and, hence, that the attractor is not a true attractor)? It turns out that this is not the case. The chaotic attractor shown in Figure 15.26 is a true attractor, in the sense that

neighbouring trajectories rapidly converge onto it—*i.e.*, after a few periods of the external drive their Poincaré sections plot out the same four-line segment shown in Figure 15.26. Thus, the exponential divergence of neighbouring trajectories, characteristic of chaotic attractors, takes place *within the attractor* itself. Obviously, this exponential divergence must come to an end when the phase-space separation of the trajectories becomes comparable to the extent of the attractor.

A dynamical system characterized by a positive Liapunov exponent, λ, has a *time horizon* beyond which regular deterministic prediction breaks down. Suppose that we measure the initial conditions of an experimental system very accurately. Obviously, no measurement is perfect: there is always some error δ_0 between our estimate and the true initial state. After a time t, the discrepancy grows to $\delta(t) \sim \delta_0 \exp(\lambda\, t)$. Let a be a measure of our tolerance: *i.e.*, a prediction within a of the true state is considered acceptable. It follows that our prediction becomes unacceptable when $\delta \gg a$, which occurs when

$$t > t_h \sim \frac{1}{\lambda} \ln\left(\frac{a}{\delta_0}\right). \tag{15.33}$$

Note the *logarithmic* dependence on δ_0. This ensures that, in practice, no matter how hard we work to reduce our initial measurement error, we cannot predict the behaviour of the system for longer than a few multiples of $1/\lambda$.

It follows, from the above discussion, that chaotic attractors are associated with motion which is essentially *unpredictable*. In other words, if we attempt to integrate the equations of motion of a chaotic system then even the slightest error made in the initial conditions will be amplified exponentially over time, and will rapidly destroy the accuracy of our prediction. Eventually, all that we will be able to say is that the motion lies somewhere on the chaotic attractor in phase-space, but exactly where it lies on the attractor at any given time will be unknown to us.

The hyper-sensitivity of chaotic systems to initial conditions is sometimes called the *butterfly effect*. The idea is that a butterfly flapping its wings in a South American rain-forest could, in principle, affect the weather in Texas (since the atmosphere exhibits chaotic dynamics). This idea was first publicized by the meteorologist Edward Lorenz, who constructed a very crude model of the convection of the atmosphere when it is heated from below by the ground.[9] Lorenz discovered, much to his surprise, that his model atmosphere exhibited chaotic motion—which, at that time, was virtually unknown. In fact, Lorenz was essentially the first scientist to fully understand the nature and ramifications of chaotic motion in physical systems. In particular, Lorenz realized that the chaotic dynamics of the atmosphere spells the doom of long-term weather forecasting: the best one can hope to achieve is to predict the weather a few days in advance ($1/\lambda$ for the atmosphere is of order a few days).

[9]E. Lorenz, *Deterministic nonperiodic flow*, J. Atmospheric Science **20**, 130 (1963).

15.11 Definition of Chaos

There is no universally agreed definition of chaos. However, most people would accept the following working definition:

> Chaos is aperiodic time-asymptotic behaviour in a deterministic system which exhibits sensitive dependence on initial conditions.

This definition contains three main elements:

1. Aperiodic time-asymptotic behaviour—this implies the existence of phase-space trajectories which do not settle down to fixed points or periodic orbits. For practical reasons, we insist that these trajectories are not too rare. We also require the trajectories to be *bounded*: *i.e.*, they should not go off to infinity.

2. Deterministic—this implies that the equations of motion of the system possess no random inputs. In other words, the irregular behaviour of the system arises from non-linear dynamics, and not from noisy driving forces.

3. Sensitive dependence on initial conditions—this implies that nearby trajectories in phase-space separate exponentially fast in time: *i.e.*, that the system has a positive Liapunov exponent.

15.12 Periodic Windows

Let us return to Figure 15.22. Recall, that this figure shows the onset of chaos, via a cascade of period-doubling bifurcations, as the quality-factor Q is gradually increased. Figure 15.31 is essentially a continuation of Figure 15.22 which shows the full extent of the chaotic region (in Q-ν space). It can be seen that the chaotic region ends abruptly when Q exceeds a critical value, which is about 1.4215. Beyond this critical value, the time-asymptotic motion appears to revert to period-1 motion (*i.e.*, the solid black region collapses to a single curve). It can also be seen that the chaotic region contains many narrow windows in which chaos reverts to periodic motion (*i.e.*, the solid black region collapses to n curves, where n is the period of the motion) for a short interval in Q. The four widest windows are indicated in the figure.

Figure 15.32 is a blow-up of the period-3 window shown in Figure 15.31. It can be seen that the window appears "out of the blue" as Q is gradually increased. However, it can also be seen that, as Q is further increased, the window breaks down, and eventually disappears, due to the action of a cascade of period-doubling bifurcations. The same basic mechanism operates here as in the original period-doubling cascade, discussed in Section 15.9, except that now the orbits are of period $3 \cdot 2^n$, instead of $2 \cdot 2^n$. Note that all of the other periodic windows seen in Figure 15.31 break down in an analogous manner, as Q is increased.

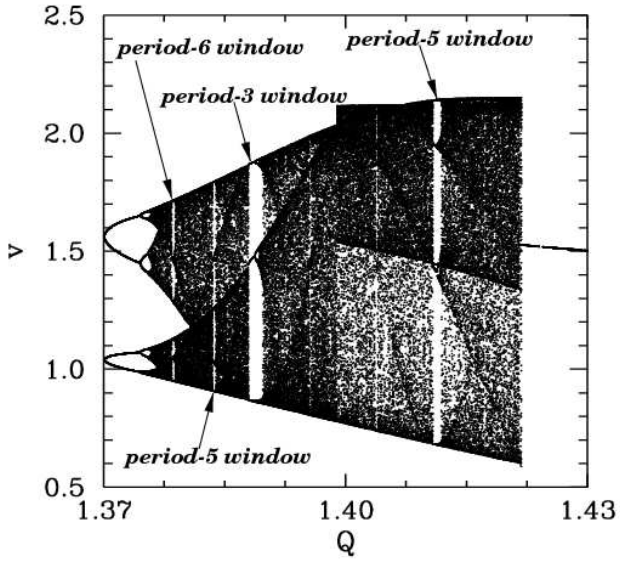


Figure 15.31: *The v-coordinate of the Poincaré section of a time-asymptotic orbit plotted against the quality-factor* Q. *Data calculated numerically for* $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, $v(0) = -0.75$, *and* $\phi = 0$.

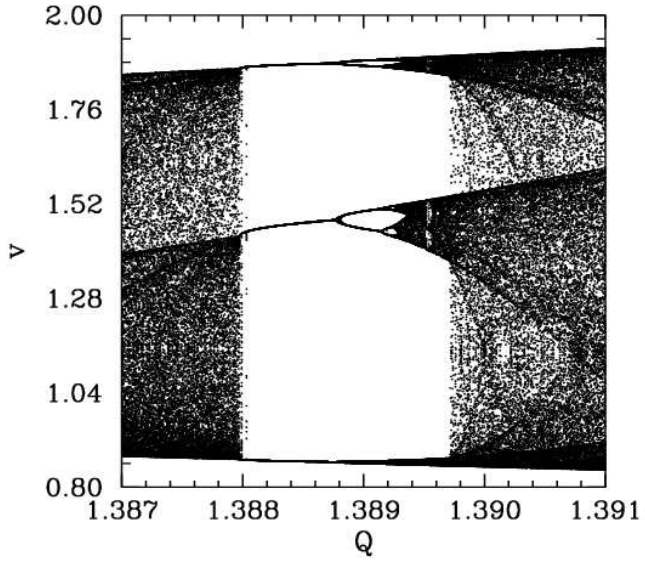


Figure 15.32: *The v-coordinate of the Poincaré section of a time-asymptotic orbit plotted against the quality-factor* Q. *Data calculated numerically for* $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, $v(0) = -0.75$, *and* $\phi = 0$.

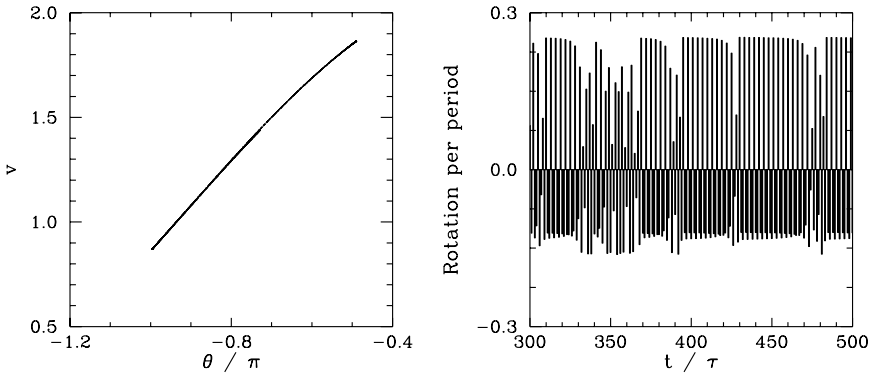


Figure 15.33: *The Poincaré section of a time-asymptotic orbit. Data calculated numerically for* $Q = 1.387976$*,* $A = 1.5$*,* $\omega = 2/3$*,* $\theta(0) = 0$*,* $v(0) = -0.75$*, and* $\phi = 0$*. Also, shown is the net rotation per period,* $\Delta\theta/2\pi$*, calculated at the Poincaré phase* $\phi = 0$*.*

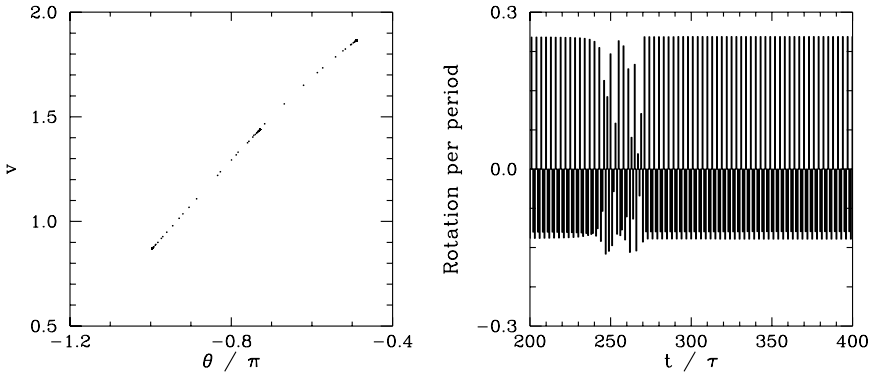


Figure 15.34: *The Poincaré section of a time-asymptotic orbit. Data calculated numerically for* $Q = 1.387977$*,* $A = 1.5$*,* $\omega = 2/3$*,* $\theta(0) = 0$*,* $v(0) = -0.75$*, and* $\phi = 0$*. Also, shown is the net rotation per period,* $\Delta\theta/2\pi$*, calculated at the Poincaré phase* $\phi = 0$*.*

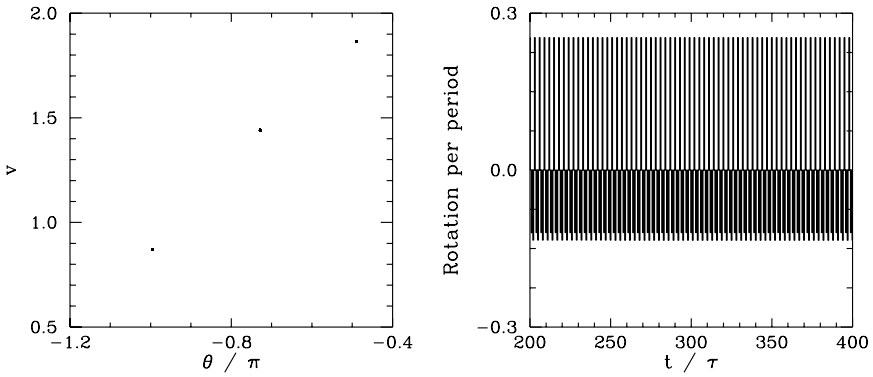


Figure 15.35: *The Poincaré section of a time-asymptotic orbit. Data calculated numerically for* $Q = 1.387978$*,* $A = 1.5$*,* $\omega = 2/3$*,* $\theta(0) = 0$*,* $\nu(0) = -0.75$*, and* $\phi = 0$*. Also, shown is the net rotation per period,* $\Delta\theta/2\pi$*, calculated at the Poincaré phase* $\phi = 0$*.*

We now understand how periodic windows break down. But, how do they appear in the first place? Figures 15.33–15.35 show details of the pendulum's time-asymptotic motion calculated *just before* the appearance of the period-3 window (shown in Figure 15.32), *just at* the appearance of the window, and *just after* the appearance of the window, respectively. It can be seen, from Figure 15.33, that just before the appearance of the window the attractor is chaotic (*i.e.*, its Poincaré section consists of a line, rather than a discrete set of points), and the time-asymptotic motion of the pendulum consists of intervals of period-3 motion interspersed with intervals of chaotic motion. Figure 15.34 shows that just at the appearance of the window the attractor loses much of its chaotic nature (*i.e.*, its Poincaré section breaks up into a series of points), and the chaotic intervals become shorter and much less frequent. Finally, Figure 15.35 shows that just after the appearance of the window the attractor collapses to a period-3 attractor, and the chaotic intervals cease altogether. All of the other periodic windows seen in Figure 15.31 appear in an analogous manner to that just described.

According to the above discussion, the typical time-asymptotic motion seen just prior to the appearance of a period-n window consists of intervals of period-n motion interspersed with intervals of chaotic motion. This type of behaviour is called *intermittency*, and is observed in a wide variety of non-linear systems. As we move away from the window, in parameter space, the intervals of periodic motion become gradually shorter and more infrequent. Eventually, they cease altogether. Likewise, as we move towards the window, the intervals of periodic motion become gradually longer and more frequent. Eventually, the whole motion becomes periodic.

In 1973, Metropolis and co-workers investigated a class of simple mathematical models which all exhibit a transition to chaos, via a cascade of period-doubling bifurcations, as some control parameter r is increased.[10] They were able to demonstrate that, for these

[10]N. Metropolis, M.L. Stein, and P.R. Stein, *On finite limit sets for transformations on the unit interval*, J. Combin. Theor. **15**, 25 (1973).

models, the order in which stable periodic orbits occur as r is increased is fixed. That is, *stable periodic attractors always occur in the same sequence* as r is varied. This sequence is called the universal or *U-sequence*. It is possible to make a fairly convincing argument that any physical system which exhibits a transition to chaos via a sequence of period-doubling bifurcations should also exhibit the U-sequence of stable periodic attractors. Up to period-6, the U-sequence is

$$1,\ 2,\ 2\times 2,\ 6,\ 5,\ 3,\ 2\times 3,\ 5,\ 6,\ 4,\ 6,\ 5,\ 6.$$

The beginning of this sequence is familiar: periods 1, 2, 2×2 are the first stages of the period-doubling cascade. (The later period-doublings give rise to periods greater than 6, and so are omitted here). The next periods, $6, 5, 3$ correspond to the first three periodic windows shown in Figure 15.31. Period 2×3 is the first component of the period-doubling cascade which breaks up the period-3 window. The next period, 5, corresponds to the last periodic window shown in Figure 15.31. The remaining periods, 6, 4, 6, 5, 6, correspond to tiny periodic windows, which, in practice, are virtually impossible to observe. It follows that our driven pendulum system exhibits the U-sequence of stable periodic orbits fairly convincingly. This sequence has also been observed experimentally in other, quite different, dynamical systems.[11] The existence of a universal sequence of stable periodic orbits in dynamical systems which exhibit a transition to chaos via a cascade of period-doubling bifurcations is another indication that chaos is a universal phenomenon.

15.13 Further Investigation

Figure 15.36 shows the complete *bifurcation diagram* for the damped, periodically driven, pendulum (with $A = 1.5$ and $\omega = 2/3$). It can be seen that the chaotic region investigated in the previous section is, in fact, the first, and least extensive, of *three* different chaotic regions.

The interval between the first and second chaotic regions is occupied by the period-1 orbits shown in Figure 15.37. Note that these orbits differ somewhat from previously encountered period-1 orbits in that the pendulum executes a complete rotation (either to the left or to the right) every period of the external drive. Now, an n, l periodic orbit is defined such that

$$\theta(t + n\,\tau) = \theta(t) + 2\pi\, l$$

for all t (after the transients have died away). It follows that all of the periodic orbits which we encountered in previous sections were $l = 0$ orbits: *i.e.*, their associated motions did not involve a net rotation of the pendulum. The orbits show in Figure 15.37 are $n = 1, l = -1$ and $n = 1, l = +1$ orbits, respectively. The existence of periodic orbits in which the pendulum undergoes a net rotation, either to the left or to the right, is another example of spatial symmetry breaking—there is nothing in the pendulum's equations of motion which distinguishes between the two possible directions of rotation.

[11] R.H. Simoyi, A. Wolf, and H.L. Swinney, *One-dimensional dynamics in a multi-component chemical reaction*, Phys. Rev. Lett. **49**, 245 (1982).

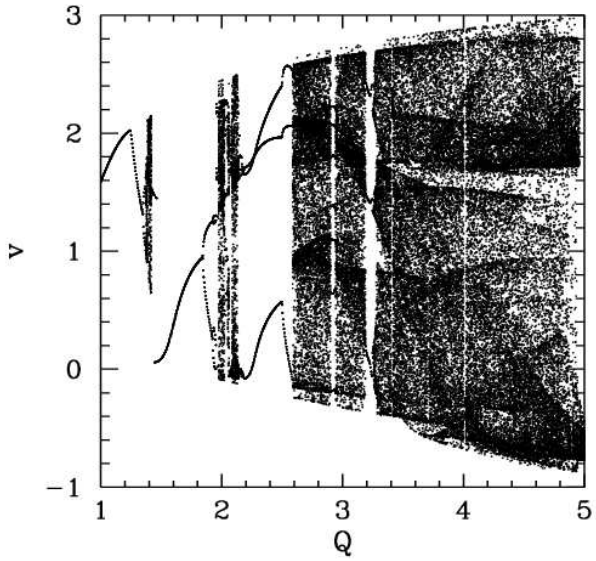


Figure 15.36: *The v-coordinate of the Poincaré section of a time-asymptotic orbit plotted against the quality-factor* Q. *Data calculated numerically for* $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, $v(0) = 0$, *and* $\phi = 0$.

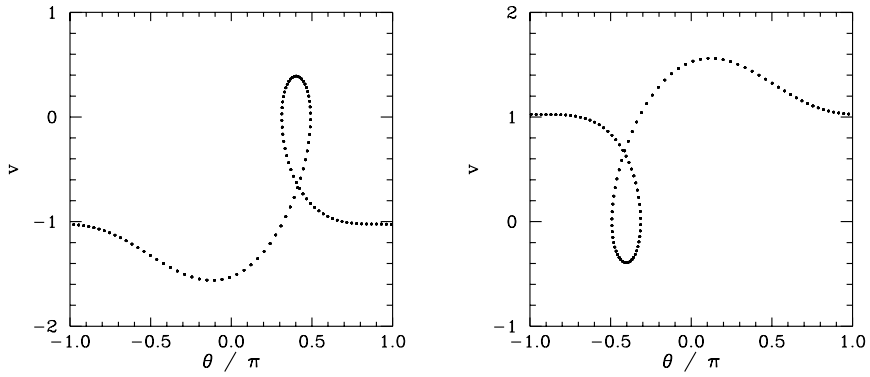


Figure 15.37: *Equally spaced (in time) points on a time-asymptotic orbit in phase-space. Data calculated numerically for* $Q = 1.5$, $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, *and* $v(0) = 0$. *Also shown is the time-asymptotic orbit calculated for the modified initial conditions* $\theta(0) = 0$, *and* $v(0) = -1$.

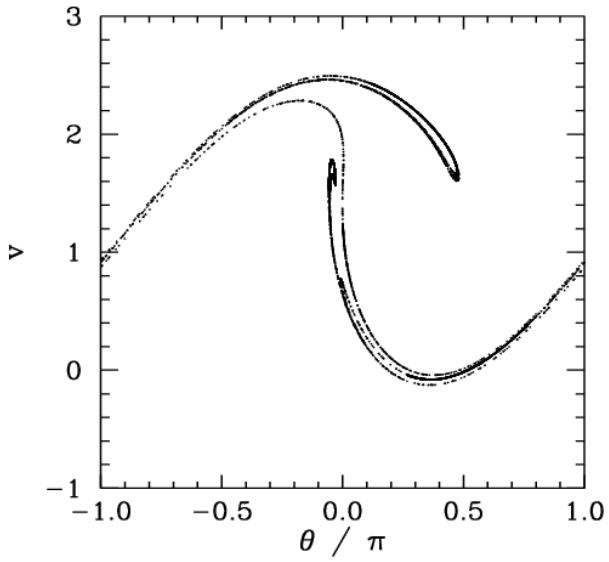


Figure 15.38: *The Poincaré section of a time-asymptotic orbit. Data calculated numerically for* $Q = 2.13$*,* $A = 1.5$*,* $\omega = 2/3$*,* $\theta(0) = 0$*,* $v(0) = 0$*, and* $\phi = 0$*.*

Figure 15.38 shows the Poincaré section of a typical attractor in the second chaotic region shown in Figure 15.36. It can be seen that this attractor is far more convoluted and extensive than the simple 4-line chaotic attractor pictured in Figure 15.26. In fact, the attractor shown in Figure 15.38 is clearly a *fractal* curve. It turns out that virtually all chaotic attractors exhibit fractal structure.

The interval between the second and third chaotic regions shown in Figure 15.36 is occupied by $n = 3$, $l = 0$ periodic orbits. Figure 15.39 shows the Poincaré section of a typical attractor in the third chaotic region. It can be seen that this attractor is even more overtly fractal in nature than that pictured in the previous figure. Note that the fractal nature of chaotic attractors is closely associated with some of their unusual properties. Trajectories on a chaotic attractor remain confined to a bounded region of phase-space, and yet they separate from their neighbours exponentially fast (at least, initially). How can trajectories diverge endlessly and still stay bounded? The basic mechanism is described below. If we imagine a blob of initial conditions in phase-space then these undergo a series of repeated *stretching and folding* episodes, as the chaotic motion unfolds. The stretching is what gives rise to the divergence of neighbouring trajectories. The folding is what ensures that the trajectories remain bounded. The net result is a phase-space structure which looks a bit like filo pastry—in other words, a fractal structure.

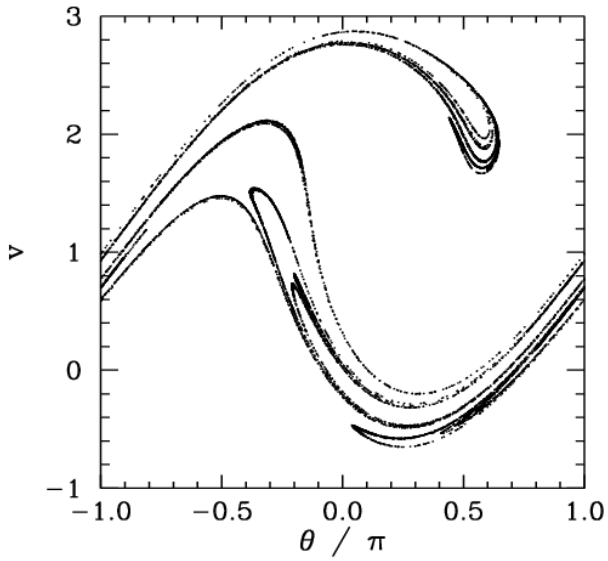


Figure 15.39: *The Poincaré section of a time-asymptotic orbit. Data calculated numerically for* $Q = 3.9$, $A = 1.5$, $\omega = 2/3$, $\theta(0) = 0$, $v(0) = 0$, *and* $\phi = 0$.

A Vector Algebra and Vector Calculus

A.1 Introduction

This appendix contains a brief outline of vector algebra and vector calculus. The essential purpose of *vector algebra* is to convert the propositions of Euclidean geometry in three-dimensional space into a convenient algebraic form. *Vector calculus* allows us to define the instantaneous velocity and acceleration of a moving object in three-dimensional space, as well as the work done when such an object travels along a general curved trajectory in a force field. Vector calculus also introduces the concept of a *scalar field*: *e.g.*, the potential energy associated with a conservative force field.

A.2 Scalars and Vectors

Many physical quantities (*e.g.*, mass, energy) are entirely defined by a numerical magnitude (in appropriate units). Such quantities, which have no directional element, are known as *scalars*. Moreover, since scalars can be represented by real numbers, it follows that they obey the familiar laws of ordinary algebra. However, there exits a second class of physical quantities (*e.g.*, velocity, acceleration, force) which are only completely defined when both a numerical magnitude *and* a direction in space is specified. Such quantities are known as *vectors*. By definition, a vector obeys the same algebra as a displacement in space, and may thus be represented geometrically by a straight-line, $\overrightarrow{PQ}$ (say), where the arrow indicates the direction of the displacement (*i.e.*, from point P to point Q)—see Figure A.1. The magnitude of the vector is represented by the length of the straight-line.

It is conventional to denote vectors by bold-faced symbols (*e.g.*, **a**, **F**) and scalars by non-bold-faced symbols (*e.g.*, r, S). The magnitude of a general vector, **a**, is denoted $|\mathbf{a}|$, or just a, and is, by definition, always greater than or equal to zero. It is convenient to define a vector with zero magnitude—this is denoted **0**, and has no direction. Finally, two vectors, **a** and **b**, are said to be equal when their magnitudes and directions are identical.

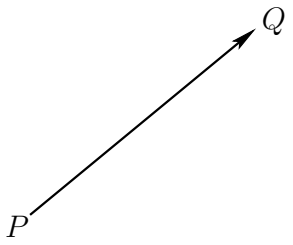


Figure A.1: *A vector.*

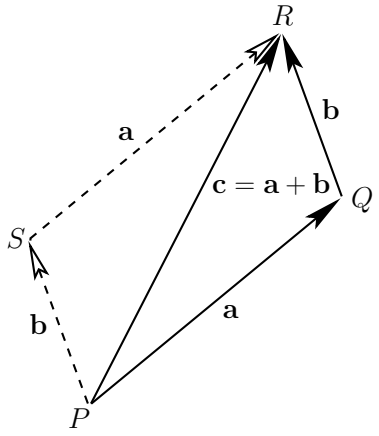


Figure A.2: *Vector addition.*

A.3 Vector Algebra

Suppose that the displacements $\overrightarrow{\mathrm{PQ}}$ and $\overrightarrow{\mathrm{QR}}$ represent the vectors **a** and **b**, respectively—see Figure A.2. It can be seen that the result of combining these two displacements is to give the net displacement $\overrightarrow{\mathrm{PR}}$. Hence, if $\overrightarrow{\mathrm{PR}}$ represents the vector **c** then we can write

$$\mathbf{c} = \mathbf{a} + \mathbf{b}. \tag{A.1}$$

This defines *vector addition*. By completing the parallelogram PQRS, we can also see that

$$\overrightarrow{\mathrm{PR}} = \overrightarrow{\mathrm{PQ}} + \overrightarrow{\mathrm{QR}} = \overrightarrow{\mathrm{PS}} + \overrightarrow{\mathrm{SR}}\,. \tag{A.2}$$

However, $\overrightarrow{\mathrm{PS}}$ has the same length and direction as $\overrightarrow{\mathrm{QR}}$, and, thus, represents the *same* vector, **b**. Likewise, $\overrightarrow{\mathrm{PQ}}$ and $\overrightarrow{\mathrm{SR}}$ both represent the vector **a**. Thus, the above equation is equivalent to

$$\mathbf{c} = \mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}. \tag{A.3}$$

We conclude that the addition of vectors is *commutative*. It can also be shown that the *associative* law holds: *i.e.*,

$$\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}. \tag{A.4}$$

The null vector, **0**, is represented by a displacement of zero length and arbitrary direction. Since the result of combining such a displacement with a finite length displacement is the same as the latter displacement by itself, it follows that

$$\mathbf{a} + \mathbf{0} = \mathbf{a}, \tag{A.5}$$

where **a** is a general vector. The negative of **a** is defined as that vector which has the same magnitude, but acts in the opposite direction, and is denoted $-\mathbf{a}$. The sum of **a** and $-\mathbf{a}$ is thus the null vector: *i.e.*,

$$\mathbf{a} + (-\mathbf{a}) = \mathbf{0}. \tag{A.6}$$

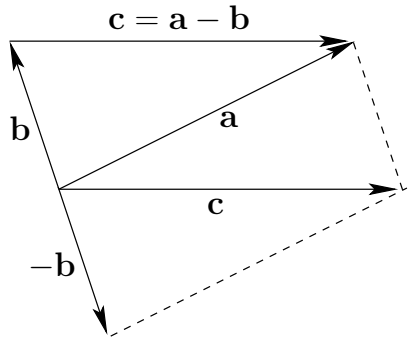


Figure A.3: *Vector subtraction.*

We can also define the difference of two vectors, **a** and **b**, as

$$\mathbf{c} = \mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b}). \tag{A.7}$$

This definition of *vector subtraction* is illustrated in Figure A.3.

If $n > 0$ is a scalar then the expression $n\,\mathbf{a}$ denotes a vector whose direction is the same as **a**, and whose magnitude is n times that of **a**. (This definition becomes obvious when n is an integer.) If n is negative then, since $n\,\mathbf{a} = |n|\,(-\mathbf{a})$, it follows that $n\,\mathbf{a}$ is a vector whose magnitude is $|n|$ times that of **a**, and whose direction is opposite to **a**. These definitions imply that if n and m are two scalars then

$$n\,(m\,\mathbf{a}) = n\,m\,\mathbf{a} = m\,(n\,\mathbf{a}), \tag{A.8}$$

$$(n+m)\,\mathbf{a} = n\,\mathbf{a} + m\,\mathbf{a}, \tag{A.9}$$

$$n\,(\mathbf{a}+\mathbf{b}) = n\,\mathbf{a} + n\,\mathbf{b}. \tag{A.10}$$

A.4 Cartesian Components of a Vector

Consider a Cartesian coordinate system $Oxyz$ consisting of an origin, O, and three mutually perpendicular coordinate axes, Ox, Oy, and Oz—see Figure A.4. Such a system is said to be *right-handed* if, when looking along the Oz direction, a 90° clockwise rotation about Oz is required to take Ox into Oy. Otherwise, it is said to be left-handed. In physics, it is conventional to always use right-handed coordinate systems.

It is convenient to define unit vectors, $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$, parallel to Ox, Oy, and Oz, respectively. Incidentally, a unit vector is a vector whose magnitude is unity. The position vector, **r**, of some general point P whose Cartesian coordinates are (x, y, z) is then given by

$$\mathbf{r} = x\,\mathbf{e}_x + y\,\mathbf{e}_y + z\,\mathbf{e}_z. \tag{A.11}$$

In other words, we can get from O to P by moving a distance x parallel to Ox, then a distance y parallel to Oy, and then a distance z parallel to Oz. Similarly, if **a** is an arbitrary vector then

$$\mathbf{a} = a_x\,\mathbf{e}_x + a_y\,\mathbf{e}_y + a_z\,\mathbf{e}_z, \tag{A.12}$$

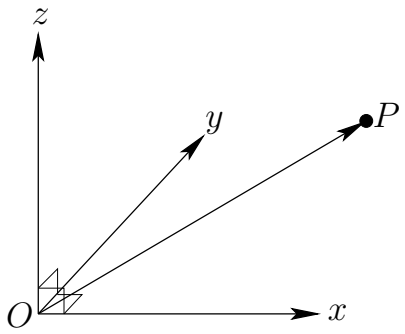


Figure A.4: *A right-handed Cartesian coordinate system.*

where a_x, a_y, and a_z are termed the *Cartesian components* of **a**. It is coventional to write $\mathbf{a} \equiv (a_x, a_y, a_z)$. It follows that $\mathbf{e}_x \equiv (1, 0, 0)$, $\mathbf{e}_y \equiv (0, 1, 0)$, and $\mathbf{e}_z \equiv (0, 0, 1)$. Of course, $\mathbf{0} \equiv (0, 0, 0)$.

According to the three-dimensional generalization of the Pythagorean theorem, the distance $OP \equiv |\mathbf{r}| = r$ is given by

$$r = \sqrt{x^2 + y^2 + z^2}. \tag{A.13}$$

By analogy, the magnitude of a general vector **a** takes the form

$$a = \sqrt{a_x^2 + a_y^2 + a_z^2}. \tag{A.14}$$

If $\mathbf{a} \equiv (a_x, a_y, a_z)$ and $\mathbf{b} \equiv (b_x, b_y, b_z)$ then it is easily demonstrated that

$$\mathbf{a} + \mathbf{b} \equiv (a_x + b_x, a_y + b_y, a_z + b_z). \tag{A.15}$$

Furthermore, if n is a scalar then it is apparent that

$$n\,\mathbf{a} \equiv (n\,a_x, n\,a_y, n\,a_z). \tag{A.16}$$

A.5 Coordinate Transformations

A Cartesian coordinate system allows position and direction in space to be represented in a very convenient manner. Unfortunately, such a coordinate system also introduces arbitrary elements into our analysis. After all, two independent observers might well choose coordinate systems with different origins, and different orientations of the coordinate axes. In general, a given vector **a** will have different sets of components in these two coordinate systems. However, the direction and magnitude of **a** are the same in both cases. Hence, the two sets of components must be related to one another in a very particular fashion. Actually, since vectors are represented by *moveable* line elements in space (*i.e.*, in Figure A.2, $\overrightarrow{PQ}$ and $\overrightarrow{SR}$ represent the *same* vector), it follows that the components of a general vector

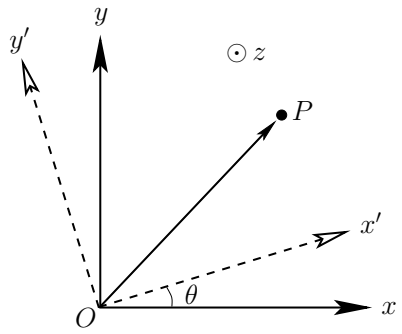


Figure A.5: *Rotation of the coordinate axes about* Oz.

are not affected by a simple shift in the origin of a Cartesian coordinate system. On the other hand, the components are modified when the coordinate axes are rotated.

Suppose that we transform to a new coordinate system, $Ox'y'z'$, which has the same origin as $Oxyz$, and is obtained by rotating the coordinate axes of $Oxyz$ through an angle θ about Oz—see Figure A.5. Let the coordinates of a general point P be (x, y, z) in $Oxyz$ and (x', y', z') in $Ox'y'z'$. According to simple trigonometry, these two sets of coordinates are related to one another via the transformation:

$$x' = x\cos\theta + y\sin\theta, \tag{A.17}$$

$$y' = -x\sin\theta + y\cos\theta, \tag{A.18}$$

$$z' = z. \tag{A.19}$$

Consider the vector displacement $\mathbf{r} \equiv \overrightarrow{OP}$. Note that this displacement is represented by the *same* symbol, **r**, in both coordinate systems, since the magnitude and direction of **r** are manifestly *independent* of the orientation of the coordinate axes. The coordinates of **r** *do* depend on the orientation of the axes: *i.e.*, $\mathbf{r} \equiv (x, y, z)$ in $Oxyz$, and $\mathbf{r} \equiv (x', y', z')$ in $Ox'y'z'$. However, they must depend in a very specific manner [*i.e.*, Equations (A.17)–(A.19)] which preserves the magnitude and direction of **r**.

The components of a general vector **a** transform in an analogous manner to Equations (A.17)–(A.19): *i.e.*,

$$a_{x'} = a_x\cos\theta + a_y\sin\theta, \tag{A.20}$$

$$a_{y'} = -a_x\sin\theta + a_y\cos\theta, \tag{A.21}$$

$$a_{z'} = a_z. \tag{A.22}$$

Moreover, there are similar transformation rules for rotation about Ox and Oy. Equations (A.20)–(A.22) effectively constitute the *definition* of a vector: *i.e.*, the three quantities (a_x, a_y, a_z) are the components of a vector provided that they transform under rotation of the coordinate axes about Oz in accordance with Equations (A.20)–(A.22). (And also transform correctly under rotation about Ox and Oy). Conversely, (a_x, a_y, a_z) *cannot* be

the components of a vector if they do not transform in accordance with Equations (A.20)–(A.22). Of course, scalar quantities are *invariant* under rotation of the coordinate axes. Thus, the individual components of a vector (a_x, say) are real numbers, but they are *not* scalars. Displacement vectors, and all vectors derived from displacements (*e.g.*, velocity, acceleration), automatically satisfy Equations (A.20)–(A.22). There are, however, other physical quantities which have both magnitude and direction, but which are not obviously related to displacements. We need to check carefully to see whether these quantities are really vectors (see Section A.8).

A.6 Scalar Product

A scalar quantity is invariant under all possible rotational transformations. The individual components of a vector are not scalars because they change under transformation. Can we form a scalar out of some combination of the components of one, or more, vectors? Suppose that we were to define the "percent" product,

$$\mathbf{a}\,\%\,\mathbf{b} \equiv a_x\,b_z + a_y\,b_x + a_z\,b_y = \text{scalar number}, \tag{A.23}$$

for general vectors $\mathbf{a}$ and $\mathbf{b}$. Is $\mathbf{a}\,\%\,\mathbf{b}$ invariant under transformation, as must be the case if it is a scalar number? Let us consider an example. Suppose that $\mathbf{a} \equiv (0,\,1,\,0)$ and $\mathbf{b} \equiv (1,\,0,\,0)$. It is easily seen that $\mathbf{a}\,\%\,\mathbf{b} = 1$. Let us now rotate the coordinate axes through 45° about Oz. In the new coordinate system, $\mathbf{a} \equiv (1/\sqrt{2},\,1/\sqrt{2},\,0)$ and $\mathbf{b} \equiv (1/\sqrt{2},\,-1/\sqrt{2},\,0)$, giving $\mathbf{a}\,\%\,\mathbf{b} = 1/2$. Clearly, $\mathbf{a}\,\%\,\mathbf{b}$ is *not* invariant under rotational transformation, so the above definition is a bad one.

Consider, now, the *dot product* or *scalar product*:

$$\mathbf{a}\cdot\mathbf{b} \equiv a_x\,b_x + a_y\,b_y + a_z\,b_z = \text{scalar number}. \tag{A.24}$$

Let us rotate the coordinate axes though θ degrees about Oz. According to Equations (A.20)–(A.22), $\mathbf{a}\cdot\mathbf{b}$ takes the form

$$\begin{aligned}\mathbf{a}\cdot\mathbf{b} &= (a_x\cos\theta + a_y\sin\theta)\,(b_x\cos\theta + b_y\sin\theta)\\ &\quad +(-a_x\sin\theta + a_y\cos\theta)\,(-b_x\sin\theta + b_y\cos\theta) + a_z\,b_z\\ &= a_x\,b_x + a_y\,b_y + a_z\,b_z\end{aligned} \tag{A.25}$$

in the new coordinate system. Thus, $\mathbf{a}\cdot\mathbf{b}$ *is* invariant under rotation about Oz. It can easily be shown that it is also invariant under rotation about Ox and Oy. We conclude that $\mathbf{a}\cdot\mathbf{b}$ is a true scalar, and that the above definition is a good one. Incidentally, $\mathbf{a}\cdot\mathbf{b}$ is the *only* simple combination of the components of two vectors which transforms like a scalar. It is easily shown that the dot product is commutative and distributive: *i.e.*,

$$\begin{aligned}\mathbf{a}\cdot\mathbf{b} &= \mathbf{b}\cdot\mathbf{a},\\ \mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) &= \mathbf{a}\cdot\mathbf{b} + \mathbf{a}\cdot\mathbf{c}.\end{aligned} \tag{A.26}$$

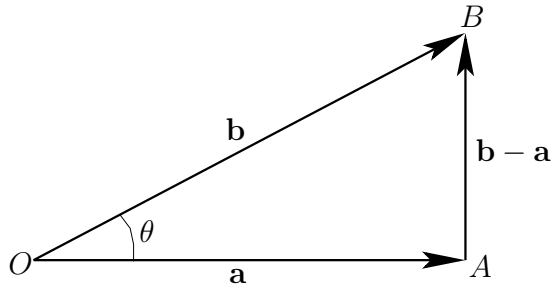


Figure A.6: *A vector triangle.*

The associative property is meaningless for the dot product, because we cannot have $(\mathbf{a}\cdot\mathbf{b})\cdot\mathbf{c}$, since $\mathbf{a}\cdot\mathbf{b}$ is scalar.

We have shown that the dot product $\mathbf{a}\cdot\mathbf{b}$ is coordinate independent. But what is the geometric significance of this? Well, in the special case where $\mathbf{a}=\mathbf{b}$, we get

$$\mathbf{a}\cdot\mathbf{b} = a_x^{\,2} + a_y^{\,2} + a_z^{\,2} = |\mathbf{a}|^2 = a^2. \tag{A.27}$$

So, the invariance of $\mathbf{a}\cdot\mathbf{a}$ is equivalent to the invariance of the magnitude of vector $\mathbf{a}$ under transformation.

Let us now investigate the general case. The length squared of AB in the vector triangle shown in Figure A.6 is

$$(\mathbf{b}-\mathbf{a})\cdot(\mathbf{b}-\mathbf{a}) = |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2\,\mathbf{a}\cdot\mathbf{b}. \tag{A.28}$$

However, according to the "cosine rule" of trigonometry,

$$(AB)^2 = (OA)^2 + (OB)^2 - 2\,(OA)\,(OB)\,\cos\theta, \tag{A.29}$$

where (AB) denotes the length of side AB. It follows that

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}|\,|\mathbf{b}|\,\cos\theta. \tag{A.30}$$

In this case, the invariance of $\mathbf{a}\cdot\mathbf{b}$ under transformation is equivalent to the invariance of the angle subtended between the two vectors. Note that if $\mathbf{a}\cdot\mathbf{b}=0$ then either $|\mathbf{a}|=0$, $|\mathbf{b}|=0$, or the vectors $\mathbf{a}$ and $\mathbf{b}$ are mutually perpendicular. The angle subtended between two vectors can easily be obtained from the dot product: *i.e.*,

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|}. \tag{A.31}$$

The work W performed by a constant force $\mathbf{F}$ which moves an object through a displacement $\mathbf{r}$ is the product of the magnitude of $\mathbf{F}$ times the displacement in the direction of $\mathbf{F}$. If the angle subtended between $\mathbf{F}$ and $\mathbf{r}$ is θ then

$$W = |\mathbf{F}|\,(|\mathbf{r}|\,\cos\theta) = \mathbf{F}\cdot\mathbf{r}. \tag{A.32}$$

The work dW performed by a non-constant force $\mathbf{f}$ which moves an object through an infinitesimal displacement $d\mathbf{r}$ in a time interval dt is $dW = \mathbf{f}\cdot d\mathbf{r}$. Thus, the rate at which the force does work on the object, which is usually referred to as the power, is $P = dW/dt = \mathbf{f}\cdot d\mathbf{r}/dt$, or $P = \mathbf{f}\cdot\mathbf{v}$, where $\mathbf{v} = d\mathbf{r}/dt$ is the object's instantaneous velocity.

A.7 Vector Product

We have discovered how to construct a *scalar* from the components of two general vectors **a** and **b**. Can we also construct a *vector* which is not just a linear combination of **a** and **b**? Consider the following definition:

$$\mathbf{a} * \mathbf{b} \equiv (a_x\, b_x,\ a_y\, b_y,\ a_z\, b_z). \tag{A.33}$$

Is $\mathbf{a} * \mathbf{b}$ a proper vector? Suppose that $\mathbf{a} = (0,\ 1,\ 0)$, $\mathbf{b} = (1,\ 0,\ 0)$. In this case, $\mathbf{a} * \mathbf{b} = \mathbf{0}$. However, if we rotate the coordinate axes through 45° about Oz then $\mathbf{a} = (1/\sqrt{2},\ 1/\sqrt{2},\ 0)$, $\mathbf{b} = (1/\sqrt{2},\ -1/\sqrt{2},\ 0)$, and $\mathbf{a} * \mathbf{b} = (1/2,\ -1/2,\ 0)$. Thus, $\mathbf{a} * \mathbf{b}$ does not transform like a vector, because its magnitude depends on the choice of axes. So, above definition is a bad one.

Consider, now, the *cross product* or *vector product*:

$$\mathbf{a} \times \mathbf{b} \equiv (a_y\, b_z - a_z\, b_y,\ a_z\, b_x - a_x\, b_z,\ a_x\, b_y - a_y\, b_x) = \mathbf{c}. \tag{A.34}$$

Does this rather unlikely combination transform like a vector? Let us try rotating the coordinate axes through an angle θ about Oz using Equations (A.20)–(A.22). In the new coordinate system,

$$\begin{aligned} c_{x'} &= (-a_x \sin\theta + a_y \cos\theta)\, b_z - a_z\, (-b_x \sin\theta + b_y \cos\theta) \\ &= (a_y\, b_z - a_z\, b_y) \cos\theta + (a_z\, b_x - a_x\, b_z) \sin\theta \\ &= c_x \cos\theta + c_y \sin\theta. \end{aligned} \tag{A.35}$$

Thus, the x-component of $\mathbf{a} \times \mathbf{b}$ transforms correctly. It can easily be shown that the other components transform correctly as well, and that all components also transform correctly under rotation about Ox and Oy. Thus, $\mathbf{a} \times \mathbf{b}$ is a proper vector. Incidentally, $\mathbf{a} \times \mathbf{b}$ is the *only* simple combination of the components of two vectors that transforms like a vector (which is non-coplanar with **a** and **b**). The cross product is *anticommutative*,

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}, \tag{A.36}$$

distributive,

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}, \tag{A.37}$$

but is *not* associative,

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \neq (\mathbf{a} \times \mathbf{b}) \times \mathbf{c}. \tag{A.38}$$

The cross product transforms like a vector, which means that it must have a well-defined direction and magnitude. We can show that $\mathbf{a} \times \mathbf{b}$ is *perpendicular* to both **a** and **b**. Consider $\mathbf{a} \cdot \mathbf{a} \times \mathbf{b}$. If this is zero then the cross product must be perpendicular to **a**. Now,

$$\begin{aligned} \mathbf{a} \cdot \mathbf{a} \times \mathbf{b} &= a_x\, (a_y\, b_z - a_z\, b_y) + a_y\, (a_z\, b_x - a_x\, b_z) + a_z\, (a_x\, b_y - a_y\, b_x) \\ &= 0. \end{aligned} \tag{A.39}$$

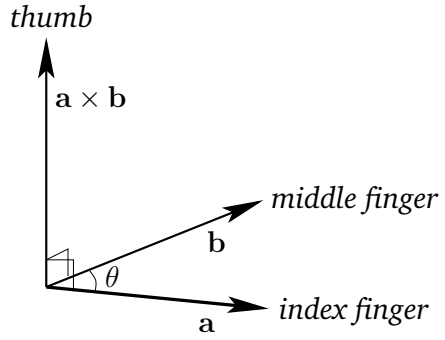


Figure A.7: *The right-hand rule for cross products. Here,* θ *is less that* 180°.

Therefore, $\mathbf{a}\times\mathbf{b}$ is perpendicular to $\mathbf{a}$. Likewise, it can be demonstrated that $\mathbf{a}\times\mathbf{b}$ is perpendicular to $\mathbf{b}$. The vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{a}\times\mathbf{b}$ form a *right-handed* set, like the unit vectors $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$. In fact, $\mathbf{e}_x\times\mathbf{e}_y=\mathbf{e}_z$. This defines a unique direction for $\mathbf{a}\times\mathbf{b}$, which is obtained from a right-hand rule—see Figure A.7.

Let us now evaluate the magnitude of $\mathbf{a}\times\mathbf{b}$. We have

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})^2 &= (a_y\,b_z-a_z\,b_y)^2+(a_z\,b_x-a_x\,b_z)^2+(a_x\,b_y-a_y\,b_x)^2\\ &= (a_x^{\,2}+a_y^{\,2}+a_z^{\,2})\,(b_x^{\,2}+b_y^{\,2}+b_z^{\,2})-(a_x\,b_x+a_y\,b_y+a_z\,b_z)^2\\ &= |\mathbf{a}|^2\,|\mathbf{b}|^2-(\mathbf{a}\cdot\mathbf{b})^2\\ &= |\mathbf{a}|^2\,|\mathbf{b}|^2-|\mathbf{a}|^2\,|\mathbf{b}|^2\,\cos^2\theta=|\mathbf{a}|^2\,|\mathbf{b}|^2\,\sin^2\theta.\end{aligned}\tag{A.40}$$

Thus,

$$|\mathbf{a}\times\mathbf{b}|=|\mathbf{a}|\,|\mathbf{b}|\,\sin\theta,\tag{A.41}$$

where θ is the angle subtended between $\mathbf{a}$ and $\mathbf{b}$. Clearly, $\mathbf{a}\times\mathbf{a}=\mathbf{0}$ for any vector, since θ is always zero in this case. Also, if $\mathbf{a}\times\mathbf{b}=\mathbf{0}$ then either $|\mathbf{a}|=0$, $|\mathbf{b}|=0$, or $\mathbf{b}$ is parallel (or antiparallel) to $\mathbf{a}$.

Consider the parallelogram defined by the vectors $\mathbf{a}$ and $\mathbf{b}$—see Figure A.8. The *scalar area* of the parallelogram is $a\,b\sin\theta$. By convention, the *vector area* has the magnitude of the scalar area, and is normal to the plane of the parallelogram, in the sense obtained from a right-hand circulation rule by rotating $\mathbf{a}$ on to $\mathbf{b}$ (through an acute angle): *i.e.*, if the fingers of the right-hand circulate in the direction of rotation then the thumb of the right-hand indicates the direction of the vector area. So, the vector area is coming out of the page in Figure A.8. It follows that

$$\mathbf{S}=\mathbf{a}\times\mathbf{b},\tag{A.42}$$

Suppose that a force $\mathbf{F}$ is applied at position $\mathbf{r}$—see Figure A.9. The torque about the origin O is the product of the magnitude of the force and the length of the lever arm OQ. Thus, the magnitude of the torque is $|\mathbf{F}|\,|\mathbf{r}|\,\sin\theta$. The direction of the torque is conventionally defined as the direction of the axis through O about which the force tries to

rotate objects, in the sense determined by a right-hand circulation rule. Hence, the torque is out of the page in Figure A.9. It follows that the vector torque is given by

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}. \tag{A.43}$$

The angular momentum, $\mathbf{l}$, of a particle of linear momentum $\mathbf{p}$ and position vector $\mathbf{r}$ is simply defined as the moment of its momentum about the origin: *i.e.*,

$$\mathbf{l} = \mathbf{r} \times \mathbf{p}. \tag{A.44}$$

A.8 Rotation

Let us try to define a rotation vector $\boldsymbol{\theta}$ whose magnitude is the angle of the rotation, θ, and whose direction is parallel to the axis of rotation, in the sense determined by a right-hand circulation rule. Unfortunately, this is not a good vector. The problem is that the addition of rotations is not commutative, whereas vector addition is commuative. Figure A.10 shows the effect of applying two successive 90° rotations, one about Ox, and the other about the Oz, to a standard six-sided die. In the left-hand case, the z-rotation is applied before the x-rotation, and *vice versa* in the right-hand case. It can be seen that the die ends up in two completely different states. In other words, the z-rotation plus the x-rotation does not equal the x-rotation plus the z-rotation. This non-commuting algebra cannot be represented by vectors. So, although rotations have a well-defined magnitude and direction, they are *not* vector quantities.

But, this is not quite the end of the story. Suppose that we take a general vector $\mathbf{a}$ and rotate it about Oz by a *small* angle $\delta\theta_z$. This is equivalent to rotating the coordinate axes about the Oz by $-\delta\theta_z$. According to Equations (A.20)–(A.22), we have

$$\mathbf{a}' \simeq \mathbf{a} + \delta\theta_z\, \mathbf{e}_z \times \mathbf{a}, \tag{A.45}$$

where use has been made of the small angle approximations $\sin\theta \simeq \theta$ and $\cos\theta \simeq 1$. The above equation can easily be generalized to allow small rotations about Ox and Oy by $\delta\theta_x$ and $\delta\theta_y$, respectively. We find that

$$\mathbf{a}' \simeq \mathbf{a} + \delta\boldsymbol{\theta} \times \mathbf{a}, \tag{A.46}$$

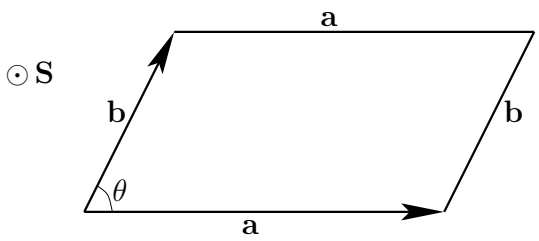


Figure A.8: *A vector parallelogram.*

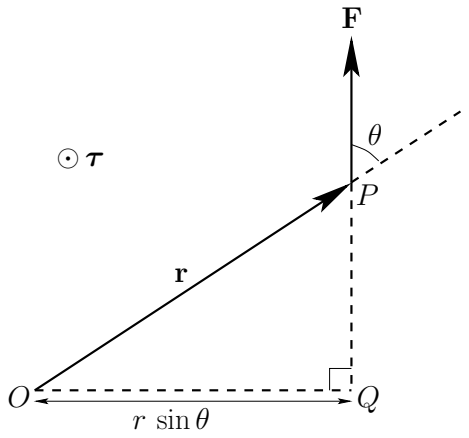


Figure A.9: *A torque.*

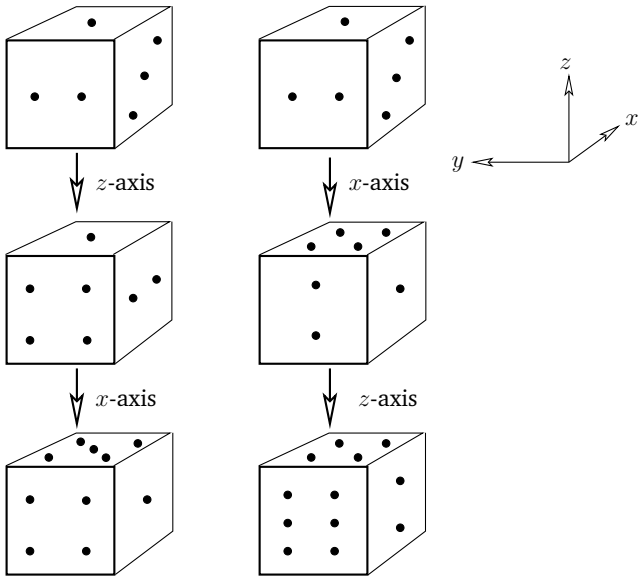


Figure A.10: *Effect of successive rotations about perpendicular axes on a six-sided die.*

where

$$\delta\theta = \delta\theta_x\,\mathbf{e}_x + \delta\theta_y\,\mathbf{e}_y + \delta\theta_z\,\mathbf{e}_z. \tag{A.47}$$

Clearly, we can define a rotation vector, $\delta\theta$, but it only works for *small* angle rotations (*i.e.*, sufficiently small that the small angle approximations of sine and cosine are good). According to the above equation, a small z-rotation plus a small x-rotation is (approximately) equal to the two rotations applied in the opposite order. The fact that infinitesimal rotation is a vector implies that angular velocity,

$$\omega = \lim_{\delta t\rightarrow 0}\frac{\delta\theta}{\delta t}, \tag{A.48}$$

must be a vector as well. Also, if $\mathbf{a}'$ is interpreted as $\mathbf{a}(t+\delta t)$ in Equation (A.46) then it follows that the equation of motion of a vector which precesses about the origin with some angular velocity ω is

$$\frac{d\mathbf{a}}{dt} = \omega\times\mathbf{a}. \tag{A.49}$$

A.9 Scalar Triple Product

Consider three vectors **a**, **b**, and **c**. The *scalar triple product* is defined $\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}$. Now, $\mathbf{b}\times\mathbf{c}$ is the vector area of the parallelogram defined by **b** and **c**. So, $\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}$ is the scalar area of this parallelogram multiplied by the component of **a** in the direction of its normal. It follows that $\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}$ is the *volume* of the parallelepiped defined by vectors **a**, **b**, and **c**—see Figure A.11. This volume is independent of how the triple product is formed from **a**, **b**, and **c**, except that

$$\mathbf{a}\cdot\mathbf{b}\times\mathbf{c} = -\mathbf{a}\cdot\mathbf{c}\times\mathbf{b}. \tag{A.50}$$

So, the "volume" is positive if **a**, **b**, and **c** form a right-handed set (*i.e.*, if **a** lies above the plane of **b** and **c**, in the sense determined from a right-hand circulation rule by rotating **b** onto **c**) and negative if they form a left-handed set. The triple product is unchanged if the dot and cross product operators are interchanged,

$$\mathbf{a}\cdot\mathbf{b}\times\mathbf{c} = \mathbf{a}\times\mathbf{b}\cdot\mathbf{c}. \tag{A.51}$$

The triple product is also invariant under any cyclic permutation of **a**, **b**, and **c**,

$$\mathbf{a}\cdot\mathbf{b}\times\mathbf{c} = \mathbf{b}\cdot\mathbf{c}\times\mathbf{a} = \mathbf{c}\cdot\mathbf{a}\times\mathbf{b}, \tag{A.52}$$

but any anti-cyclic permutation causes it to change sign,

$$\mathbf{a}\cdot\mathbf{b}\times\mathbf{c} = -\mathbf{b}\cdot\mathbf{a}\times\mathbf{c}. \tag{A.53}$$

The scalar triple product is zero if any two of **a**, **b**, and **c** are parallel, or if **a**, **b**, and **c** are coplanar.

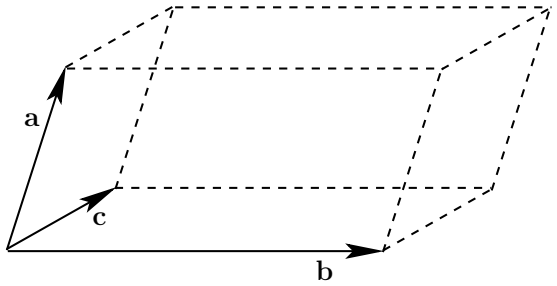


Figure A.11: *A vector parallelepiped.*

If **a**, **b**, and **c** are non-coplanar then any vector **r** can be written in terms of them: *i.e.*,

$$\mathbf{r} = \alpha\,\mathbf{a} + \beta\,\mathbf{b} + \gamma\,\mathbf{c}. \tag{A.54}$$

Forming the dot product of this equation with $\mathbf{b}\times\mathbf{c}$, we then obtain

$$\mathbf{r}\cdot\mathbf{b}\times\mathbf{c} = \alpha\,\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}, \tag{A.55}$$

so

$$\alpha = \frac{\mathbf{r}\cdot\mathbf{b}\times\mathbf{c}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}}. \tag{A.56}$$

Analogous expressions can be written for β and γ. The parameters α, β, and γ are uniquely determined provided $\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}\neq 0$: *i.e.*, provided that the three vectors are non-coplanar.

A.10 Vector Triple Product

For three vectors **a**, **b**, and **c**, the *vector triple product* is defined $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$. The brackets are important because $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})\neq(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$. In fact, it can be demonstrated that

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) \equiv (\mathbf{a}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\,\mathbf{c} \tag{A.57}$$

and

$$(\mathbf{a}\times\mathbf{b})\times\mathbf{c} \equiv (\mathbf{a}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{b}\cdot\mathbf{c})\,\mathbf{a}. \tag{A.58}$$

Let us try to prove the first of the above theorems. The left-hand side and the right-hand side are both proper vectors, so if we can prove this result in one particular coordinate system then it must be true in general. Let us take convenient axes such that Ox lies along **b**, and **c** lies in the x-y plane. It follows that $\mathbf{b}\equiv(b_x,\,0,\,0)$, $\mathbf{c}\equiv(c_x,\,c_y,\,0)$, and $\mathbf{a}\equiv(a_x,\,a_y,\,a_z)$. The vector $\mathbf{b}\times\mathbf{c}$ is directed along Oz: *i.e.*, $\mathbf{b}\times\mathbf{c}\equiv(0,\,0,\,b_x\,c_y)$. Hence, $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ lies in the x-y plane: *i.e.*, $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})\equiv(a_y\,b_x\,c_y,\,-a_x\,b_x\,c_y,\,0)$. This is the left-hand side of Equation (A.57) in our convenient coordinate system. To evaluate the right-hand side, we need $\mathbf{a}\cdot\mathbf{c} = a_x\,c_x + a_y\,c_y$ and $\mathbf{a}\cdot\mathbf{b} = a_x\,b_x$. It follows that the right-hand side is

$$\begin{aligned}\text{RHS} &= (\,[a_x\,c_x + a_y\,c_y]\,b_x,\,0,\,0) - (a_x\,b_x\,c_x,\,a_x\,b_x\,c_y,\,0)\\ &= (a_y\,c_y\,b_x,\,-a_x\,b_x\,c_y,\,0) = \text{LHS},\end{aligned} \tag{A.59}$$

which proves the theorem.

A.11 Vector Calculus

Suppose that vector $\mathbf{a}$ varies with time, so that $\mathbf{a} = \mathbf{a}(t)$. The time derivative of the vector is defined

$$\frac{d\mathbf{a}}{dt} = \lim_{\delta t \rightarrow 0} \left[\frac{\mathbf{a}(t+\delta t) - \mathbf{a}(t)}{\delta t} \right]. \tag{A.60}$$

When written out in component form this becomes

$$\frac{d\mathbf{a}}{dt} \equiv \left(\frac{da_x}{dt}, \frac{da_y}{dt}, \frac{da_z}{dt} \right). \tag{A.61}$$

Suppose that $\mathbf{a}$ is, in fact, the product of a scalar $\phi(t)$ and another vector $\mathbf{b}(t)$. What now is the time derivative of $\mathbf{a}$? We have

$$\frac{da_x}{dt} = \frac{d}{dt}(\phi\, b_x) = \frac{d\phi}{dt}\, b_x + \phi\, \frac{db_x}{dt}, \tag{A.62}$$

which implies that

$$\frac{d\mathbf{a}}{dt} = \frac{d\phi}{dt}\, \mathbf{b} + \phi\, \frac{d\mathbf{b}}{dt}. \tag{A.63}$$

Moreover, it is easily demonstrated that

$$\frac{d}{dt}\,(\mathbf{a}\cdot\mathbf{b}) = \frac{d\mathbf{a}}{dt}\cdot\mathbf{b} + \mathbf{a}\cdot\frac{d\mathbf{b}}{dt}, \tag{A.64}$$

and

$$\frac{d}{dt}\,(\mathbf{a}\times\mathbf{b}) = \frac{d\mathbf{a}}{dt}\times\mathbf{b} + \mathbf{a}\times\frac{d\mathbf{b}}{dt}. \tag{A.65}$$

Hence, it can be seen that the laws of vector differentiation are analogous to those in conventional calculus.

A.12 Line Integrals

Consider a two-dimensional function $f(x,y)$ which is defined for all x and y. What is meant by the integral of f along a given curve joining the points P and Q in the x-y plane? Well, we first draw out f as a function of length l along the path—see Figure A.12. The integral is then simply given by

$$\int_P^Q f(x,y)\, dl = \text{Area under the curve}, \tag{A.66}$$

where $dl = \sqrt{dx^2 + dy^2}$.

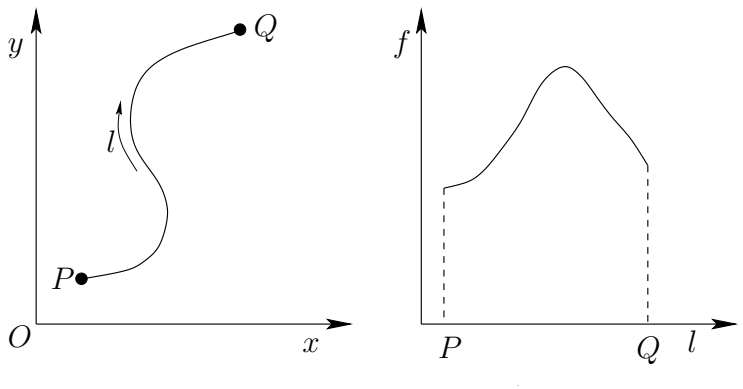


Figure A.12: *A line integral.*

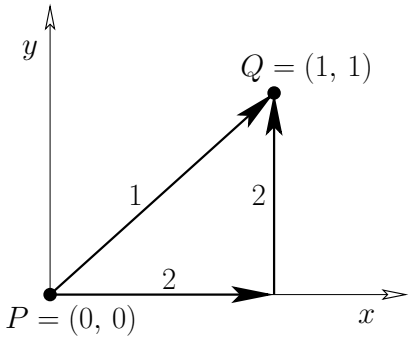


Figure A.13: *An example line integral.*

As an example of this, consider the integral of $f(x, y) = x\,y^2$ between P and Q along the two routes indicated in Figure A.13. Along route 1 we have $x = y$, so $dl = \sqrt{2}\,dx$. Thus,

$$\int_P^Q x\,y^2\,dl = \int_0^1 x^3\,\sqrt{2}\,dx = \frac{\sqrt{2}}{4}. \tag{A.67}$$

The integration along route 2 gives

$$\begin{aligned}\int_P^Q x\,y^2\,dl &= \int_0^1 x\,y^2\,dx\Big|_{y=0} + \int_0^1 x\,y^2\,dy\Big|_{x=1} \\ &= 0 + \int_0^1 y^2\,dy = \frac{1}{3}.\end{aligned} \tag{A.68}$$

Note that the integral depends on the route taken between the initial and final points.

The most common type of line integral is that in which the contributions from dx and dy are evaluated separately, rather that through the path length dl: *i.e.*,

$$\int_P^Q \left[f(x, y)\,dx + g(x, y)\,dy\right]. \tag{A.69}$$

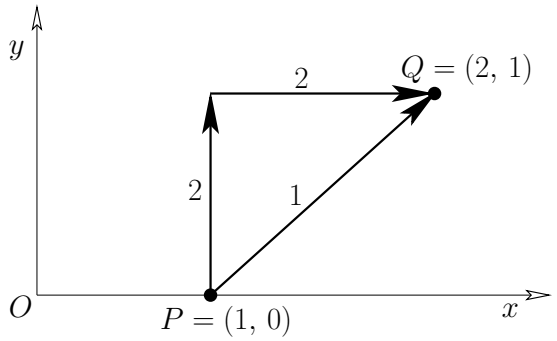


Figure A.14: *An example line integral.*

As an example of this, consider the integral

$$\int_P^Q \left[y\,dx + x^3\,dy \right] \tag{A.70}$$

along the two routes indicated in Figure A.14. Along route 1 we have $x = y + 1$ and $dx = dy$, so

$$\int_P^Q \left[y\,dx + x^3\,dy \right] = \int_0^1 \left[y\,dy + (y+1)^3\,dy \right] = \frac{17}{4}. \tag{A.71}$$

Along route 2,

$$\int_P^Q \left[y\,dx + x^3\,dy \right] = \int_0^1 x^3\,dy \Big|_{x=1} + \int_1^2 y\,dx \Big|_{y=1} = \frac{7}{4}. \tag{A.72}$$

Again, the integral depends on the path of integration.

Suppose that we have a line integral which does *not* depend on the path of integration. It follows that

$$\int_P^Q (f\,dx + g\,dy) = F(Q) - F(P) \tag{A.73}$$

for some function F. Given $F(P)$ for one point P in the x-y plane, then

$$F(Q) = F(P) + \int_P^Q (f\,dx + g\,dy) \tag{A.74}$$

defines $F(Q)$ for all other points in the plane. We can then draw a contour map of $F(x, y)$. The line integral between points P and Q is simply the change in height in the contour map between these two points:

$$\int_P^Q (f\,dx + g\,dy) = \int_P^Q dF(x, y) = F(Q) - F(P). \tag{A.75}$$

Thus,

$$dF(x, y) = f(x, y)\,dx + g(x, y)\,dy. \tag{A.76}$$

For instance, if $F = x^3\,y$ then $dF = 3\,x^2\,y\,dx + x^3\,dy$ and

$$\int_P^Q \left(3\,x^2\,y\,dx + x^3\,dy\right) = \left[x^3\,y\right]_P^Q \tag{A.77}$$

is independent of the path of integration.

It is clear that there are two distinct types of line integral. Those which depend only on their endpoints and not on the path of integration, and those which depend both on their endpoints and the integration path. Later on, we shall learn how to distinguish between these two types (see Section A.15).

A.13 Vector Line Integrals

A *vector field* is defined as a set of vectors associated with each point in space. For instance, the velocity $\mathbf{v}(\mathbf{r})$ in a moving liquid (*e.g.*, a whirlpool) constitutes a vector field. By analogy, a *scalar field* is a set of scalars associated with each point in space. An example of a scalar field is the temperature distribution $T(\mathbf{r})$ in a furnace.

Consider a general vector field $\mathbf{A}(\mathbf{r})$. Let $d\mathbf{r} \equiv (dx, dy, dz)$ be the vector element of line length. Vector line integrals often arise as

$$\int_P^Q \mathbf{A}\cdot d\mathbf{r} = \int_P^Q (A_x\,dx + A_y\,dy + A_z\,dz). \tag{A.78}$$

For instance, if $\mathbf{A}$ is a force-field then the line integral is the work done in going from P to Q.

As an example, consider the work done by a repulsive inverse-square central field, $\mathbf{F} = -\mathbf{r}/|r^3|$. The element of work done is $dW = \mathbf{F}\cdot d\mathbf{r}$. Take $P = (\infty, 0, 0)$ and $Q = (a, 0, 0)$. Route 1 is along the x-axis, so

$$W = \int_\infty^a \left(-\frac{1}{x^2}\right)\,dx = \left[\frac{1}{x}\right]_\infty^a = \frac{1}{a}. \tag{A.79}$$

The second route is, firstly, around a large circle ($r =$ constant) to the point $(a, \infty, 0)$, and then parallel to the y-axis—see Figure A.15. In the first part, no work is done, since $\mathbf{F}$ is perpendicular to $d\mathbf{r}$. In the second part,

$$W = \int_\infty^0 \frac{-y\,dy}{(a^2+y^2)^{3/2}} = \left[\frac{1}{(y^2+a^2)^{1/2}}\right]_\infty^0 = \frac{1}{a}. \tag{A.80}$$

In this case, the integral is independent of the path. However, not all vector line integrals are path independent.

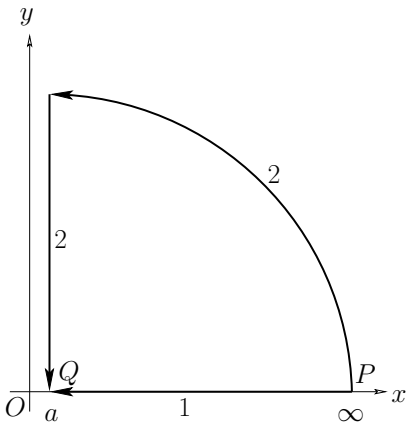


Figure A.15: *An example vector line integral.*

A.14 Volume Integrals

A volume integral takes the form

$$\iiint_V f(x,y,z)\,dV, \tag{A.81}$$

where V is some volume, and $dV = dx\,dy\,dz$ is a small volume element. The volume element is sometimes written $d^3\mathbf{r}$, or even $d\tau$.

As an example of a volume integral, let us evaluate the center of gravity of a solid pyramid. Suppose that the pyramid has a square base of side a, a height a, and is composed of material of uniform density. Let the centroid of the base lie at the origin, and let the apex lie at $(0, 0, a)$. By symmetry, the center of mass lies on the line joining the centroid to the apex. In fact, the height of the center of mass is given by

$$\overline{z} = \iiint z\,dV \Big/ \iiint dV. \tag{A.82}$$

The bottom integral is just the volume of the pyramid, and can be written

$$\begin{aligned}\iiint dV &= \int_0^a dz \int_{-(a-z)/2}^{(a-z)/2} dy \int_{-(a-z)/2}^{(a-z)/2} dx = \int_0^a (a-z)^2\,dz = \int_0^a (a^2 - 2\,a\,z + z^2)\,dz\\ &= \left[a^2\,z - a\,z^2 + z^3/3\right]_0^a = \frac{1}{3}\,a^3.\end{aligned} \tag{A.83}$$

Here, we have evaluated the z-integral last because the limits of the x- and y- integrals are z-dependent. The top integral takes the form

$$\begin{aligned}\iiint z\,dV &= \int_0^a z\,dz \int_{-(a-z)/2}^{(a-z)/2} dy \int_{-(a-z)/2}^{(a-z)/2} dx = \int_0^a z\,(a-z)^2\,dz = \int_0^a (z\,a^2 - 2\,a\,z^2 + z^3)\,dz\\ &= \left[a^2\,z^2/2 - 2\,a\,z^3/3 + z^4/4\right]_0^a = \frac{1}{12}\,a^4.\end{aligned} \tag{A.84}$$

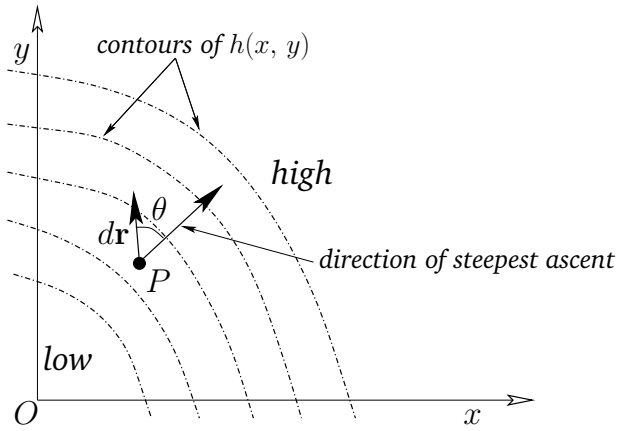


Figure A.16: *A two-dimensional gradient.*

Thus,

$$\bar{z} = \frac{1}{12}\,a^4 \Big/ \frac{1}{3}\,a^3 = \frac{1}{4}\,a. \tag{A.85}$$

In other words, the center of mass of a pyramid lies one quarter of the way between the centroid of the base and the apex.

A.15 Gradient

A one-dimensional function $f(x)$ has a gradient df/dx which is defined as the slope of the tangent to the curve at x. We wish to extend this idea to cover scalar fields in two and three dimensions.

Consider a two-dimensional scalar field $h(x, y)$, which is (say) height above sea-level in a hilly region. Let $d\mathbf{r} \equiv (dx, dy)$ be an element of horizontal distance. Consider dh/dr, where dh is the change in height after moving an infinitesimal distance $d\mathbf{r}$. This quantity is somewhat like the one-dimensional gradient, except that dh depends on the *direction* of $d\mathbf{r}$, as well as its magnitude. In the immediate vicinity of some point P, the slope reduces to an inclined plane—see Figure A.16. The largest value of dh/dr is straight up the slope. It is easily shown that for any other direction

$$\frac{dh}{dr} = \left(\frac{dh}{dr}\right)_{\max} \cos\theta, \tag{A.86}$$

where θ is the angle shown in Figure A.16. Let us define a two-dimensional vector, **grad** h, called the *gradient* of h, whose magnitude is $(dh/dr)_{\max}$, and whose direction is the direction of steepest ascent. The $\cos\theta$ variation exhibited in the above expression ensures that the component of **grad** h in any direction is equal to dh/dr for that direction.

The component of dh/dr in the x-direction can be obtained by plotting out the profile of h at constant y, and then finding the slope of the tangent to the curve at given x. This

quantity is known as the *partial derivative* of h with respect to x at constant y, and is denoted $(\partial h/\partial x)_y$. Likewise, the gradient of the profile at constant x is written $(\partial h/\partial y)_x$. Note that the subscripts denoting constant-x and constant-y are usually omitted, unless there is any ambiguity. If follows that in component form

$$\mathbf{grad}\, h \equiv \left(\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y}\right). \tag{A.87}$$

Now, the equation of the tangent plane at $P = (x_0, y_0)$ is

$$h_T(x, y) = h(x_0, y_0) + \alpha\,(x - x_0) + \beta\,(y - y_0). \tag{A.88}$$

This has the same local gradients as $h(x, y)$, so

$$\alpha = \frac{\partial h}{\partial x}, \qquad \beta = \frac{\partial h}{\partial y}, \tag{A.89}$$

by differentiation of the above. For small $dx = x - x_0$ and $dy = y - y_0$, the function h is coincident with the tangent plane, so

$$dh = \frac{\partial h}{\partial x}\,dx + \frac{\partial h}{\partial y}\,dy. \tag{A.90}$$

But, $\mathbf{grad}\, h \equiv (\partial h/\partial x, \partial h/\partial y)$ and $d\mathbf{r} \equiv (dx, dy)$, so

$$dh = \mathbf{grad}\, h \cdot d\mathbf{r}. \tag{A.91}$$

Incidentally, the above equation demonstrates that $\mathbf{grad}\, h$ is a proper vector, since the left-hand side is a scalar, and, according to the properties of the dot product, the right-hand side is also a scalar provided that $d\mathbf{r}$ and $\mathbf{grad}\, h$ are both proper vectors ($d\mathbf{r}$ is an obvious vector, because it is directly derived from displacements).

Consider, now, a three-dimensional temperature distribution $T(x, y, z)$ in (say) a reaction vessel. Let us define $\mathbf{grad}\, T$, as before, as a vector whose magnitude is $(dT/dr)_{max}$, and whose direction is the direction of the maximum gradient. This vector is written in component form

$$\mathbf{grad}\, T \equiv \left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}, \frac{\partial T}{\partial z}\right). \tag{A.92}$$

Here, $\partial T/\partial x \equiv (\partial T/\partial x)_{y,z}$ is the gradient of the one-dimensional temperature profile at constant y and z. The change in T in going from point P to a neighbouring point offset by $d\mathbf{r} \equiv (dx, dy, dz)$ is

$$dT = \frac{\partial T}{\partial x}\,dx + \frac{\partial T}{\partial y}\,dy + \frac{\partial T}{\partial z}\,dz. \tag{A.93}$$

In vector form, this becomes

$$dT = \mathbf{grad}\, T \cdot d\mathbf{r}. \tag{A.94}$$

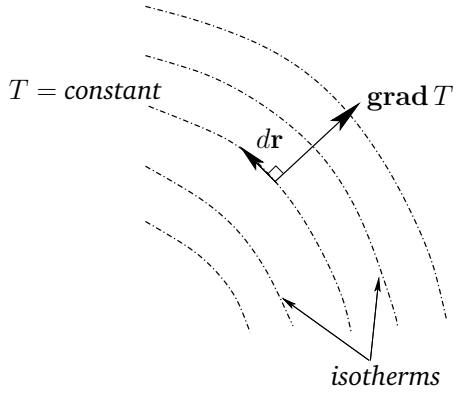


Figure A.17: *Isotherms.*

Suppose that $dT = 0$ for some $d\mathbf{r}$. It follows that

$$dT = \mathbf{grad}\, T \cdot d\mathbf{r} = 0. \tag{A.95}$$

So, $d\mathbf{r}$ is perpendicular to $\mathbf{grad}\, T$. Since $dT = 0$ along so-called "isotherms" (*i.e.*, contours of the temperature), we conclude that the isotherms (contours) are everywhere perpendicular to $\mathbf{grad}\, T$—see Figure A.17.

It is, of course, possible to integrate dT. For instance, the line integral of dT between points P and Q is written

$$\int_P^Q dT = \int_P^Q \mathbf{grad}\, T \cdot d\mathbf{r} = T(Q) - T(P). \tag{A.96}$$

This integral is clearly independent of the path taken between P and Q, so $\int_P^Q \mathbf{grad}\, T \cdot d\mathbf{r}$ must be path independent.

Consider a vector field $\mathbf{A}(\mathbf{r})$. In general, the line integral $\int_P^Q \mathbf{A} \cdot d\mathbf{r}$ depends on the path taken between the end points, but for some special vector fields the integral is path independent. Such fields are called *conservative* fields. It can be shown that if $\mathbf{A}$ is a conservative field then $\mathbf{A} = \mathbf{grad}\, V$ for some scalar field V. The proof of this is straightforward. Keeping P fixed, we have

$$\int_P^Q \mathbf{A} \cdot d\mathbf{r} = V(Q), \tag{A.97}$$

where $V(Q)$ is a well-defined function, due to the path independent nature of the line integral. Consider moving the position of the end point by an infinitesimal amount dx in the x-direction. We have

$$V(Q + dx) = V(Q) + \int_Q^{Q+dx} \mathbf{A} \cdot d\mathbf{r} = V(Q) + A_x\, dx. \tag{A.98}$$

Hence,

$$\frac{\partial V}{\partial x} = A_x, \tag{A.99}$$

with analogous relations for the other components of **A**. It follows that

$$\mathbf{A} = \mathbf{grad}\, V. \tag{A.100}$$

In Newtonian dynamics, the force due to gravity is a good example of a conservative field. Now, if $\mathbf{A}(\mathbf{r})$ is a force-field then $\int \mathbf{A}\cdot \mathrm{d}\mathbf{r}$ is the work done in traversing some path. If **A** is conservative then

$$\oint \mathbf{A}\cdot \mathrm{d}\mathbf{r} = 0, \tag{A.101}$$

where $\oint$ corresponds to the line integral around a *closed* loop. The fact that zero net work is done in going around a closed loop is equivalent to the conservation of energy (which is why conservative fields are called "conservative"). A good example of a non-conservative field is the force due to friction. Clearly, a frictional system loses energy in going around a closed cycle, so $\oint \mathbf{A}\cdot \mathrm{d}\mathbf{r} \neq 0$.

A.16 Grad Operator

It is useful to define the vector *operator*

$$\nabla \equiv \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right), \tag{A.102}$$

which is usually called the *grad* or *del* operator. This operator acts on everything to its right in a expression, until the end of the expression or a closing bracket is reached. For instance,

$$\mathbf{grad}\, f = \nabla f \equiv \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right). \tag{A.103}$$

For two scalar fields ϕ and ψ,

$$\mathbf{grad}\,(\phi\,\psi) = \phi\ \mathbf{grad}\,\psi + \psi\ \mathbf{grad}\,\phi \tag{A.104}$$

can be written more succinctly as

$$\nabla(\phi\,\psi) = \phi\,\nabla\psi + \psi\,\nabla\phi. \tag{A.105}$$

Suppose that we rotate the coordinate axes through an angle θ about Oz. By analogy with Equations (A.17)–(A.19), the old coordinates $(x,\ y,\ z)$ are related to the new ones $(x',\ y',\ z')$ via

$$x = x'\cos\theta - y'\sin\theta, \tag{A.106}$$

$$y = x'\sin\theta + y'\cos\theta, \tag{A.107}$$

$$z = z'. \tag{A.108}$$

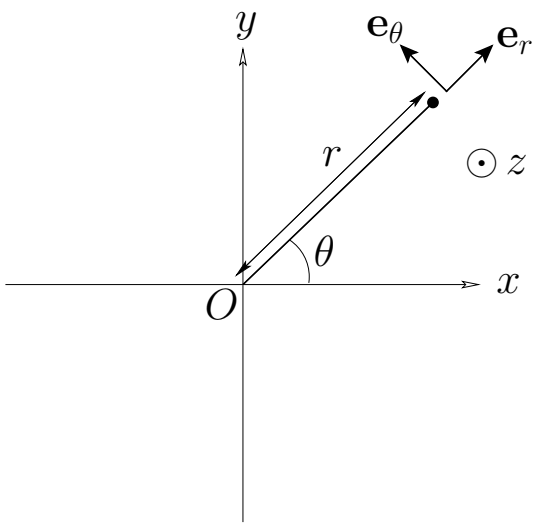


Figure A.18: *Cylindrical polar coordinates.*

Now,

$$\frac{\partial}{\partial x'} = \left(\frac{\partial x}{\partial x'}\right)_{y',z'} \frac{\partial}{\partial x} + \left(\frac{\partial y}{\partial x'}\right)_{y',z'} \frac{\partial}{\partial y} + \left(\frac{\partial z}{\partial x'}\right)_{y',z'} \frac{\partial}{\partial z}, \tag{A.109}$$

giving

$$\frac{\partial}{\partial x'} = \cos\theta \, \frac{\partial}{\partial x} + \sin\theta \, \frac{\partial}{\partial y}, \tag{A.110}$$

and

$$\nabla_{x'} = \cos\theta \, \nabla_x + \sin\theta \, \nabla_y. \tag{A.111}$$

It can be seen, from Equations (A.20)–(A.22), that the differential operator ∇ transforms in an analogous manner to a vector. This is another proof that ∇f is a good vector.

A.17 Curvilinear Coordinates

In the *cylindrical* coordinate system, the Cartesian coordinates x and y are replaced by $r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}(y/x)$. Here, r is the perpendicular distance from the z-axis, and θ the angle subtended between the perpendicular radius vector and the x-axis—see Figure A.18. A general vector $\mathbf{A}$ is thus written

$$\mathbf{A} = A_r \, \mathbf{e}_r + A_\theta \, \mathbf{e}_\theta + A_z \, \mathbf{e}_z, \tag{A.112}$$

where $\mathbf{e}_r = \nabla r/|\nabla r|$ and $\mathbf{e}_\theta = \nabla\theta/|\nabla\theta|$—see Figure A.18. Note that the unit vectors $\mathbf{e}_r$, $\mathbf{e}_\theta$, and $\mathbf{e}_z$ are mutually orthogonal. Hence, $A_r = \mathbf{A}\cdot\mathbf{e}_r$, *etc.* The volume element in this coordinate system is $d^3\mathbf{r} = r\,dr\,d\theta\,dz$. Moreover, the gradient of a general scalar field $V(\mathbf{r})$ takes the form

$$\nabla V = \frac{\partial V}{\partial r}\,\mathbf{e}_r + \frac{1}{r}\frac{\partial V}{\partial \theta}\,\mathbf{e}_\theta + \frac{\partial V}{\partial z}\,\mathbf{e}_z. \tag{A.113}$$

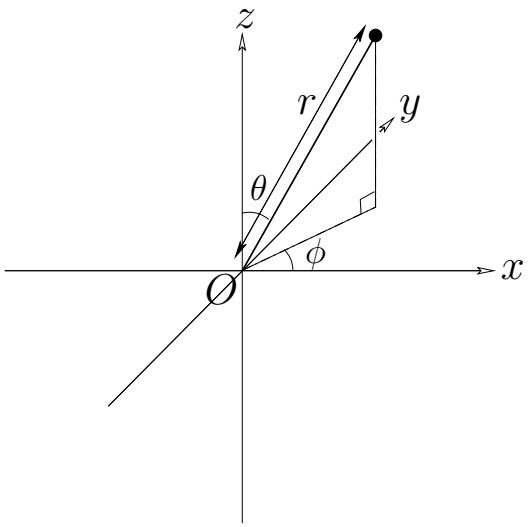


Figure A.19: *Spherical polar coordinates.*

In the *spherical* coordinate system, the Cartesian coordinates x, y, and z are replaced by $r = \sqrt{x^2 + y^2 + z^2}$, $\theta = \cos^{-1}(z/r)$, and $\phi = \tan^{-1}(y/x)$. Here, r is the radial distance from the origin, θ the angle subtended between the radius vector and the z-axis, and ϕ the angle subtended between the projection of the radius vector onto the x-y plane and the x-axis—see Figure A.19. Note that r and θ in the spherical system are *not* the same as their counterparts in the cylindrical system. A general vector $\mathbf{A}$ is written

$$\mathbf{A} = A_r\,\mathbf{e}_r + A_\theta\,\mathbf{e}_\theta + A_\phi\,\mathbf{e}_\phi, \tag{A.114}$$

where $\mathbf{e}_r = \nabla r/|\nabla r|$, $\mathbf{e}_\theta = \nabla\theta/|\nabla\theta|$, and $\mathbf{e}_\phi = \nabla\phi/|\nabla\phi|$. The unit vectors $\mathbf{e}_r$, $\mathbf{e}_\theta$, and $\mathbf{e}_\phi$ are mutually orthogonal. Hence, $A_r = \mathbf{A}\cdot\mathbf{e}_r$, *etc.* The volume element in this coordinate system is $d^3\mathbf{r} = r^2\sin\theta\,dr\,d\theta\,d\phi$. Moreover, the gradient of a general scalar field $V(\mathbf{r})$ takes the form

$$\nabla V = \frac{\partial V}{\partial r}\,\mathbf{e}_r + \frac{1}{r}\frac{\partial V}{\partial\theta}\,\mathbf{e}_\theta + \frac{1}{r\sin\theta}\frac{\partial V}{\partial\phi}\,\mathbf{e}_\phi. \tag{A.115}$$

A.18 Exercises

A.1. The position vectors of the four points A, B, C, and D are $\mathbf{a}$, $\mathbf{b}$, $3\,\mathbf{a} + 2\,\mathbf{b}$, and $\mathbf{a} - 3\,\mathbf{b}$, respectively. Express $\overrightarrow{AC}$, $\overrightarrow{DB}$, $\overrightarrow{BC}$, and $\overrightarrow{CD}$ in terms of $\mathbf{a}$ and $\mathbf{b}$.

A.2. Prove the trigonometric law of sines

$$\frac{\sin a}{A} = \frac{\sin b}{B} = \frac{\sin c}{C}$$

using vector methods. Here, a, b, and c are the three angles of a plane triangle, and A, B, and C the lengths of the corresponding opposite sides.

A.3. Demonstrate using vectors that the diagonals of a parallelogram bisect one another. In addition, show that if the diagonals of a quadrilateral bisect one another then it is a parallelogram.

A.4. From the inequality

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}|\,|\mathbf{b}|\,\cos\theta \leq |\mathbf{a}|\,|\mathbf{b}|$$

deduce the triangle inequality

$$|\mathbf{a}+\mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|.$$

A.5. Find the scalar product $\mathbf{a}\cdot\mathbf{b}$ and the vector product $\mathbf{a}\times\mathbf{b}$ when

(a) $\mathbf{a} = \mathbf{e}_x + 3\,\mathbf{e}_y - \mathbf{e}_z, \quad \mathbf{b} = 3\,\mathbf{e}_x + 2\,\mathbf{e}_y + \mathbf{e}_z,$

(b) $\mathbf{a} = \mathbf{e}_x - 2\,\mathbf{e}_y + \mathbf{e}_z, \quad \mathbf{b} = 2\,\mathbf{e}_x + \mathbf{e}_y + \mathbf{e}_z.$

A.6. Which of the following statements regarding the three general vectors **a**, **b**, and **c** are true?

(a) $\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) = (\mathbf{b}\times\mathbf{a})\cdot\mathbf{c}$.

(b) $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\times\mathbf{b})\times\mathbf{c}$.

(c) $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\,\mathbf{c}$.

(d) $\mathbf{d} = \lambda\,\mathbf{a} + \mu\,\mathbf{b}$ implies that $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{d} = 0$.

(e) $\mathbf{a}\times\mathbf{c} = \mathbf{b}\times\mathbf{c}$ implies that $\mathbf{c}\cdot\mathbf{a} - \mathbf{c}\cdot\mathbf{b} = c\,|\mathbf{a}-\mathbf{b}|$.

(f) $(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{b}) = [\mathbf{b}\cdot(\mathbf{c}\times\mathbf{a})]\,\mathbf{b}$.

A.7. Prove that the length of the shortest straight-line from point **a** to the straight-line joining points **b** and **c** is

$$\frac{|\mathbf{a}\times\mathbf{b} + \mathbf{b}\times\mathbf{c} + \mathbf{c}\times\mathbf{a}|}{|\mathbf{b}-\mathbf{c}|}.$$

A.8. Identify the following surfaces:

(a) $|\mathbf{r}| = a$,

(b) $\mathbf{r}\cdot\mathbf{n} = b$,

(c) $\mathbf{r}\cdot\mathbf{n} = c\,|\mathbf{r}|$,

(d) $|\mathbf{r} - (\mathbf{r}\cdot\mathbf{n})\,\mathbf{n}| = d$.

Here, **r** is the position vector, a, b, c, and d are positive constants, and **n** is a fixed unit vector.

A.9. Let **a**, **b**, and **c** be coplanar vectors related via

$$\alpha\,\mathbf{a} + \beta\,\mathbf{b} + \gamma\,\mathbf{c} = \mathbf{0},$$

where α, β, and γ are not all zero. Show that the condition for the points with position vectors $u\,\mathbf{a}$, $v\,\mathbf{b}$, and $w\,\mathbf{c}$ to be colinear is

$$\frac{\alpha}{u} + \frac{\beta}{v} + \frac{\gamma}{w} = 0.$$

A.10. If $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{r}$ are any vectors, demonstrate that $\mathbf{a} = \mathbf{q} + \lambda\,\mathbf{r}$, $\mathbf{b} = \mathbf{r} + \mu\,\mathbf{p}$, and $\mathbf{c} = \mathbf{p} + \nu\,\mathbf{q}$ are coplanar provided that $\lambda\,\mu\,\nu = -1$, where λ, μ, and ν are scalars. Show that this condition is satisfied when $\mathbf{a}$ is perpendicular to $\mathbf{p}$, $\mathbf{b}$ to $\mathbf{q}$, and $\mathbf{c}$ to $\mathbf{r}$.

A.11. The vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are non-coplanar, and form a non-orthogonal vector base. The vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, defined by

$$\mathbf{A} = \frac{\mathbf{b}\times\mathbf{c}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}},$$

plus cyclic permutations, are said to be *reciprocal vectors*. Show that

$$\mathbf{a} = (\mathbf{B}\times\mathbf{C})/(\mathbf{A}\cdot\mathbf{B}\times\mathbf{C}),$$

plus cyclic permutations.

A.12. In the notation of the previous question, demonstrate that the plane passing through points $\mathbf{a}/\alpha$, $\mathbf{b}/\beta$, and $\mathbf{c}/\gamma$ is normal to the direction of the vector

$$\mathbf{h} = \alpha\,\mathbf{A} + \beta\,\mathbf{B} + \gamma\,\mathbf{C}.$$

In addition, show that the perpendicular distance of the plane from the origin is $|\mathbf{h}|^{-1}$.

A.13. Evaluate $\oint \mathbf{A}\cdot d\mathbf{r}$ for

$$\mathbf{A} = \frac{x\,\mathbf{e}_x + y\,\mathbf{e}_y}{\sqrt{x^2+y^2}}$$

around the square whose sides are $x = 0$, $x = a$, $y = 0$, $y = a$.

A.14. Find the gradients of the following scalar functions of the position vector $\mathbf{r} = (x, y, z)$:

(a) $\mathbf{k}\cdot\mathbf{r}$,

(b) $|\mathbf{r}|^n$,

(c) $|\mathbf{r}-\mathbf{k}|^{-n}$,

(d) $\cos(\mathbf{k}\cdot\mathbf{r})$.

Here, $\mathbf{k}$ is a fixed vector.