CLASSICAL DYNAMICS OF PARTICLES AND SYSTEMS

CHAPTER 1 - Matrices, Vectors, and Vector Calculus

1.1 Introduction

Physical phenomena can be discussed concisely and elegantly through the use of vector methods.

In applying physical "laws" to particular situations, the results must be independent of whether we choose a rectangular or bipolar cylindrical coordinate system.

The results must also be independent of the exact choice of origin for the coordinates.

The use of vectors gives us this independence.

A given physical law will still be correctly represented no matter which coordinate system we decide is most convenient to describe a particular problem.

Also, the use of vector notation provides an extremely compact method of expressing even the most complicated results.

In elementary treatments of vectors, the discussion may start with the statement that "a vector is a quantity that can be represented as a directed line segment."

To be sure, this type of development will yield correct results, and it is even beneficial to impart a certain feeling for the physical nature of a vector

We assume that the reader is familiar with this type of development, but we forego the approach here because we wish to emphasize the relationship that a vector bears to a coordinate transformation.

Therefore, we introduce matrices and matrix notation to describe not only the transformation but the vector as well. We also introduce a type of notation that is readily adapted to the use of tensors, although we do not encounter these objects until the normal course of events requires their use.

We do not attempt a complete exposition of vector methods; instead, we consider only those topics necessary for a study of mechanical systems.

Thus in this chapter, we treat the fundamentals of matrix and vector algebra and vector calculus.

1.2 Concept of a Scalar

Consider the array of particles shown in Figure 1-1a.

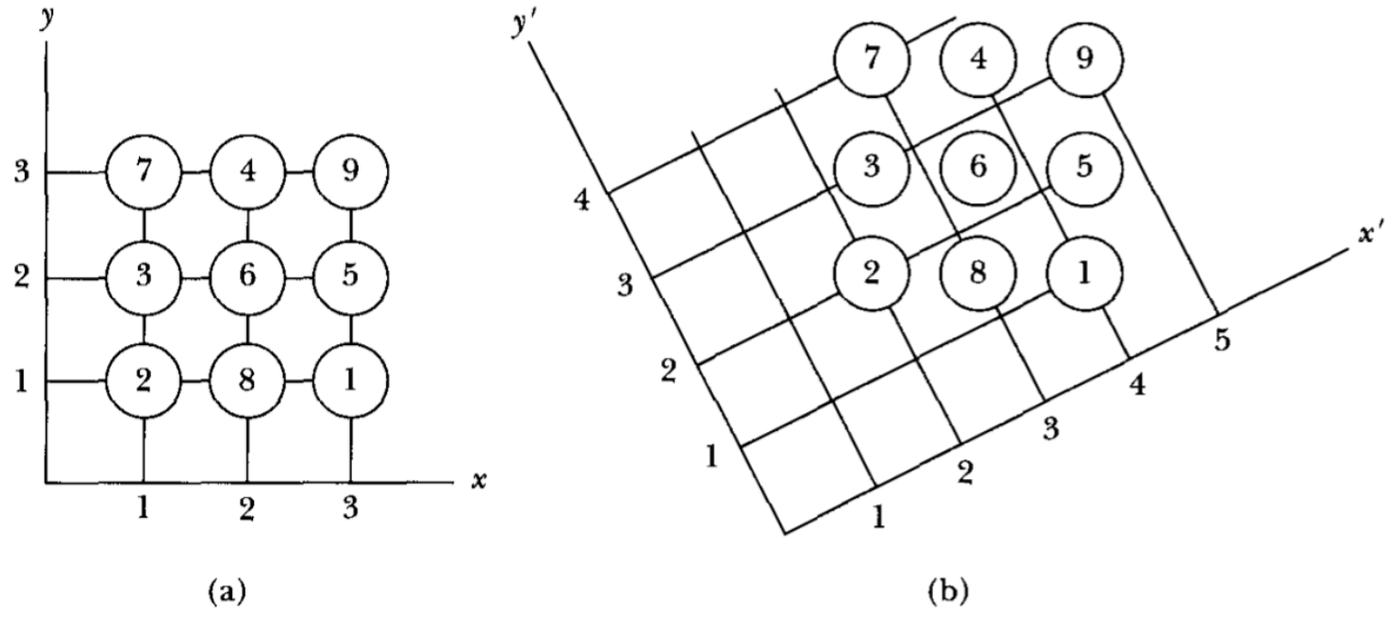


FIGURE 1-1 An array of particles in two different coordinate systems.

Each particle of the array is labeled according to its mass, say, in grams.

The coordinate axes are shown so that we can specify a particular particle by a pair of numbers (x, y).

The mass M of the particle at (x, y) can be expressed as M(x, y); thus the mass of the particle at x = 2, y = 3 can be written as M(x = 2, y = 5) = 4.

Now consider the axes rotated and displaced in the manner shown in Figure 1-lb.

The 4 g mass is now located at x' = 4, y' = 3.5; that is, the mass is specified by M(x' = 4, y' = 3.5) = 4.

And, in general.

$$M(x, y) = M(x', y') \tag{1.1}$$

because the mass of any particle is not affected by a change in the coordinate axes.

Quantities that are *invariant under coordinate transformation*—those that obey an equation of this type—are termed **scalars**.

Although we can describe the mass of a particle (or the temperature, or the speed, etc.) relative to any coordinate system by the same number, some physical properties associated with the particle (such as the direction of motion of the particle or the direction of a force that may act on the particle) cannot be specified in such a simple manner.

The description of these more complicated quantities requires the use of **vectors**.

Just as a scalar is defined as a quantity that remains invariant under a coordinate transformation, a vector may also be defined in terms of transformation properties.

We begin by considering how the coordinates of a point change when the coordinate system rotates around its origin.

1.3 Coordinate Transformations

Consider a point P with coordinates (x_1, x_2, x_3) with respect to a certain coordinate system.

Next consider a different coordinate system, one that can be generated from the original system by a simple rotation; let the coordinates of the point P with respect to the new coordinate system be (x'_1, x'_2, x'_3).

The situation is illustrated for a two-dimensional case in Figure 1-2

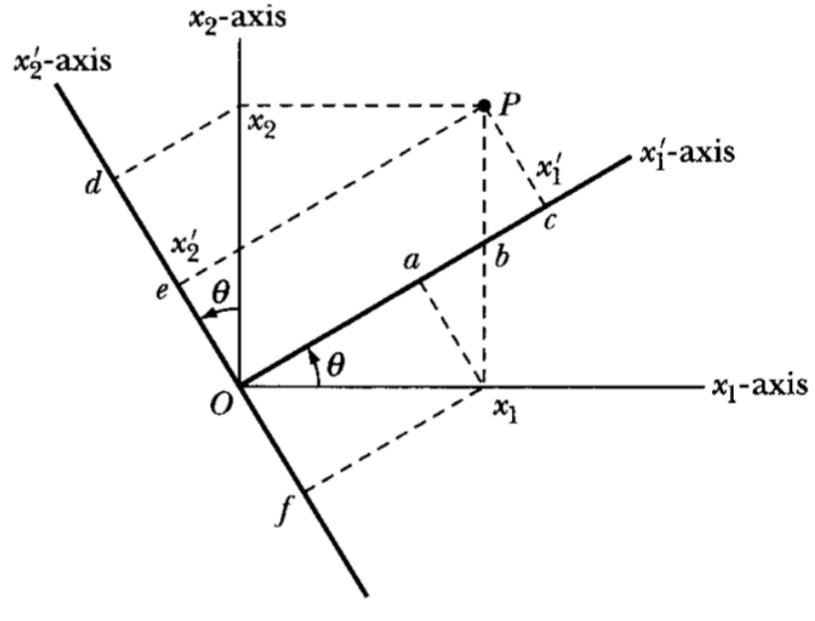


FIGURE 1-2 The position of a point P can be represented in two coordinate systems, one rotated from the other.

The new coordinate x'_1 is the sum of the projection of x_1 onto the x'_1-axis (the line $\overline{Oa}$) plus the projection of x_2 onto the x'_1-axis (the line $\overline{ab} + \overline{bc}$); that is,

$$\begin{aligned} x'_1 &= x_1 \cos\theta + x_2 \sin\theta \\ &= x_1 \cos\theta + x_2 \cos\left(\frac{\pi}{2} - \theta\right) \end{aligned} \tag{1.2a}$$

The coordinate x'_2 is the sum of similar projections: $x'_2 = \overline{Od} - \overline{de}$, but the line $\overline{de}$ is also equal to the line $\overline{Of}$.

Therefore

$$\begin{aligned} x'_2 &= -x_1 \sin\theta + x_2 \cos\theta \\ &= x_1 \cos\left(\frac{\pi}{2} + \theta\right) + x_2 \cos\theta \end{aligned} \tag{1.2b}$$

Let us introduce the following notation: we write the angle between the x'_1-axis and the x_1-axis as (x'_1, x_1), and in general, the angle between the x'_i-axis and the x_j-axis is denoted by (x'_i, x_j).

Furthermore, we define a set of numbers

$$\lambda_{ij} \equiv \cos(x'_i, x_j) \tag{1.3}$$

Therefore, for Figure 1-2, we have

$$\left.\begin{aligned}
\lambda_{11} &= \cos(x_1', x_1) = \cos\theta \\
\lambda_{12} &= \cos(x_1', x_2) = \cos\left(\frac{\pi}{2} - \theta\right) = \sin\theta \\
\lambda_{21} &= \cos(x_2', x_1) = \cos\left(\frac{\pi}{2} + \theta\right) = -\sin\theta \\
\lambda_{22} &= \cos(x_2', x_2) = \cos\theta
\end{aligned}\right\} \tag{1.4}$$

The equations of transformation (Equation 1.2) now become

$$x_1' = x_1\cos(x_1', x_1) + x_2\cos(x_1', x_2) \tag{1.5a}$$
$$= \lambda_{11}x_1 + \lambda_{12}x_2$$
$$x_2' = x_1\cos(x_2', x_1) + x_2\cos(x_2', x_2) \tag{1.5b}$$
$$= \lambda_{21}x_1 + \lambda_{22}x_2$$

Thus, in general, for three dimensions we have

$$\left.\begin{aligned}
x_1' &= \lambda_{11}x_1 + \lambda_{12}x_2 + \lambda_{13}x_3 \\
x_2' &= \lambda_{21}x_1 + \lambda_{22}x_2 + \lambda_{23}x_3 \\
x_3' &= \lambda_{31}x_1 + \lambda_{32}x_2 + \lambda_{33}x_3
\end{aligned}\right\} \tag{1.6}$$

or, in summation notation.

$$x_i' = \sum_{j=1}^{3} \lambda_{ij} x_j, \quad i = 1, 2, 3 \tag{1.7}$$

The inverse transformation is

$$\begin{aligned} x_1 &= x_1' \cos(x_1', x_1) + x_2' \cos(x_2', x_1) + x_3' \cos(x_3', x_1) \\ &= \lambda_{11} x_1' + \lambda_{21} x_2' + \lambda_{31} x_3' \end{aligned}$$

or, in general,

$$x_i = \sum_{j=1}^{3} \lambda_{ji} x_j', \quad i = 1, 2, 3 \tag{1.8}$$

The quantity λ_{ij} is called the **direction cosine** of the x'_i-axis relative to the x_j-axis.

It is convenient to arrange the λ_{ij} into a square array called a **matrix**.

The boldface symbol $\boldsymbol{\lambda}$ denotes the totality of the individual elements λ_{ij} when arranged as follows:

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{pmatrix} \tag{1.9}$$

Once we find the direction cosines relating the two sets of coordinate axes.

Equations 1.7 and 1.8 give the general rules for specifying the coordinates of a point in either system.

When $\boldsymbol{\lambda}$ is defined this way and when it specifies the transformation properties of the coordinates of a point, it is called a **transformation matrix** or a **rotation matrix**.

EXAMPLE: 1.1

A point P is represented in the (x_1, x_2, x_3) system by P(2, 1, 3).

In another coordinate system, the same point is represented as P(x'_1, x'_2, x'_3) where x_2 has been rotated toward x_3 around the x_1-axis by an angle of 30° (Figure 1-3).

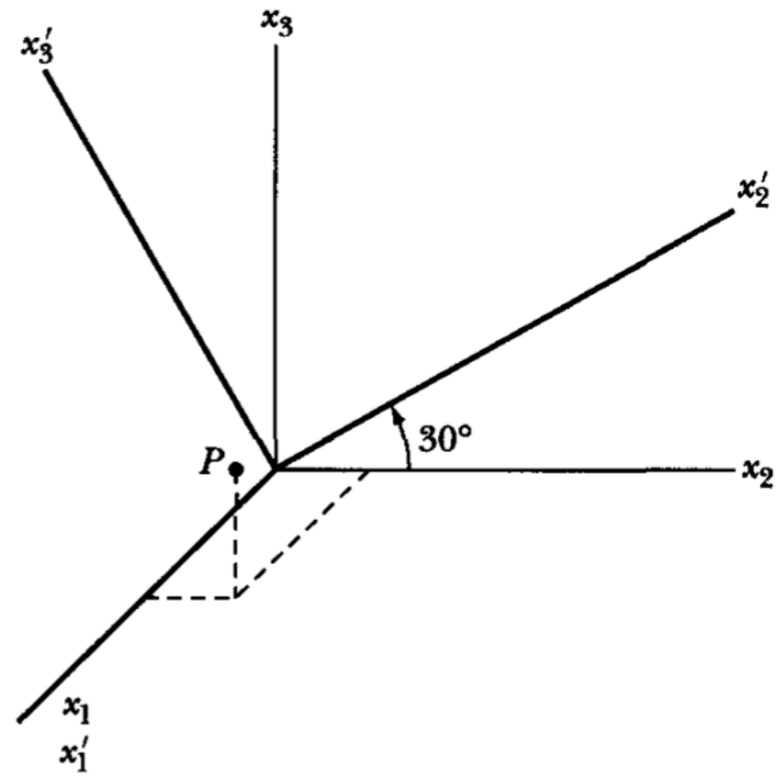


FIGURE 1-3 Example 1.1. A point *P* is represented in two coordinate systems, one rotated from the other by 30°.

Find the rotation matrix and determine P(x'_1, x'_2, x'_3).

Solution. The direction cosines λ_{ij} can be determined from Figure 1-3 using the definition of Equation 1.3.

$$\lambda_{11} = \cos(x'_1, x_1) = \cos(0°) = 1$$
$$\lambda_{12} = \cos(x'_1, x_2) = \cos(90°) = 0$$
$$\lambda_{13} = \cos(x'_1, x_3) = \cos(90°) = 0$$
$$\lambda_{21} = \cos(x'_2, x_1) = \cos(90°) = 0$$
$$\lambda_{22} = \cos(x'_2, x_2) = \cos(30°) = 0.866$$
$$\lambda_{23} = \cos(x'_2, x_3) = \cos(90° - 30°) = \cos(60°) = 0.5$$
$$\lambda_{31} = \cos(x'_3, x_1) = \cos(90°) = 0$$
$$\lambda_{32} = \cos(x'_3, x_2) = \cos(90° + 30°) = -0.5$$
$$\lambda_{33} = \cos(x'_3, x_3) = \cos(30°) = 0.866$$

$$\boldsymbol{\lambda} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.866 & 0.5 \\ 0 & -0.5 & 0.866 \end{pmatrix}$$

and using Equation 1.7, P(x'$_1$, x'$_2$, x'$_3$) is

$$x'_1 = \lambda_{11}x_1 + \lambda_{12}x_2 + \lambda_{13}x_3 = x_1 = 2$$
$$x'_2 = \lambda_{21}x_1 + \lambda_{22}x_2 + \lambda_{23}x_3 = 0.866x_2 + 0.5x_3 = 2.37$$
$$x'_3 = \lambda_{31}x_1 + \lambda_{32}x_2 + \lambda_{33}x_3 = -0.5x_2 + 0.866x_3 = 2.10$$

Notice that the rotation operator preserves the length of the position vector.

$$r = \sqrt{x_1^2 + x_2^2 + x_3^2} = \sqrt{x_1'^2 + x_2'^2 + x_3'^2} = 3.74$$

1.4 Properties of Rotation Matrices

To begin the discussion of rotation matrices, we must recall two trigonometric results. Consider, as in Figure l-4a, a line segment extending in a certain direction in space.

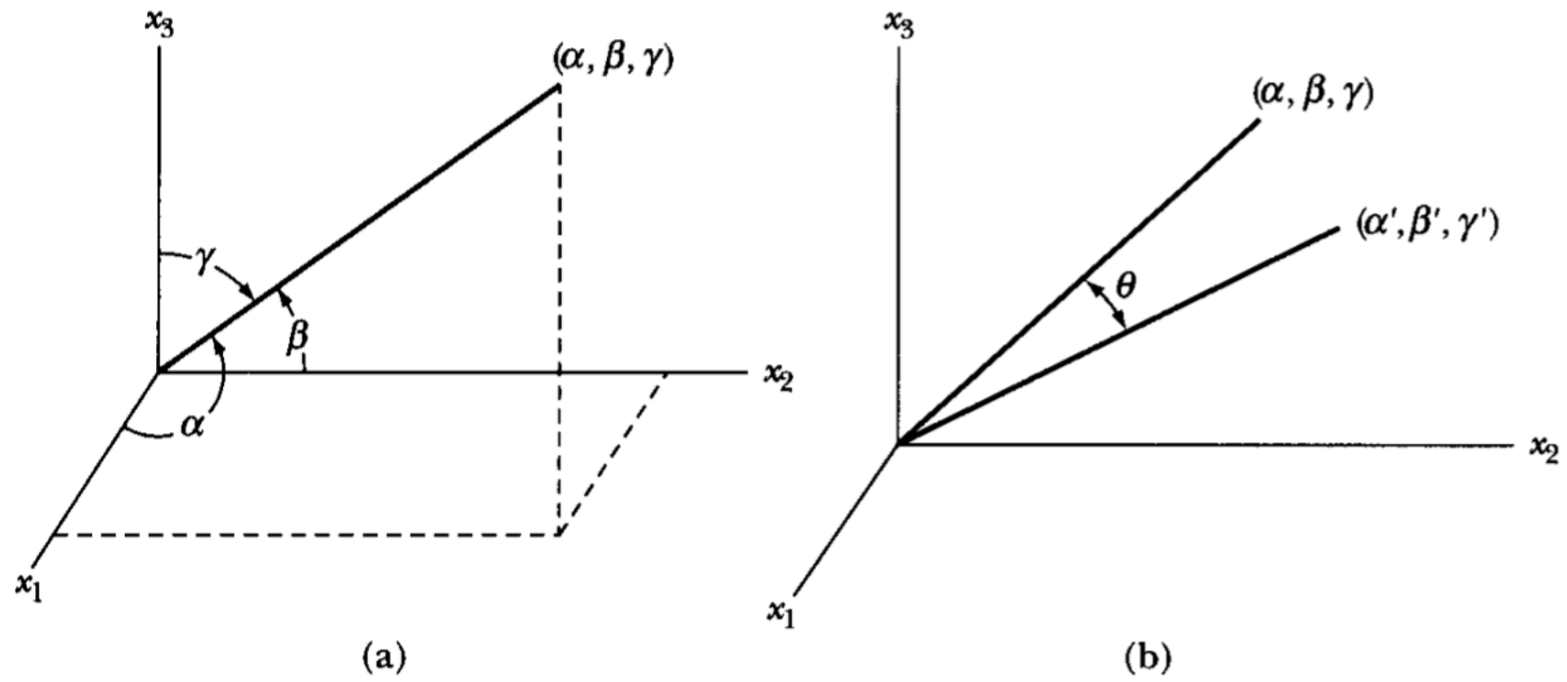


FIGURE 1-4 (a) A line segment is defined by angles (α, β, γ) from the coordinate axes. (b) Another line segment is added that is defined by angles $(\alpha', \beta', \gamma')$.

We choose an origin for our coordinate system that lies at some point on the line. The line then makes certain definite angles with each of the coordinate axes; we let the angles made with the x_1-, x_2-, x_3-axes be α, β, γ . The quantities of interest are the cosines of these angles; $\cos\alpha, \cos\beta, \cos\gamma$. These quantities are called the **direction cosines** of the line. The first result we need is the identity

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1 \tag{1.10}$$

Second, if we have two lines with direction cosines $\cos\alpha, \cos\beta, \cos\gamma$ and $\cos\alpha', \cos\beta', \cos\gamma'$, then the cosine of the angle θ between these lines (see Figure l-4b) is given by

$$\cos\theta = \cos\alpha\cos\alpha' + \cos\beta\cos\beta' + \cos\gamma\cos\gamma' \tag{1.11}$$

With a set of axes x_1, x_2, x_3, let us now perform an arbitrary rotation about some axis through the origin. In the new position, we label the axes x'_1, x'_2, x'_3.

The coordinate rotation may be specified by giving the cosines of all the angles between the various axes, in other words, by the λ_{ij} .

Not all of the nine quantities λ_{ij} , are independent; in fact, six relations exist among the λ_{ij} , so only three are independent. We find these six relations by using the trigonometric results stated in Equations 1.10 and 1.11.

First, the x_1-axis maybe considered alone to be a line in the (x_1, x_2, x_3) coordinate system; the direction cosines of this line are ($\lambda_{11}, \lambda_{12}, \lambda_{13}$). Similarly, the direction cosines of the x_2-axis in the (x_1, x_2, x_3) system are given by ($\lambda_{21}, \lambda_{22}, \lambda_{23}$). Because the angle between the x'_1-axis and the x'_2-axis is $\pi/2$, we have, from Equation 1.11,

$$\lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22} + \lambda_{13}\lambda_{23} = \cos\theta = \cos(\pi/2) = 0$$

or

$$\sum_j \lambda_{1j}\,\lambda_{2j} = 0$$

All summations run from 1 to 3.

And, in general.

$$\sum_j \lambda_{ij}\,\lambda_{kj} = 0, \quad i \neq k \tag{1.12a}$$

Equation 1.12a gives three (one for each value of i or k) of the six relations among the λ_{ij} .

Because the sum of the squares of the direction cosines of a line equals unity (Equation 1.10), we have for the x_j-axis in the (x_1, x_2, x_3) system,

$$\lambda_{11}^2 + \lambda_{12}^2 + \lambda_{13}^2 = 1$$

or

$$\sum_j \lambda_{1j}^2 = \sum_j \lambda_{1j}\lambda_{1j} = 1$$

And, in general,

$$\sum_j \lambda_{ij}\lambda_{kj} = 1, \quad i = k \tag{1.12b}$$

which are the remaining three relations among the λ_{ij} .

We may combine the results given by Equations 1.12a and 1.12b as

$$\boxed{\sum_j \lambda_{ij}\lambda_{kj} = \delta_{ik}} \tag{1.13}$$

where δ_{ik} is the **Kronecker delta symbol**

$$\delta_{ik} = \begin{cases} 0, & \text{if } i \neq k \\ 1, & \text{if } i = k \end{cases} \tag{1.14}$$

The validity of Equation 1.13 depends on the coordinate axes in each of the systems being mutually perpendicular. Such systems are said to be **orthogonal** and Equation 1.13 is the **orthogonality condition**. The transformation matrix $\boldsymbol{\lambda}$ specifying the rotation of any orthogonal coordinate system must then obey Equation 1.13.

If we were to consider the x_i-axes as lines in the x'_i coordinate system and perform a calculation analogous to our preceding calculations, we would find the relation

$$\boxed{\sum_i \lambda_{ij}\lambda_{ik} = \delta_{jk}} \tag{1.15}$$

The two orthogonality relations we have derived (Equations 1.13 and 1.15) appear to be different. (Note: In Equation 1.13 the summation is over the second indices of the λ_{ij}, whereas in Equation 1.15 the summation is over the first indices.) Thus, it seems that we have an overdetermined system: twelve equations in nine unknowns since each of the orthogonality relations represents six equations. Such is not the case, however, because Equations 1.13 and 1.15 are not actually different. In fact, the validity of either of these equations implies the validity of the other. This is clear on physical grounds (because the transformations between the two coordinate systems in either direction are equivalent), and we omit a formal proof. We regard either Equation 1.13 or 1.15 as providing the orthogonality relations for our systems of coordinates.

In the preceding discussion regarding the transformation of coordinates and the properties of rotation matrices, we considered the point P to be fixed and allowed the coordinate axes to be rotated. This interpretation is not unique; we could equally well have maintained the axes fixed and allowed the point to rotate (always keeping constant the distance to the origin). In either event, the transformation matrix is the same. For example, consider the two cases illustrated in Figures l-5a and b. In Figure l-5a, the axes x_1 and x_2 are reference axes, and the x'_1- and x'_2-axes have been obtained by a rotation through an angle θ.

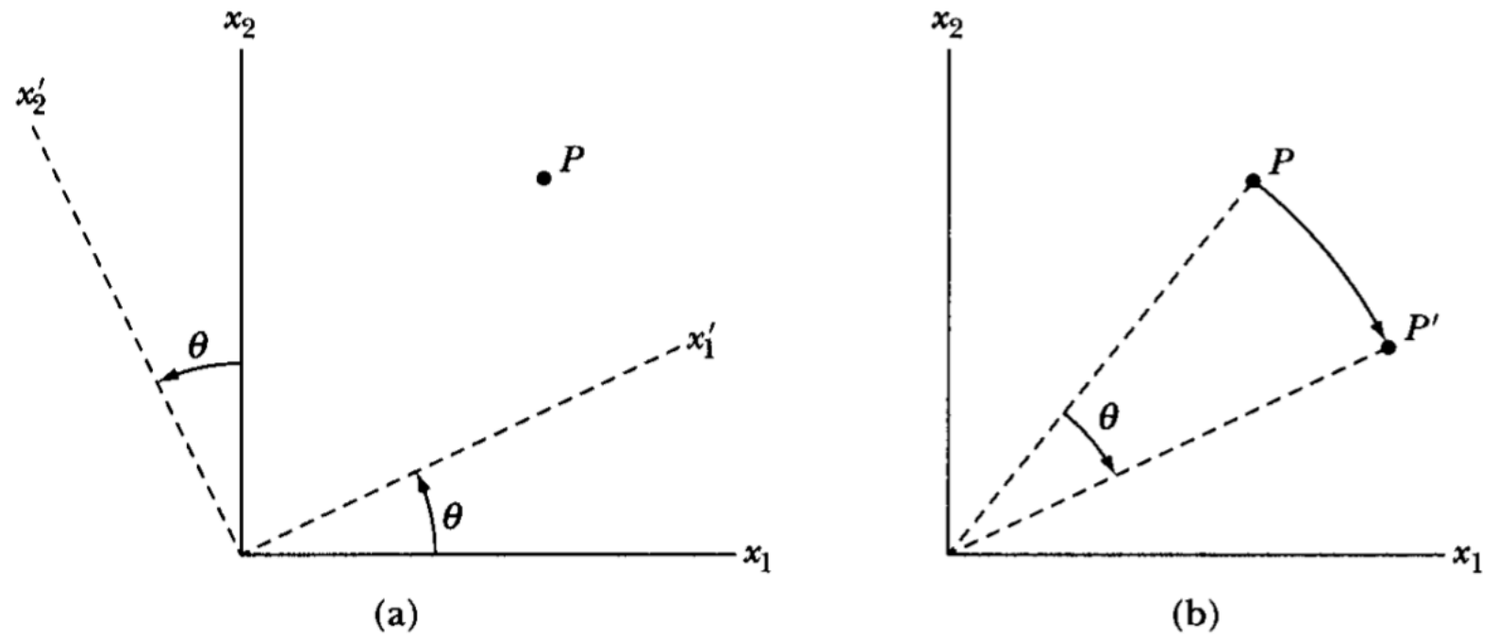


FIGURE 1-5 (a) The coordinate axes x_1, x_2 are rotated by angle θ, but the point P remains fixed. (b) In this case, the coordinates of point P are rotated to a new point P', but not the coordinate system.

Therefore, the coordinates of the point P with respect to the rotated axes may be found (see Equations 1.2a and 1.2b) from

$$\left.\begin{aligned} x_1' &= x_1 \cos\theta + x_2 \sin\theta \\ x_2' &= -x_1 \sin\theta + x_2 \cos\theta \end{aligned}\right\} \tag{1.16}$$

However, if the axes are fixed and the point P is allowed to rotate (as in Figure l-5b) through an angle θ about the origin (but in the opposite sense from that of the rotated axes), then the coordinates of P' are exactly those given by Equation 1.16. Therefore, we may elect to say either that the transformation acts on the point giving a new state of the point expressed with respect to a fixed coordinate system (Figure l-5b) or that the transformation acts on the frame of reference (the coordinate system), as in Figure l-5a. Mathematically, the interpretations are entirely equivalent.

1.5 Matrix Operations

The matrix λ given in Equation 1.9 has equal numbers of rows and columns and is therefore called a **square matrix**. A matrix need not be square. In fact, the coordinates of a point may be written as a **column** matrix

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{1.17a}$$

or as a **row** matrix

$$\mathbf{x} = (x_1 \quad x_2 \quad x_3) \tag{1.17b}$$

We must now establish rules to multiply two matrices. These rules must be consistent with Equations 1.7 and 1.8 when we choose to express the x_i and the x'_i in matrix form. Let us take a column matrix for the coordinates; then we have the following equivalent expressions:

$$x'_i = \sum_j \lambda_{ij} x_j \tag{1.18a}$$

$$\mathbf{x}' = \boldsymbol{\lambda}\mathbf{x} \tag{1.18b}$$

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{1.18c}$$

$$\left.\begin{aligned} x'_1 &= \lambda_{11}x_1 + \lambda_{12}x_2 + \lambda_{13}x_3 \\ x'_2 &= \lambda_{21}x_1 + \lambda_{22}x_2 + \lambda_{23}x_3 \\ x'_3 &= \lambda_{31}x_1 + \lambda_{32}x_2 + \lambda_{33}x_3 \end{aligned}\right\} \tag{1.18d}$$

Equations 1.18a-d completely specify the operation of matrix multiplication for a matrix of three rows and three columns operating on a matrix of three rows and one column. (To be consistent with standard matrix convention we choose **x** and **x**' to be column matrices; multiplication of the type shown in Equation 1.18c is not defined if **x** and **x**' are row matrices.) We must now extend our definition of multiplication to include matrices with arbitrary numbers of rows and columns.

The multiplication of a matrix **A** and a matrix **B** is defined only if the number of columns of **A** is equal to the number of rows of **B**. (The number of rows of **A** and the number of columns of **B** are each arbitrary.) Therefore, in analogy with Equation 1.18a, the product **AB** is given by

$$\mathbf{C} = \mathbf{AB}$$
$$C_{ij} = [\mathbf{AB}]_{ij} = \sum_k A_{ik} B_{kj} \tag{1.19}$$

As an example, let the two matrices **A** and **B** be

$$\mathbf{A} = \begin{pmatrix} 3 & -2 & 2 \\ 4 & -3 & 5 \end{pmatrix}$$
$$\mathbf{B} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & j \end{pmatrix}$$

We multiply the two matrices by

$$\mathbf{AB} = \begin{pmatrix} 3 & -2 & 2 \\ 4 & -3 & 5 \end{pmatrix} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & j \end{pmatrix} \tag{1.20}$$

The product of the two matrices, **C**, is

$$\mathbf{C} = \mathbf{AB} = \begin{pmatrix} 3a - 2d + 2g & 3b - 2e + 2h & 3c - 2f + 2j \\ 4a - 3d + 5g & 4b - 3e + 5h & 4c - 3f + 5j \end{pmatrix} \tag{1.21}$$

To obtain the $\mathbf{C}_{ij}$ element in the *i*th row and *j*th column, we first set the two matrices adjacent as we did in Equation 1.20 in the order **A** and then **B**. We then multiply the individual elements in the *i*th row of **A**, one by one from left to right, times the corresponding elements in the *j*th column of **B**, one by one from top to bottom. We add all these products, and the sum is the $\mathbf{C}_{ij}$ element. Now it is easier to see why a matrix **A** with *m* rows and *n* columns must be multiplied times another matrix **B** with *n* rows and any number of columns, say *p*. The result is a matrix **C** of *m* rows and *p* columns.

EXAMPLE 1.2

Find the product **AB** of the two matrices listed below:

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 3 \\ -2 & 2 & 4 \\ -1 & -3 & -4 \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} -1 & -2 \\ 1 & 2 \\ 3 & 4 \end{pmatrix}$$

Solution. We follow the example of Equations 1.20 and 1.21 to multiply the two matrices together.

$$\mathbf{AB} = \begin{pmatrix} 2 & 1 & 3 \\ -2 & 2 & 4 \\ -1 & -3 & -4 \end{pmatrix}\begin{pmatrix} -1 & -2 \\ 1 & 2 \\ 3 & 4 \end{pmatrix}$$

$$\mathbf{AB} = \begin{pmatrix} -2+1+9 & -4+2+12 \\ 2+2+12 & 4+4+16 \\ 1-3-12 & 2-6-16 \end{pmatrix} = \begin{pmatrix} 8 & 10 \\ 16 & 24 \\ -14 & -20 \end{pmatrix}$$

The result of multiplying a 3 x 3 matrix times a 3 x 2 matrix is a 3 x 2 matrix.

It should be evident from Equation 1.19 that matrix multiplication is not commutative. Thus, if **A** and **B** are both square matrices, then the sums

$$\sum_k A_{ik} B_{kj} \quad \text{and} \quad \sum_k B_{ik} A_{kj}$$

are both defined, but, in general, they will not be equal.

EXAMPLE 1.3

Show that the multiplication of the matrices **A** and **B** in this example is non-commutative.

Solution. If **A** and **B** are the matrices

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ -1 & 3 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} -1 & 2 \\ 4 & -2 \end{pmatrix}$$

then

$$\mathbf{AB} = \begin{pmatrix} 2 & 2 \\ 13 & -8 \end{pmatrix} \quad \text{but} \quad \mathbf{BA} = \begin{pmatrix} -4 & 5 \\ 10 & -2 \end{pmatrix}$$

thus

$$\mathbf{AB} \neq \mathbf{BA}$$

1.6 Further Definitions

A **transposed matrix** is a matrix derived from an original matrix by interchange of rows and columns. We denote the **transpose** of a matrix **A** by $\mathbf{A}^t$. According to the definition, we have

$$\boxed{\lambda_{ij}^t = \lambda_{ji}} \tag{1.22}$$

Evidently,

$$(\boldsymbol{\lambda}^t)^t = \boldsymbol{\lambda} \tag{1.23}$$

Equation 1.8 may therefore be written as any of the following equivalent expressions:

$$x_i = \sum_j \lambda_{ji} x_j' \tag{1.24a}$$

$$x_i = \sum_j \lambda_{ij}^t x_j' \tag{1.24b}$$

$$\mathbf{x} = \boldsymbol{\lambda}^t \mathbf{x}' \tag{1.24c}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} \\ \lambda_{12} & \lambda_{22} & \lambda_{32} \\ \lambda_{13} & \lambda_{23} & \lambda_{33} \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} \tag{1.24d}$$

The identity matrix is that matrix which, when multiplied by another matrix, leaves the latter unaffected. Thus

$$\mathbf{1A} = \mathbf{A}, \quad \mathbf{B1} = \mathbf{B} \tag{1.25}$$

that is.

$$\mathbf{1A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \mathbf{A}$$

Let us consider the orthogonal rotation matrix $\boldsymbol{\lambda}$ for the case of two dimensions:

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}$$

Then

$$\boldsymbol{\lambda}\boldsymbol{\lambda}^t = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} \begin{pmatrix} \lambda_{11} & \lambda_{21} \\ \lambda_{12} & \lambda_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_{11}^2 + \lambda_{12}^2 & \lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22} \\ \lambda_{21}\lambda_{11} + \lambda_{22}\lambda_{12} & \lambda_{21}^2 + \lambda_{22}^2 \end{pmatrix}$$

Using the orthogonality relation (Equation 1.13), we find

$$\lambda_{11}^2 + \lambda_{12}^2 = \lambda_{21}^2 + \lambda_{22}^2 = 1$$

$$\lambda_{21}\lambda_{11} + \lambda_{22}\lambda_{12} = \lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22} = 0$$

so that for the special case of the orthogonal rotation matrix $\boldsymbol{\lambda}$ we have

$$\boldsymbol{\lambda}\boldsymbol{\lambda}^t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{1} \tag{1.26}$$

This result is not valid for matrices in general. It is true only for orthogonal matrices.

The **inverse** of a matrix is defined as that matrix which, when multiplied by the original matrix, produces the identity matrix. The inverse of the matrix $\boldsymbol{\lambda}$ is denoted by $\boldsymbol{\lambda}^{-1}$:

$$\boldsymbol{\lambda}\,\boldsymbol{\lambda}^{-1} = \mathbf{1} \tag{1.27}$$

By comparing Equations 1.26 and 1.27, we find

$$\boxed{\boldsymbol{\lambda}^t = \boldsymbol{\lambda}^{-1}} \quad \text{for orthogonal matrices} \tag{1.28}$$

Therefore, the transpose and the inverse of the rotation matrix $\boldsymbol{\lambda}$ are identical. In fact, the transpose of any orthogonal matrix is equal to its inverse.

To summarize some of the rules of matrix algebra:

1. Matrix multiplication is not commutative in general:

$$\mathbf{AB} \neq \mathbf{BA} \tag{1.29a}$$

The special case of the multiplication of a matrix and its inverse is commutative:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{1} \tag{1.29b}$$

The identity matrix always commutes:

$$\mathbf{1A} = \mathbf{A1} = \mathbf{A} \tag{1.29c}$$

2. Matrix multiplication is associative:

$$[\mathbf{AB}]\mathbf{C} = \mathbf{A}[\mathbf{BC}] \tag{1.30}$$

3. Matrix addition is performed by adding corresponding elements of the two matrices. The components of **C** from the addition **C** = **A** + **B** are

$$C_{ij} = A_{ij} + B_{ij} \tag{1.31}$$

Addition is defined only if **A** and **B** have the same dimensions.

1.7 Geometrical Significance of Transformation Matrices

Consider coordinate axes rotated counterclockwise through an angle of 90° about the x_3-axis, as in Figure 1-6.

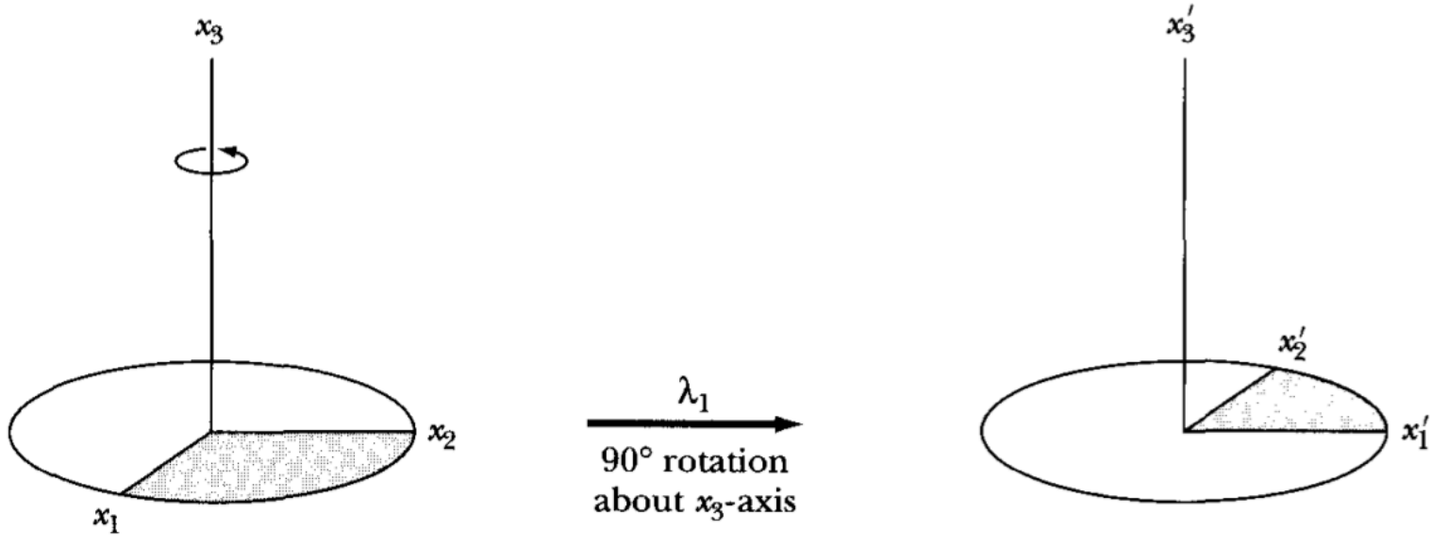


FIGURE 1-6 Coordinate system x_1, x_2, x_3 is rotated 90° counter-clockwise (ccw) about the x_3-axis. This is consistent with the right-hand rule of rotation.

We determine the sense of the rotation by looking along the positive portion of the axis of rotation at the plane being rotated. This definition is then consistent with the "right-hand rule," in which the positive direction of advance of a right-hand screw when turned in the same sense.

In such a rotation, $x'_1 = x_2$, $x'_2 = -x_1$, $x'_3 = x_3$. The only nonvanishing cosines are

$$\cos(x_1', x_2) = \quad 1 = \lambda_{12}$$
$$\cos(x_2', x_1) = -1 = \lambda_{21}$$
$$\cos(x_3', x_3) = \quad 1 = \lambda_{33}$$

so the λ matrix for this case is

$$\boldsymbol{\lambda}_1 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Next consider the counterclockwise rotation through 90° about the x_1-axis, as in Figure 1-7.

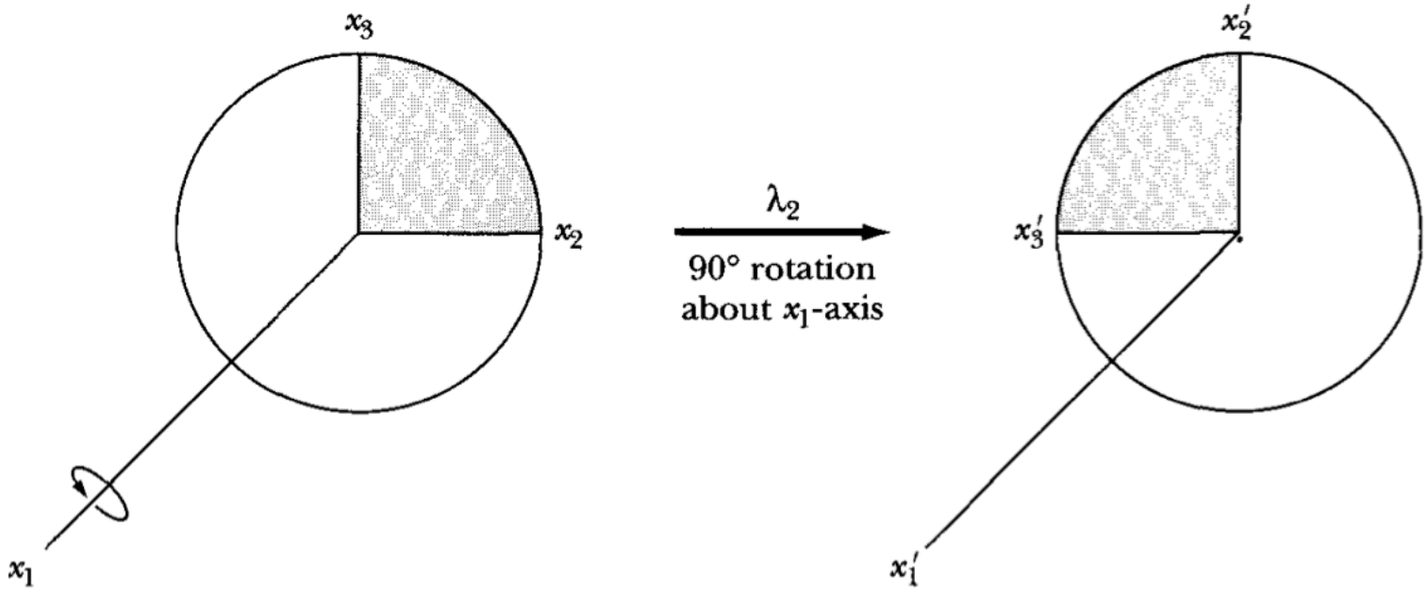


FIGURE 1-7 Coordinate system x_1, x_2, x_3 is rotated 90° ccw about the x_1-axis.

We have $x'_1 = x_1$, $x'_2 = x_3$, $x'_3 = -x_2$, and the transformation matrix is

$$\boldsymbol{\lambda}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

To find the transformation matrix for the combined transformation for rotation about the X_3-axis, followed by rotation about the new x'_1-axis (see Figure 1-8), we have

$$\mathbf{x}' = \boldsymbol{\lambda}_1 \mathbf{x} \tag{1.32a}$$

and

$$\mathbf{x}'' = \boldsymbol{\lambda}_2 \mathbf{x}' \tag{1.32b}$$

or

$$\mathbf{x}'' = \boldsymbol{\lambda}_2 \boldsymbol{\lambda}_1 \mathbf{x} \tag{1.33a}$$

$$\begin{pmatrix} x_1'' \\ x_2'' \\ x_3'' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_3 \\ x_1 \end{pmatrix} \tag{1.33b}$$

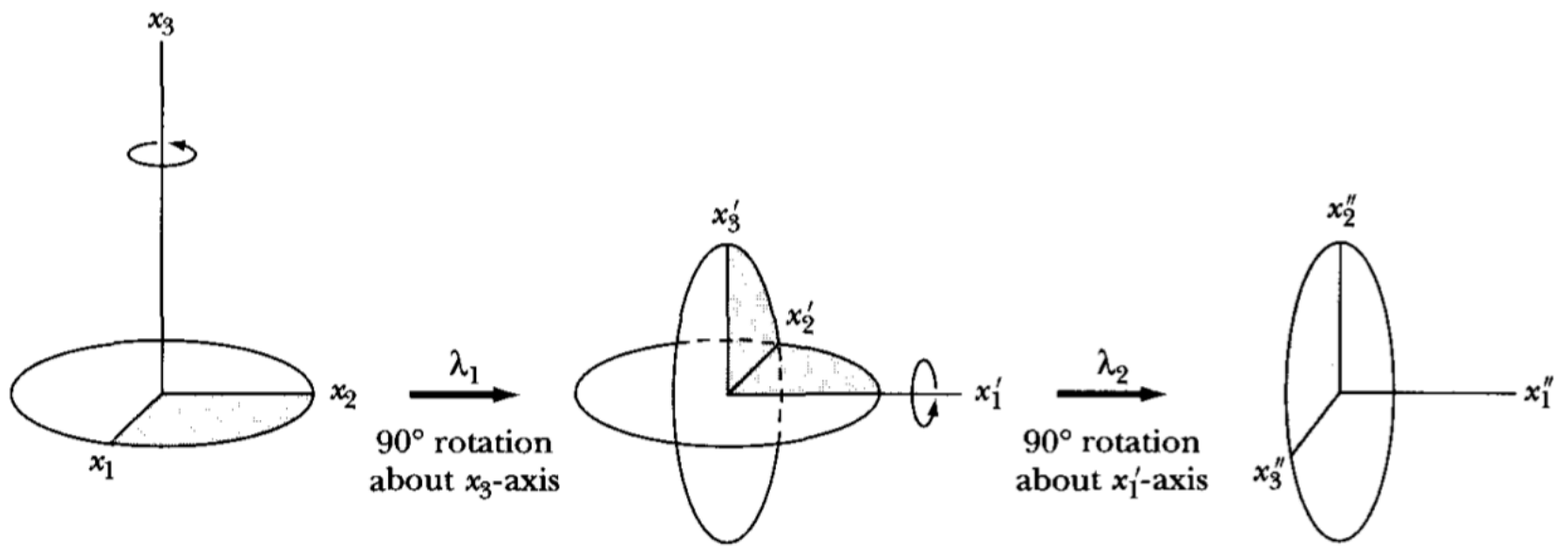


FIGURE 1-8 Coordinate system x_1, x_2, x_3 is rotated 90° ccw about the x_3-axis followed by a 90° rotation about the intermediate x_1'-axis.

Therefore, the two rotations alreadv described mav be represented by a single transformation matrix:

$$\boldsymbol{\lambda}_3 = \boldsymbol{\lambda}_2\boldsymbol{\lambda}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{1.34}$$

and the final orientation is specified by $x''_1 = x_2$, $x''_2 = x_3$, $x''_3 = x_1$. Note that the order in which the transformation matrices operate on X is important because the multiplication is not commutative. In the other order,

$$
\begin{aligned}
\boldsymbol{\lambda}_4 &= \boldsymbol{\lambda}_1\boldsymbol{\lambda}_2 \\
&= \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \neq \boldsymbol{\lambda}_3
\end{aligned}
\tag{1.35}
$$

and an entirely different orientation results. Figure 1-9 illustrates the different final orientations of a parallelepiped that undergoes rotations corresponding to two rotation matrices $\boldsymbol{\lambda}_A, \boldsymbol{\lambda}_B$ when successive rotations are made in different order. The upper portion of the figure represents the matrix product $\boldsymbol{\lambda}_B\boldsymbol{\lambda}_A$, and the lower portion represents the product $\boldsymbol{\lambda}_A\boldsymbol{\lambda}_B$.

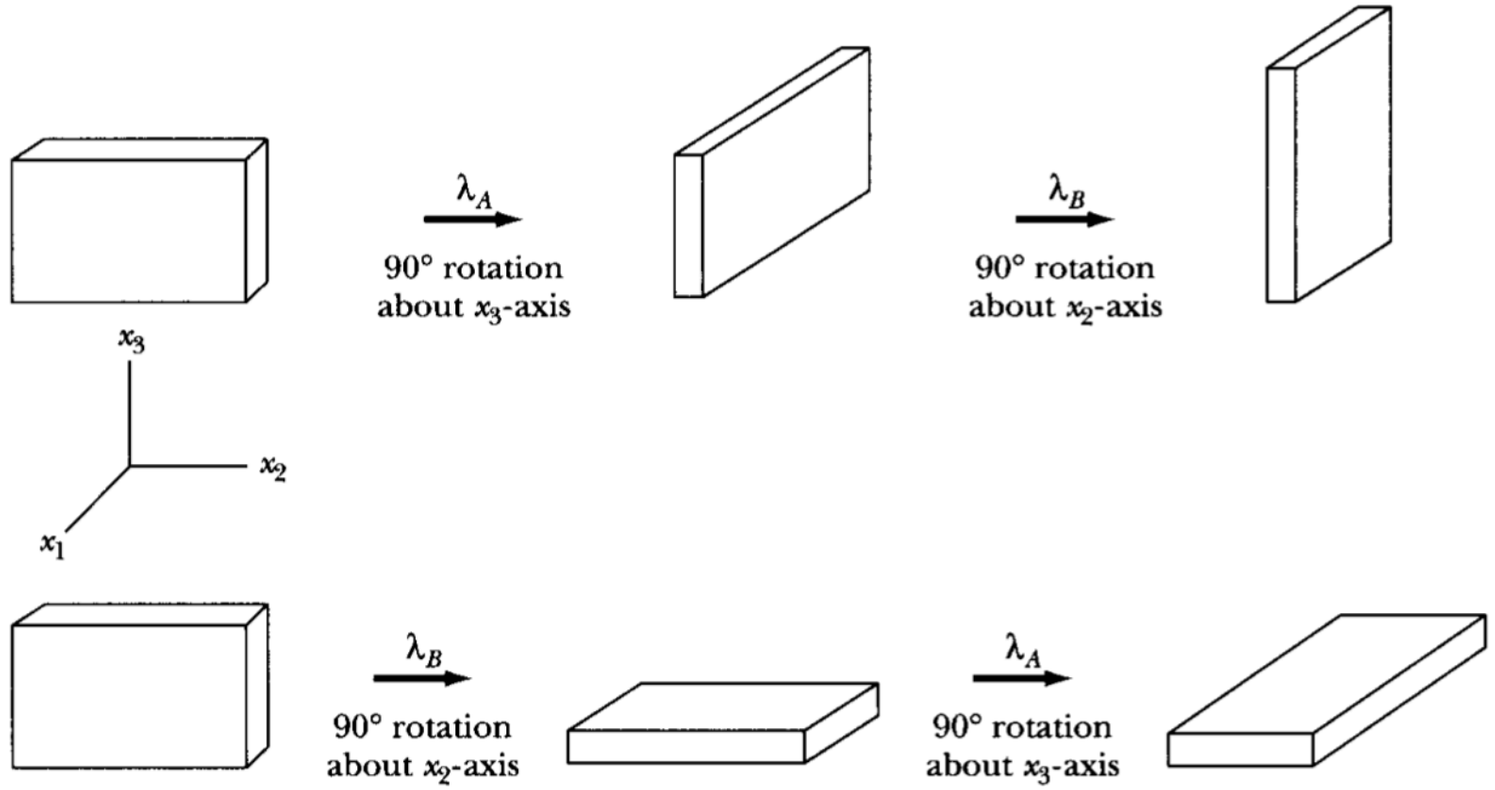


FIGURE 1-9 A parallelepiped undergoes two successive rotations in different order. The results are different.

Next, consider the coordinate rotation pictured in Figure 1-10 (which is the same as that in Figure 1-2).

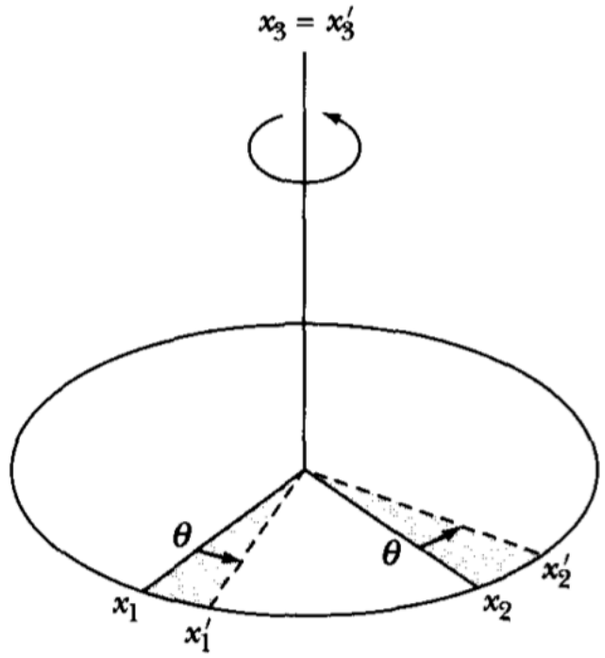


FIGURE 1-10 Coordinate system x_1, x_2, x_3 is rotated an angle θ ccw about the x_3-axis.

The elements of the transformation matrix in two dimensions are given by the following cosines:

$$\cos(x_1', x_1) = \cos\theta = \lambda_{11}$$
$$\cos(x_1', x_2) = \cos\left(\frac{\pi}{2} - \theta\right) = \sin\theta = \lambda_{12}$$
$$\cos(x_2', x_1) = \cos\left(\frac{\pi}{2} + \theta\right) = -\sin\theta = \lambda_{21}$$
$$\cos(x_2', x_2) = \cos\theta = \lambda_{22}$$

Therefore, the matrix is

$$\boldsymbol{\lambda}_5 = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{1.36a}$$

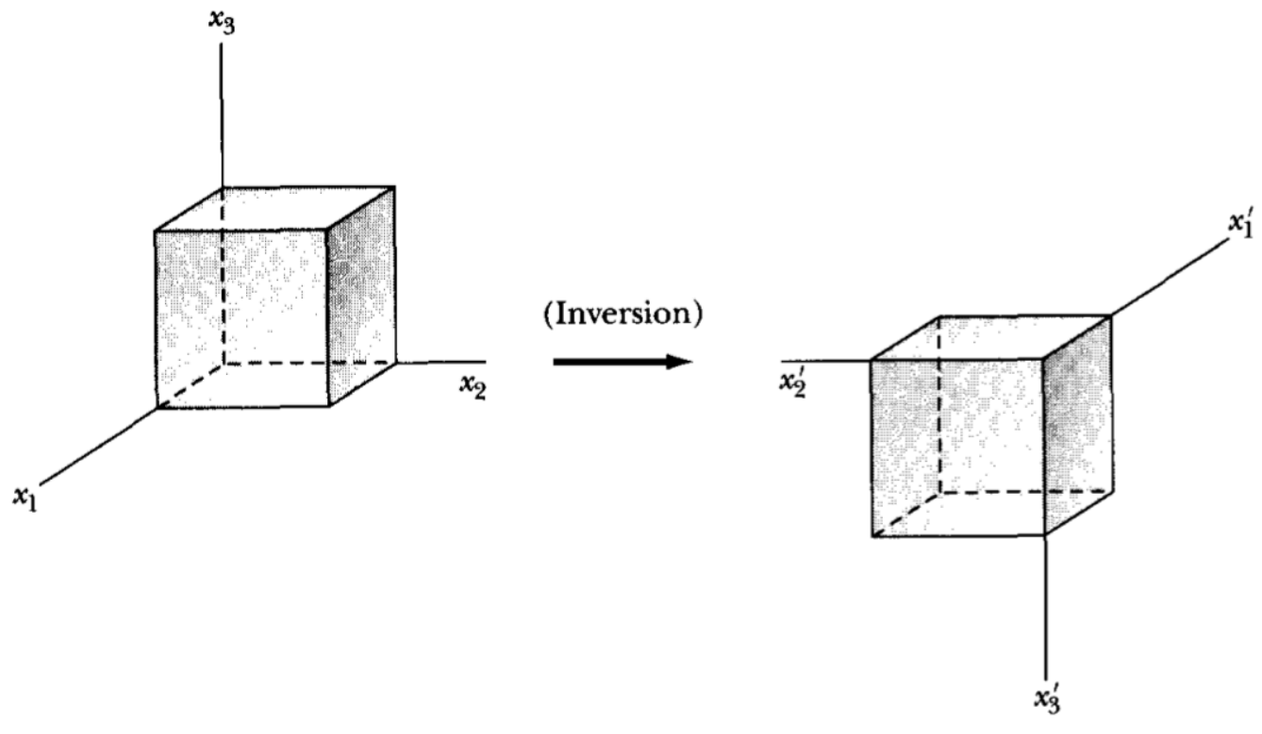
If this rotation were a three-dimensional rotation with $x'_3 = x_3$ would have the following additional cosines:

$$\cos(x'_1, x_3) = 0 = \lambda_{13}$$
$$\cos(x'_2, x_3) = 0 = \lambda_{23}$$
$$\cos(x'_3, x_3) = 1 = \lambda_{33}$$
$$\cos(x'_3, x_1) = 0 = \lambda_{31}$$
$$\cos(x'_3, x_2) = 0 = \lambda_{32}$$

and the three-dimensional transformation matrix is

$$\boldsymbol{\lambda}_5 = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1.36b}$$

As a final example, consider the transformation that results in the reflection through the origin of all the axes, as in Figure 1-11



Such a transformation is called an **inversion**. In such a case, $x'_1 = -x_1$, $x'_2 = -x_2$, $x'_3 = -x_3$, and

$$\boldsymbol{\lambda}_6 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{1.37}$$

In the preceding examples, we defined the transformation matrix $\boldsymbol{\lambda}_3$ to be the result of two successive rotations, each of which was an orthogonal transformation: $\boldsymbol{\lambda}_3 = \boldsymbol{\lambda}_2\boldsymbol{\lambda}_1$. We can prove that the successive application of orthogonal transformations always results in an orthogonal transformation. We write

$$x_i' = \sum_j \lambda_{ij} x_j, \qquad x_k'' = \sum_i \mu_{ki} x_i'$$

Combining these expressions, we obtain

$$x_k'' = \sum_j \left(\sum_i \mu_{ki} \lambda_{ij} \right) x_j$$

$$= \sum_j [\boldsymbol{\mu\lambda}]_{kj} x_j$$

Thus, we accomplish the transformation from x_i, to x''_i by operating on x_i, with the ($\boldsymbol{\mu\lambda}$) matrix. The combined transformation will then be shown to be orthogonal if ($\boldsymbol{\mu\lambda}$)^t = ($\boldsymbol{\mu\lambda}$)$^{-1}$. The transpose of a product matrix is the product of the transposed matrices taken in reverse order; that is, (**AB**)^t = **B**^{t} **A**^{t}. Therefore

$$(\boldsymbol{\mu\lambda})^t = \boldsymbol{\lambda}^t\boldsymbol{\mu}^t \tag{1.38}$$

But, because $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, are orthogonal. $\boldsymbol{\lambda}^t = \boldsymbol{\lambda}^{-1}$ and $\boldsymbol{\mu}^t = \boldsymbol{\mu}^{-1}$. Multiplying the above equation by $\boldsymbol{\mu\lambda}$ from the right, we obtain

$$
\begin{aligned}
(\boldsymbol{\mu\lambda})^t\boldsymbol{\mu\lambda} &= \boldsymbol{\lambda}^t\boldsymbol{\mu}^t\boldsymbol{\mu\lambda} \\
&= \boldsymbol{\lambda}^t\mathbf{1}\boldsymbol{\lambda} \\
&= \boldsymbol{\lambda}^t\boldsymbol{\lambda} \\
&= \mathbf{1} \\
&= (\boldsymbol{\mu\lambda})^{-1}\boldsymbol{\mu\lambda}
\end{aligned}
$$

Hence

$$
(\boldsymbol{\mu\lambda}^t) = (\boldsymbol{\mu\lambda})^{-1} \tag{1.39}
$$

and the $\boldsymbol{\mu\lambda}$ matrix is orthogonal.

The determinants of all the rotation matrices in the preceding examples can be calculated according to the standard rule for the evaluation of determinants of second or third order:

$$
|\boldsymbol{\lambda}| = \begin{vmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{vmatrix} = \lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21} \tag{1.40}
$$

$$
\begin{aligned}
|\boldsymbol{\lambda}| &= \begin{vmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{vmatrix} \\
&= \lambda_{11}\begin{vmatrix} \lambda_{22} & \lambda_{23} \\ \lambda_{32} & \lambda_{33} \end{vmatrix} - \lambda_{12}\begin{vmatrix} \lambda_{21} & \lambda_{23} \\ \lambda_{31} & \lambda_{33} \end{vmatrix} + \lambda_{13}\begin{vmatrix} \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \end{vmatrix}
\end{aligned} \tag{1.41}
$$

where the third-order determinant has been expanded in minors of the first row. Therefore, we find, for the rotation matrices used in this section.

$$|\boldsymbol{\lambda}_1| = |\boldsymbol{\lambda}_2| = \cdots = |\boldsymbol{\lambda}_5| = 1$$

but

$$|\boldsymbol{\lambda}_6| = -1$$

Thus, all those transformations resulting from rotations starting from the original set of axes have determinants equal to +1. But an inversion cannot be generated by any series of rotations, and the determinant of an inversion matrix is equal to -1. Orthogonal transformations, the determinant of whose matrices is +1, are called **proper rotations**; those with determinants equal to -1 are called **improper rotations**. *All* orthogonal matrices must have a determinant equal to either +1 or -1. Here, we confine our attention to the effect of proper rotations and do not concern ourselves with the special properties of vectors manifest in improper rotations.

EXAMPLE 1.4

Show that $|\boldsymbol{\lambda}_2| = 1$ and $|\boldsymbol{\lambda}_6| = -1$.

Solution.

$$|\boldsymbol{\lambda}_2| = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{vmatrix} = +1 \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix} = 0 - (-1) = 1$$

$$|\boldsymbol{\lambda}_6| = \begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{vmatrix} = -1 \begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix} = -1(1 - 0) = -1$$

1.8 Definitions of a Scalar and a Vector in Terms of Transformation Properties

Consider a coordinate transformation of the type

$$x_i' = \sum_j \lambda_{ij} x_j \tag{1.42}$$

with

$$\sum_j \lambda_{ij} \lambda_{kj} = \delta_{ik} \tag{1.43}$$

If, under such a transformation, a quantity ϕ is unaffected, then ϕ is called a **scalar** (or **scalar invariant**).

If a set of quantities (A_1, A_2, A_3) is transformed from the x_i system to the x'_i-system by a transformation matrix $\boldsymbol{\lambda}$ with the result

$$\boxed{A_i' = \sum_j \lambda_{ij} A_j} \tag{1.44}$$

then the quantities A_i transform as the coordinates of a point (i.e., according to Equation 1.42), and the quantity $\mathbf{A} = (A_1, A_2, A_3)$ is termed a **vector**.

1.9 Elementary Scalar and Vector Operations

In the following, **A** and **B** are vectors (with components A_i and B_j) and ϕ, ψ, and ξ are scalars.

Addition

$$A_i + B_i = B_i + A_i \qquad \text{Commutative law} \tag{1.45}$$

$$A_i + (B_i + C_i) = (A_i + B_i) + C_i \qquad \text{Associative law} \tag{1.46}$$

$$\phi + \psi = \psi + \phi \qquad \text{Commutative law} \tag{1.47}$$

$$\phi + (\psi + \xi) = (\phi + \psi) + \xi \qquad \text{Associative law} \tag{1.48}$$

Multiplication by a scalar ξ

$$\xi\mathbf{A} = \mathbf{B} \quad \text{is a vector} \tag{1.49}$$

$$\xi\phi = \psi \quad \text{is a scalar} \tag{1.50}$$

Equation 1.49 can be proved as follows:

$$\begin{aligned} B_i' &= \sum_j \lambda_{ij} B_j = \sum_j \lambda_{ij} \xi A_j \\ &= \xi \sum_j \lambda_{ij} A_j = \xi A_i' \end{aligned} \tag{1.51}$$

and $\xi\mathbf{A}$ transforms as a vector. Similarly, $\xi\phi$ transforms as a scalar

1.10 Scalar Product of Two Vectors

The multiplication of two vectors **A** and **B** to form the scalar product is defined to be

$$\boxed{\mathbf{A}\cdot\mathbf{B} = \sum_i A_i B_i} \tag{1.52}$$

where the dot between **A** and **B** denotes scalar multiplication; this operation is sometimes called the **dot product**.

The vector **A** has components A_1, A_2, A_3, and the magnitude (or length) of **A** is given by

$$|\mathbf{A}| = +\sqrt{A_1^2 + A_2^2 + A_3^2} \equiv A \tag{1.53}$$

where the magnitude is indicated by $|\mathbf{A}|$ or, if there is no possibility of confusion, simply by A. Dividing both sides of Equation 1.52 by AB, we have

$$\frac{\mathbf{A}\cdot\mathbf{B}}{AB}=\sum_i \frac{A_i}{A}\frac{B_i}{B} \tag{1.54}$$

A_1/A is the cosine of the angle α between the vector **A** and the x_1-axis (see Figure 1-12).

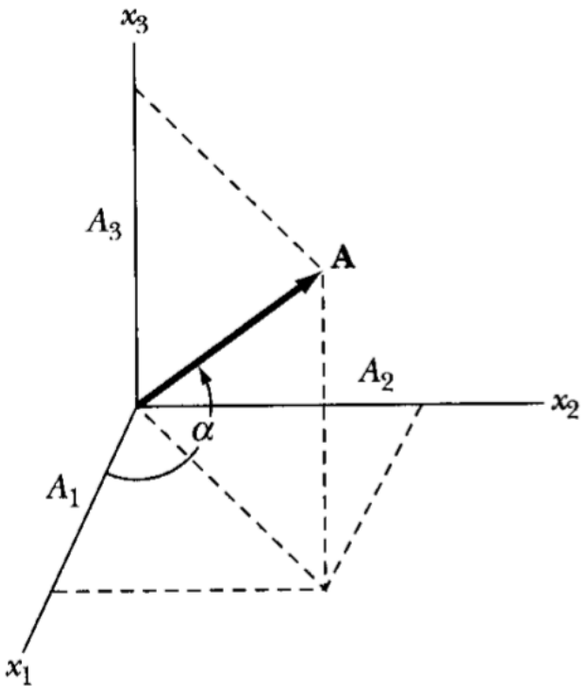


FIGURE 1-12 A vector **A** is shown in coordinate system x_1, x_2, x_3 with its vector components A_1, A_2, and A_3. The vector **A** is oriented at an angle α with the x_1-axis.

In general, A_i/A and B_i/B are the direction cosines Λ_i^A and Λ_i^B of the vectors **A** and **B**:

$$\frac{\mathbf{A}\cdot\mathbf{B}}{AB}=\sum_i \Lambda_i^A\Lambda_i^B \tag{1.55}$$

The sum $\sum_i \Lambda_i^A\Lambda_i^B$ is just the cosine of the angle between A and B (see Equation 1.11):

$$\cos(\mathbf{A},\mathbf{B})=\sum_i \Lambda_i^A\Lambda_i^B$$

or

$$\boxed{\mathbf{A}\cdot\mathbf{B}=AB\cos(\mathbf{A},\mathbf{B})} \tag{1.56}$$

That the product **A**·**B** is indeed a scalar may be shown as follows. **A** and **B** transform as vectors:

$$A'_i = \sum_j \lambda_{ij} A_j, \qquad B'_i = \sum_k \lambda_{ik} B_k \tag{1.57}$$

Therefore the product **A'**·**B'** becomes

$$\mathbf{A}' \cdot \mathbf{B}' = \sum_i A'_i B'_i$$

$$= \sum_i \left(\sum_j \lambda_{ij} A_j \right) \left(\sum_k \lambda_{ik} B_k \right)$$

Rearranging the summations, we can write

$$\mathbf{A}' \cdot \mathbf{B}' = \sum_{j,k} \left(\sum_i \lambda_{ij} \lambda_{ik} \right) A_j B_k$$

But according to the orthogonality condition, the term in parentheses is just δ_{jk} . Thus,

$$\mathbf{A}' \cdot \mathbf{B}' = \sum_j \left(\sum_k \delta_{jk} A_j B_k \right)$$

$$= \sum_j A_j B_j$$

$$= \mathbf{A} \cdot \mathbf{B} \tag{1.58}$$

Because the value of the product is unaltered by the coordinate transformation, the product must be a scalar.

Notice that the distance from the origin to the point (x_1, x_2, x_3) defined by the vector **A**, called the **position vector**, is given by

$$|\mathbf{A}| = \sqrt{\mathbf{A}\cdot\mathbf{A}} = \sqrt{x_1^2 + x_2^2 + x_3^2} = \sqrt{\sum_i x_i^2}$$

Similarly the distance from the point (x_1, x_2, x_3) to another point $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ defined by the vector **B** is

$$\sqrt{\sum_i (x_i - \bar{x}_i)^2} = \sqrt{(\mathbf{A}-\mathbf{B})\cdot(\mathbf{A}-\mathbf{B})} = |\mathbf{A}-\mathbf{B}|$$

That is, we can define the vector connecting any point with any other point as the difference of the position vectors that define the individual -points, as in Figure 1-13.

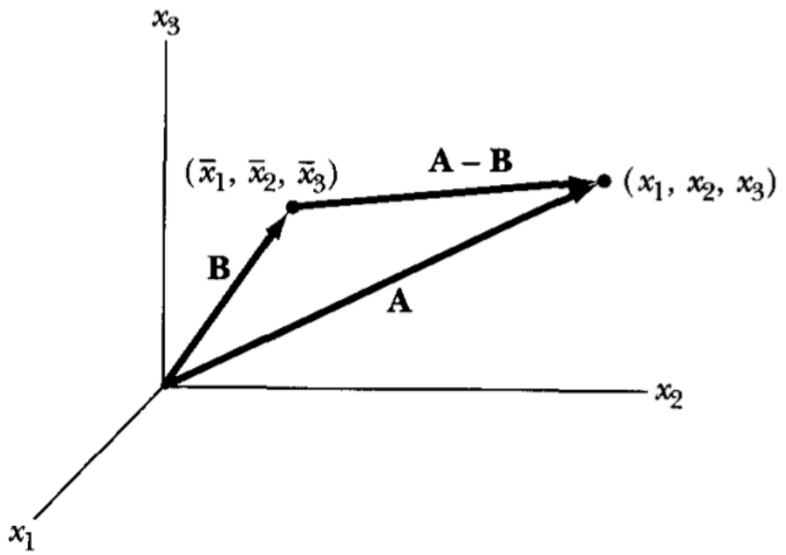


FIGURE 1-13 The vector **A** is the position vector of point (x_1, x_2, x_3), and vector **B** is the position vector of point $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$. The vector **A** − **B** is the position vector from $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ to (x_1, x_2, x_3).

The distance between the points is then the magnitude of the difference vector. And because this magnitude is the square root of a scalar product, it is invariant to a coordinate transformation. This is an important fact and can be summarized by the statement that *orthogonal transformations are distance-preserving transformations*. Also, the angle between two vectors is preserved under an orthogonal transformation. These two results are essential if we are to successfully apply transformation theory to physical situations.

The scalar product obeys the commutative and distributive laws:

$$\mathbf{A}\cdot\mathbf{B} = \sum_i A_i B_i = \sum_i B_i A_i = \mathbf{B}\cdot\mathbf{A} \tag{1.59}$$

$$\begin{aligned}\mathbf{A}\cdot(\mathbf{B}+\mathbf{C}) &= \sum_i A_i (B+C)_i = \sum_i A_i (B_i + C_i) \\ &= \sum_i (A_i B_i + A_i C_i) = (\mathbf{A}\cdot\mathbf{B}) + (\mathbf{A}\cdot\mathbf{C})\end{aligned} \tag{1.60}$$

1.11 Unit Vectors

Sometimes we want to describe a vector in terms of the components along the three coordinate axes together with a convenient specification of these axes. For this purpose, we introduce **unit vectors**, which are vectors having a length equal to the unit of length used along the particular coordinate axes. For example, the unit vector along the radial direction described by the vector $\mathbf{R}$ is $\mathbf{e}_R = \mathbf{R}/(|\mathbf{R}|)$.

There are several variants of the symbols for unit vectors; examples of the most common sets are $(\mathbf{i}, \mathbf{j}, \mathbf{k})$, $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi)$, and $(\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$. The following ways of expressing the vector $\mathbf{A}$ are equivalent:

$$\left.\begin{aligned}&\mathbf{A} = (A_1, A_2, A_3) \quad \text{or} \quad \mathbf{A} = \mathbf{e}_1 A_1 + \mathbf{e}_2 A_2 + \mathbf{e}_3 A_3 = \sum_i \mathbf{e}_i A_i \\ &\text{or} \quad \mathbf{A} = A_1\mathbf{i} + A_2\mathbf{j} + A_3\mathbf{k}\end{aligned}\right\} \tag{1.61}$$

Although the unit vectors $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ and $(\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$ are somewhat easier to use, we tend to use unit vectors such as $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, because of the ease of summation notation. We obtain the components of the vector $\mathbf{A}$ by projection onto the axes:

$$A_i = \mathbf{e}_i \cdot \mathbf{A} \tag{1.62}$$

We have seen (Equation 1.56) that the scalar product of two vectors has a magnitude equal to the product of the individual magnitudes multiplied by the cosine of the angle between the vectors:

$$\mathbf{A} \cdot \mathbf{B} = AB \cos(\mathbf{A}, \mathbf{B}) \tag{1.63}$$

If any two unit vectors are orthogonal, we have

$$\boxed{\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}} \tag{1.64}$$

EXAMPLE 1.5

Two position vectors are expressed in Cartesian coordinates as $\mathbf{A} = \mathbf{i} + 2\mathbf{j} - 2\mathbf{k}$ and $\mathbf{B} = 4\mathbf{i} + 2\mathbf{j} - 3\mathbf{k}$. Find the magnitude of the vector from point A to point B, the angle 6 between **A** and **B**, and the component of **B** in the direction of **A**.

Solution. The vector from point A to point B is **B** - **A** (see Figure 1.13)

$$\mathbf{B} - \mathbf{A} = 4\mathbf{i} + 2\mathbf{j} - 3\mathbf{k} - (\mathbf{i} + 2\mathbf{j} - 2\mathbf{k}) = 3\mathbf{i} - \mathbf{k}$$

$$|\mathbf{B} - \mathbf{A}| = \sqrt{9 + 1} = \sqrt{10}$$

From Equation 1.56

$$\cos\theta = \frac{\mathbf{A} \cdot \mathbf{B}}{AB} = \frac{(\mathbf{i} + 2\mathbf{j} - 2\mathbf{k}) \cdot (4\mathbf{i} + 2\mathbf{j} - 3\mathbf{k})}{\sqrt{9}\sqrt{29}}$$

$$\cos\theta = \frac{4 + 4 + 6}{3(\sqrt{29})} = 0.867$$

$$\theta = 30°$$

The component of **B** in the direction of **A** is $B\cos\theta$ and, from Equation 1.56,

$$B\cos\theta = \frac{\mathbf{A} \cdot \mathbf{B}}{A} = \frac{14}{3} = 4.67$$

1.12 Vector Product of Two Vectors

We next consider another method of combining two vectors—the **vector product** (sometimes called the **cross product**). In most respects, the vector product of two vectors behaves like a vector, and we shall treat it as such. The vector product of **A** and **B** is denoted by a bold cross x,

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} \tag{1.65}$$

where **C** is the vector resulting from this operation. The components of **C** are defined by the relation

$$\boxed{C_i \equiv \sum_{j,k} \varepsilon_{ijk} A_j B_k} \tag{1.66}$$

where the symbol ε_{ijk} is the **permutation symbol** or (**Levi-Civita density**) and has the following properties:

$$\left.\begin{aligned} 0, &\quad \text{if any index is equal to any other index} \\ \varepsilon_{ijk} = +1, &\quad \text{if } i, j, k \text{ form an } \textit{even} \text{ permutation of } 1, 2, 3 \\ -1, &\quad \text{if } i, j, k \text{ form an } \textit{odd} \text{ permutation of } 1, 2, 3 \end{aligned}\right\} \tag{1.67}$$

An even permutation has an even number of exchanges of position of two symbols. Cyclic permutations (for example, $123 \to 231 \to 312$) are always even. Thus

$$\varepsilon_{122} = \varepsilon_{313} = \varepsilon_{211} = 0, \quad \text{etc.}$$
$$\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = +1$$
$$\varepsilon_{132} = \varepsilon_{213} = \varepsilon_{321} = -1$$

Using the preceding notation, the components of **C** can be explicitly evaluated. For the first subscript equal to 1, the only nonvanishing ε_{ijk} are ε_{123} and ε_{132} - that is, for j, k = 2, 3 in either order. Therefore

$$C_1 = \sum_{j,k} \varepsilon_{1jk} A_j B_k = \varepsilon_{123} A_2 B_3 + \varepsilon_{132} A_3 B_2$$
$$= A_2 B_3 - A_3 B_2 \tag{1.68a}$$

Similarly,

$$C_2 = A_3 B_1 - A_1 B_3 \tag{1.68b}$$
$$C_3 = A_1 B_2 - A_2 B_1 \tag{1.68c}$$

Consider now the expansion of the quantity $[AB \sin(\mathbf{A}, \mathbf{B})]^2 = (AB \sin\theta)^2$:

$$A^2 B^2 \sin^2\theta = A^2 B^2 - A^2 B^2 \cos^2\theta$$
$$= \left(\sum_i A_i^2\right)\left(\sum_i B_i^2\right) - \left(\sum_i A_i B_i\right)^2$$
$$= (A_2 B_3 - A_3 B_2)^2 + (A_3 B_1 - A_1 B_3)^2 + (A_1 B_2 - A_2 B_1)^2 \tag{1.69}$$

where the last equality requires some algebra. Identifying the components of **C** in the last expression, we can write

$$(AB \sin\theta)^2 = C_1^2 + C_2^2 + C_3^2 = |\mathbf{C}^2| = C^2 \tag{1.70}$$

If we take the positive square root of both sides of this equation,

$$C = AB \sin\theta \tag{1.71}$$

This equation states that if **C** = **A** x **B**, the magnitude of **C** is equal to the product of the magnitudes of **A** and **B** multiplied by the sine of the angle between them. Geometrically, AB $\sin\theta$ is the area of the parallelogram defined by the vectors **A** and **B** and the angle between them, as in Figure 1-14.

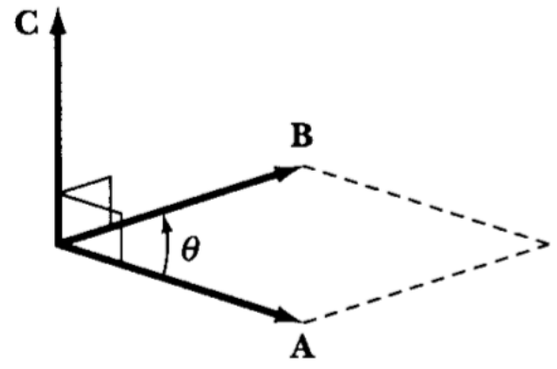


FIGURE 1-14 The magnitude of the vector **C** determined by $\mathbf{C} = \mathbf{A} \times \mathbf{B}$ has a magnitude given by the area of the parallelogram $AB \sin\theta$, where θ is the angle between the vectors **A** and **B**.

EXAMPLE 1.6

Show by using Equations 1.52 and 1.66 that

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{D}) = \mathbf{D} \cdot (\mathbf{A} \times \mathbf{B}) \tag{1.72}$$

Solution. Using Equation 1.66, we have

$$(\mathbf{B} \times \mathbf{D})_i = \sum_{j,k} \varepsilon_{ijk} B_j D_k$$

Using Equation 1.52, we have

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{D}) = \sum_{i,j,k} \varepsilon_{ijk} A_i B_j D_k \tag{1.73}$$

Similarly, for the right-hand side of Equation 1.72, we have

$$\mathbf{D} \cdot (\mathbf{A} \times \mathbf{B}) = \sum_{i,j,k} \varepsilon_{ijk} D_i A_j B_k$$

From the definition (Equation 1.67) of ε_{ijk} , we can interchange two adjacent indices of ε_{ijk} , which changes the sign.

$$\mathbf{D}\cdot(\mathbf{A}\times\mathbf{B}) = \sum_{i,j,k} - \varepsilon_{jik} D_i A_j B_k$$
$$= \sum_{i,j,k} \varepsilon_{jki} A_j B_k D_i \tag{1.74}$$

Because the indices i, j, k are dummy and can be renamed, the right-hand sides of Equations 1.73 and 1.74 are identical, and Equation 1.72 is proved. Equation 1.72 can also be written as $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{D}) = (\mathbf{A}\times\mathbf{B})\cdot\mathbf{D}$, indicating that the scalar and vector products can be interchanged as long as the vectors stay in the order **A**, **B**, **D**. Notice that, if we let **B** = **A**, we have

$$\mathbf{A}\cdot(\mathbf{A}\times\mathbf{D}) = \mathbf{D}\cdot(\mathbf{A}\times\mathbf{A}) = 0$$

showing that $\mathbf{A}\times\mathbf{D}$ must be perpendicular to **A**.

$\mathbf{A}\times\mathbf{B}$ (i.e., **C**) is perpendicular to the plane defined by A and B because $\mathbf{A}\cdot(\mathbf{A}\times\mathbf{B}) = 0$ and $\mathbf{B}\cdot(\mathbf{A}\times\mathbf{B}) = 0$. Because a plane area can be represented by a vector normal to the plane and of magnitude equal to the area, **C** is evidently such a vector. The positive direction of **C** is chosen to be the direction of advance of a right-hand screw when rotated from **A** to **B**.

The definition of the vector product is now complete; components, magnitude, and geometrical interpretation have been given. We may therefore reasonably expect that **C** is indeed a vector The ultimate test, however, is to examine the transformation properties of **C**, and **C** does, in fact, transform as a vector under a proper rotation.

We should note the following properties of the vector product that result from the definitions:

$$\text{(a)} \quad \mathbf{A}\times\mathbf{B} = -\mathbf{B}\times\mathbf{A} \tag{1.75}$$

but, in general,

$$\text{(b)} \quad \mathbf{A}\times(\mathbf{B}\times\mathbf{C}) \neq (\mathbf{A}\times\mathbf{B})\times\mathbf{C} \tag{1.76}$$

Another important result is

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C} \tag{1.77}$$

EXAMPLE 1.7

Find the product of (**A** x **B**)•(**C** x **D**).

Solution.

$$(\mathbf{A} \times \mathbf{B})_i = \sum_{j,k} \varepsilon_{ijk} A_j B_k$$

$$(\mathbf{C} \times \mathbf{D})_i = \sum_{l,m} \varepsilon_{ilm} C_l D_m$$

The scalar product is then computed according to Equation 1.52:

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = \sum_i \left(\sum_{j,k} \varepsilon_{ijk} A_j B_k \right) \left(\sum_{l,m} \varepsilon_{ilm} C_l D_m \right)$$

Rearranging the summations, we have

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = \sum_{\substack{l,m \\ j,k}} \left(\sum_i \varepsilon_{jki} \varepsilon_{lmi} \right) A_j B_k C_l D_m$$

where the indices of the ε_{ijk} have been permuted (twice each so that no sign change occurs) to place in the third position the index over which the sum is carried out. We can now use an important property of the ε_{ijk}:

$$\boxed{\sum_k \varepsilon_{ijk} \varepsilon_{lmk} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}} \tag{1.78}$$

We therefore have

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = \sum_{\substack{j,k \\ l,m}} (\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}) A_j B_k C_l D_m$$

Carrying out the summations over j and k, the Kronecker deltas reduce the expression to

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = \sum_{l,m} (A_l B_m C_l D_m - A_m B_l C_l D_m)$$

This equation can be rearranged to obtain

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = \left(\sum_l A_l C_l\right)\left(\sum_m B_m D_m\right) - \left(\sum_l B_l C_l\right)\left(\sum_m A_m D_m\right)$$

Because each term in parentheses on the right-hand side is just a scalar product, we have, finally,

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}) - (\mathbf{B} \cdot \mathbf{C})(\mathbf{A} \cdot \mathbf{D})$$

The orthogonality of the unit vectors $\mathbf{e}_i$ requires the vector product to be

$$\mathbf{e}_i \times \mathbf{e}_j = \mathbf{e}_k \quad i, j, k \text{ in cyclic order} \tag{1.79a}$$

We can now use the permutation symbol to express this result as

$$\boxed{\mathbf{e}_i \times \mathbf{e}_j = \sum_k \mathbf{e}_k \varepsilon_{ijk}} \tag{1.79b}$$

The vector product $\mathbf{C} = \mathbf{A} \times \mathbf{B}$, for example, can now be expressed as

$$\mathbf{C} = \sum_{i,j,k} \varepsilon_{ijk} \mathbf{e}_i A_j B_k \tag{1.80a}$$

By direct expansion and comparison with Equation 1.80a, we can verify a determinantal expression for the vector product:

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix} \tag{1.80b}$$

We state the following identities without proof:

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}) \equiv \mathbf{ABC} \tag{1.81}$$

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C} \tag{1.82}$$

$$\left.\begin{aligned} (\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) &= \mathbf{A} \cdot [\mathbf{B} \times (\mathbf{C} \times \mathbf{D})] \\ &= \mathbf{A} \cdot [(\mathbf{B} \cdot \mathbf{D})\mathbf{C} - (\mathbf{B} \cdot \mathbf{C})\mathbf{D}] \\ &= (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}) - (\mathbf{A} \cdot \mathbf{D})(\mathbf{B} \cdot \mathbf{C}) \end{aligned}\right\} \tag{1.83}$$

$$\left.\begin{aligned} (\mathbf{A} \times \mathbf{B}) \times (\mathbf{C} \times \mathbf{D}) &= [(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{D}]\mathbf{C} - [(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C}]\mathbf{D} \\ &= (\mathbf{ABD})\mathbf{C} - (\mathbf{ABC})\mathbf{D} = (\mathbf{ACD})\mathbf{B} - (\mathbf{BCD})\mathbf{A} \end{aligned}\right\} \tag{1.84}$$

1.13 Differentiation of a Vector with Respect to a Scalar

If a scalar function $\phi = \phi(s)$ is differentiated with respect to the scalar variable s, then, because neither part of the derivative can change under a coordinate transformation, the derivative itself cannot change and must therefore be a scalar; that is, in the x_i and x'_i-coordinate systems, $\phi = \phi'$ and s = s', so $d\phi = d\phi'$ and ds = ds'. Hence

$$\frac{d\phi}{ds} = \frac{d\phi'}{ds'} = \left(\frac{d\phi}{ds}\right)'$$

Similarly, we can formally define the differentiation of a vector **A** with respect to a scalar s. The components of **A** transform according to

$$A'_i = \sum_j \lambda_{ij} A_j \tag{1.85}$$

Therefore, on differentiation, we obtain (because the λ_{ij} , are independent of s')

$$\frac{dA'_i}{ds'} = \frac{d}{ds'} \sum_j \lambda_{ij} A_j = \sum_j \lambda_{ij} \frac{dA_j}{ds'}$$

Because s and s' are identical, we have

$$\frac{dA'_i}{ds'} = \left(\frac{dA_i}{ds}\right)' = \sum_j \lambda_{ij} \left(\frac{dA_j}{ds}\right)$$

Thus the quantities dA_j/ds transform as do the components of a vector and hence are the components of a vector, which we can write as d**A**/ds.

We can give a geometrical interpretation to the vector d**A**/ds as follows.

First, for d**A**/ds to exist, **A** must be a continuous function of the variable s: **A** = **A**(s). Suppose this function is represented by the continuous curve Γ in Figure 1-15; at the point P, the variable has the value s, and at Q it has the value $s + \Delta s$.

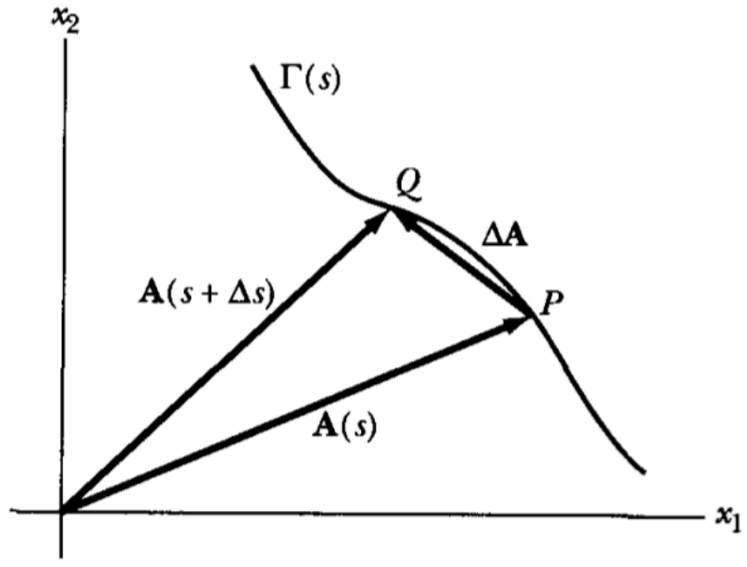


FIGURE 1-15 The vector $\mathbf{A}(s)$ traces out the function $\Gamma(s)$ as the variable s changes.

The derivative of **A** with respect to s is then given in standard fashion by

$$\frac{d\mathbf{A}}{ds} = \lim_{\Delta s \to 0} \frac{\Delta \mathbf{A}}{\Delta s} = \lim_{\Delta s \to 0} \frac{\mathbf{A}(s + \Delta s) - \mathbf{A}(s)}{\Delta s} \tag{1.86a}$$

The derivatives of vector sums and products obey the rules of ordinary vector calculus. For example,

$$\frac{d}{ds}(\mathbf{A} + \mathbf{B}) = \frac{d\mathbf{A}}{ds} + \frac{d\mathbf{B}}{ds} \tag{1.86b}$$

$$\frac{d}{ds}(\mathbf{A} \cdot \mathbf{B}) = \mathbf{A} \cdot \frac{d\mathbf{B}}{ds} + \frac{d\mathbf{A}}{ds} \cdot \mathbf{B} \tag{1.86c}$$

$$\frac{d}{ds}(\mathbf{A} \times \mathbf{B}) = \mathbf{A} \times \frac{d\mathbf{B}}{ds} + \frac{d\mathbf{A}}{ds} \times \mathbf{B} \tag{1.86d}$$

$$\frac{d}{ds}(\phi\mathbf{A}) = \phi \frac{d\mathbf{A}}{ds} + \frac{d\phi}{ds}\mathbf{A} \tag{1.86e}$$

and similarly for total differentials and for partial derivatives.

1.14 Examples of Derivatives - Velocity and Acceleration

Of particular importance in the development of the dynamics of point particles (and of systems of particles) is the representation of the motion of these particles by vectors. For such an approach, we require vectors to represent the position, velocity, and acceleration of a given particle. It is customary to specify the position of a particle with respect to a certain reference frame by a vector **r**, which is in general a function of time: **r** = **r**(t). The velocity vector **v** and the acceleration vector **a** are defined according to

$$\mathbf{v} \equiv \frac{d\mathbf{r}}{dt} = \dot{\mathbf{r}} \tag{1.87}$$

$$\mathbf{a} \equiv \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2} = \ddot{\mathbf{r}} \tag{1.88}$$

where a single dot above a symbol denotes the first time derivative, and two dots denote the second time derivative. In rectangular coordinates, the expressions for **r**, **v**, and **a** are

$$\left.\begin{aligned}
\mathbf{r} &= x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 = \sum_i x_i\mathbf{e}_i && \text{Position} \\
\mathbf{v} &= \dot{\mathbf{r}} = \sum_i \dot{x}_i\mathbf{e}_i = \sum_i \frac{dx_i}{dt}\mathbf{e}_i && \text{Velocity} \\
\mathbf{a} &= \dot{\mathbf{v}} = \ddot{\mathbf{r}} = \sum_i \ddot{x}_i\,\mathbf{e}_i = \sum_i \frac{d^2x_i}{dt^2}\mathbf{e}_i && \text{Acceleration}
\end{aligned}\right\} \tag{1.89}$$

Calculating these quantities in rectangular coordinates is straightforward because the unit vectors $\mathbf{e}_i$ are constant in time. In nonrectangular coordinate systems, however, the unit vectors at the position of the particle as it moves in space are not necessarily constant in time, and the components of the time derivatives of **r** are no longer simple relations, as in Equation 1.89. We do not discuss general curvilinear coordinate systems here, but plane polar coordinates, spherical coordinates, and cylindrical coordinates are of sufficient importance to warrant a discussion of velocity and acceleration in these coordinate systems.

Refer to the figures in Appendix F for the geometry of these coordinate systems.

To express v and a in plane polar coordinates, consider the situation in Figure 1-16. A point moves along the curve s(t) and in the time interval $t_2 - t_1 = dt$ moves from $P^{(1)}$ to $P^{(2)}$. The unit vectors, $\mathbf{e}_r$ and $\mathbf{e}_\theta$, which are orthogonal, change from $\mathbf{e}_r^{(1)}$ to $\mathbf{e}_\theta^{(1)}$ and from $\mathbf{e}_r^{(2)}$ to $\mathbf{e}_\theta^{(2)}$. The change in $\mathbf{e}_r$ is

$$\mathbf{e}_r^{(2)} - \mathbf{e}_r^{(1)} = d\mathbf{e}_r \tag{1.90}$$

which is a vector normal to $\mathbf{e}_r$ (and, therefore, in the direction of $\mathbf{e}_\theta$). Similarly, the change in $\mathbf{e}_\theta$ is

$$\mathbf{e}_\theta^{(2)} - \mathbf{e}_\theta^{(1)} = d\mathbf{e}_\theta \tag{1.91}$$

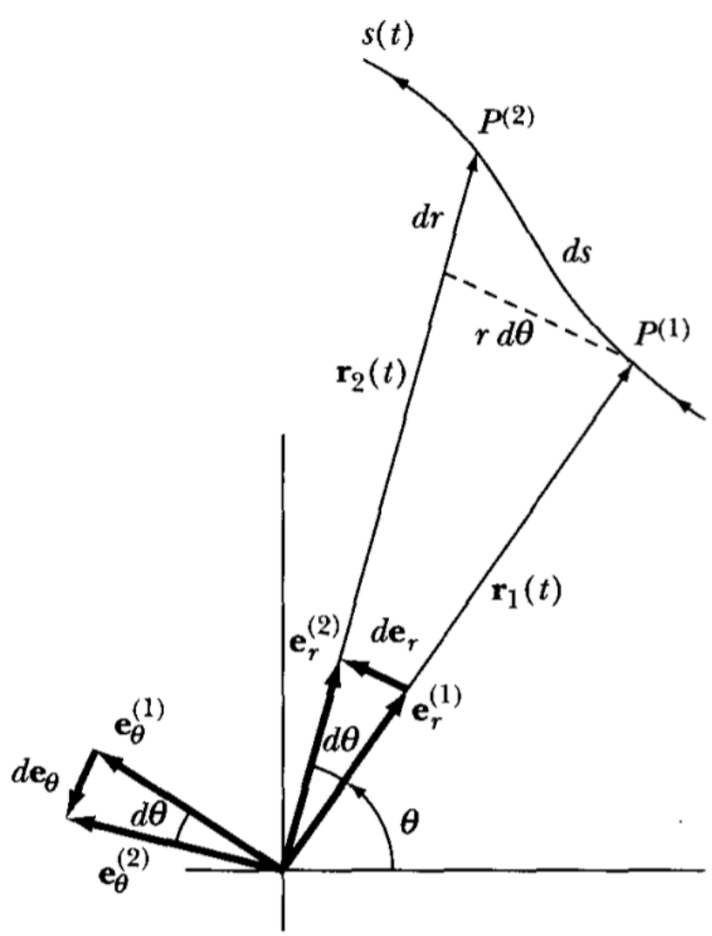


FIGURE 1-16 An object traces out the curve $s(t)$ over time. The unit vectors $\mathbf{e}_r$ and $\mathbf{e}_\theta$ and their differentials are shown for two position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$.

which is a vector normal to $\mathbf{e}_\theta$. We can then write

$$d\mathbf{e}_r = d\theta\mathbf{e}_\theta \tag{1.92}$$

and

$$d\mathbf{e}_\theta = -d\theta\mathbf{e}_r \tag{1.93}$$

where the minus sign enters the second relation because $d\mathbf{e}_\theta$ is directed opposite to $\mathbf{e}_r$ (see Figure 1-16).

Equations 1.92 and 1.93 are perhaps easier to see by referring to Figure 1-16. In this case, $d\mathbf{e}_r$ subtends an angle $d\theta$ with unit sides, so it has a magnitude of $d\theta$. It also points in the direction of $\mathbf{e}_\theta$, so we have $d\mathbf{e}_r = d\theta\mathbf{e}_\theta$. Similarly, $d\mathbf{e}_\theta$ subtends an angle $d\theta$ with unit sides, so it also has a magnitude of $d\theta$, but from Figure 1-16 we see that $d\mathbf{e}_\theta$ points in the direction of - $\mathbf{e}_r$ so we have $d\mathbf{e}_\theta = -d\theta\mathbf{e}_r$.

Dividing each side of Equations 1.92 and 1.93 by dt, we have

$$\boxed{\dot{\mathbf{e}}_r = \dot{\theta}\mathbf{e}_\theta} \tag{1.94}$$

$$\boxed{\dot{\mathbf{e}}_\theta = -\dot{\theta}\mathbf{e}_r} \tag{1.95}$$

If we express **v** as

$$\begin{aligned}\mathbf{v} = \frac{d\mathbf{r}}{dt} &= \frac{d}{dt}(r\mathbf{e}_r) \\ &= \dot{r}\mathbf{e}_r + r\dot{\mathbf{e}}_r\end{aligned} \tag{1.96}$$

we have immediately, using Equation 1.94,

$$\boxed{\mathbf{v} = \dot{\mathbf{r}} = \dot{r}\mathbf{e}_r + r\dot{\theta}\mathbf{e}_\theta} \tag{1.97}$$

so that the velocity is resolved into a *radial* component $\dot{r}$ and an *angular* (or *transverse*) component $r\dot{\theta}$.

A second differentiation yields the acceleration:

$$\begin{aligned}\mathbf{a} &= \frac{d}{dt}(\dot{r}\mathbf{e}_r + r\dot{\theta}\,\mathbf{e}_\theta) \\ &= \ddot{r}\mathbf{e}_r + \dot{r}\dot{\mathbf{e}}_r + \dot{r}\dot{\theta}\mathbf{e}_\theta + r\ddot{\theta}\,\mathbf{e}_\theta + r\dot{\theta}\dot{\mathbf{e}}_\theta \\ &= (\ddot{r} - r\dot{\theta}^2)\mathbf{e}_r + (r\ddot{\theta} + 2\dot{r}\dot{\theta})\mathbf{e}_\theta\end{aligned} \tag{1.98}$$

so that the acceleration s resolved into a radial component $(\ddot{r} - r\dot{\theta}^2)$ and an angular (or transverse) component $(r\ddot{\theta} + 2\dot{r}\dot{\theta})$.

The expressions for $d\mathbf{s}$, ds^2, v^2, and $\mathbf{v}$ in the three most important coordinate systems (see also Appendix F) are

Rectangular coordinates (x, y, z)

$$\left.\begin{aligned} d\mathbf{s} &= dx_1\mathbf{e}_1 + dx_2\mathbf{e}_2 + dx_3\mathbf{e}_3 \\ ds^2 &= dx_1^2 + dx_2^2 + dx_3^2 \\ v^2 &= \dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2 \\ \mathbf{v} &= \dot{x}_1\mathbf{e}_1 + \dot{x}_2\mathbf{e}_2 + \dot{x}_3\mathbf{e}_3 \end{aligned}\right\} \tag{1.99}$$

Spherical coordinates (r, θ, ϕ)

$$\left.\begin{aligned} d\mathbf{s} &= dr\mathbf{e}_r + rd\theta\mathbf{e}_\theta + r\sin\theta\, d\phi\mathbf{e}_\phi \\ ds^2 &= dr^2 + r^2d\theta^2 + r^2\sin^2\theta\, d\phi^2 \\ v^2 &= \dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2 \\ \mathbf{v} &= \dot{r}\mathbf{e}_r + r\dot{\theta}\mathbf{e}_\theta + r\sin\theta\,\dot{\phi}\mathbf{e}_\phi \end{aligned}\right\} \tag{1.100}$$

Cylindrical coordinates (r, ϕ, z)

$$\left.\begin{aligned} d\mathbf{s} &= dr\mathbf{e}_r + rd\phi\mathbf{e}_\phi + dz\mathbf{e}_z \\ ds^2 &= dr^2 + r^2d\phi^2 + dz^2 \\ v^2 &= \dot{r}^2 + r^2\dot{\phi}^2 + \dot{z}^2 \\ \mathbf{v} &= \dot{r}\mathbf{e}_r + r\dot{\phi}\mathbf{e}_\phi + \dot{z}\mathbf{e}_z \end{aligned}\right\} \tag{1.101}$$

The expressions for plane polar coordinates result from Equation 1.100 by $d\phi = 0$.)

EXAMPLE: 1.8

Find the components of the acceleration vector **a** in cylindrical coordinates.

Solution. The velocity components in cylindrical coordinates were given in Equation 1.101. The acceleration is determined by taking the time derivative of **v**.

$$\mathbf{a} = \frac{d}{dt}\,\mathbf{v} = \frac{d}{dt}(\dot{r}\mathbf{e}_r + r\dot{\phi}\mathbf{e}_\phi + \dot{z}\mathbf{e}_z)$$

$$= \ddot{r}\mathbf{e}_r + \dot{r}\dot{\mathbf{e}}_r + \dot{r}\dot{\phi}\mathbf{e}_\phi + r\ddot{\phi}\mathbf{e}_\phi + r\dot{\phi}\dot{\mathbf{e}}_\phi + \ddot{z}\mathbf{e}_z + \dot{z}\dot{\mathbf{e}}_z$$

We need to find the time derivative of the unit vectors $\mathbf{e}_r, \mathbf{e}_\phi$, and $\mathbf{e}_z$. The cylindrical coordinate system is shown in Figure 1-17, and in terms .of the (x, y, z) components, the unit vectors $\mathbf{e}_r, \mathbf{e}_\phi$, and $\mathbf{e}_z$ are

$$\mathbf{e}_r = (\cos\phi, \sin\phi, 0)$$
$$\mathbf{e}_\phi = (-\sin\phi, \cos\phi, 0)$$
$$\mathbf{e}_z = (0, 0, 1)$$

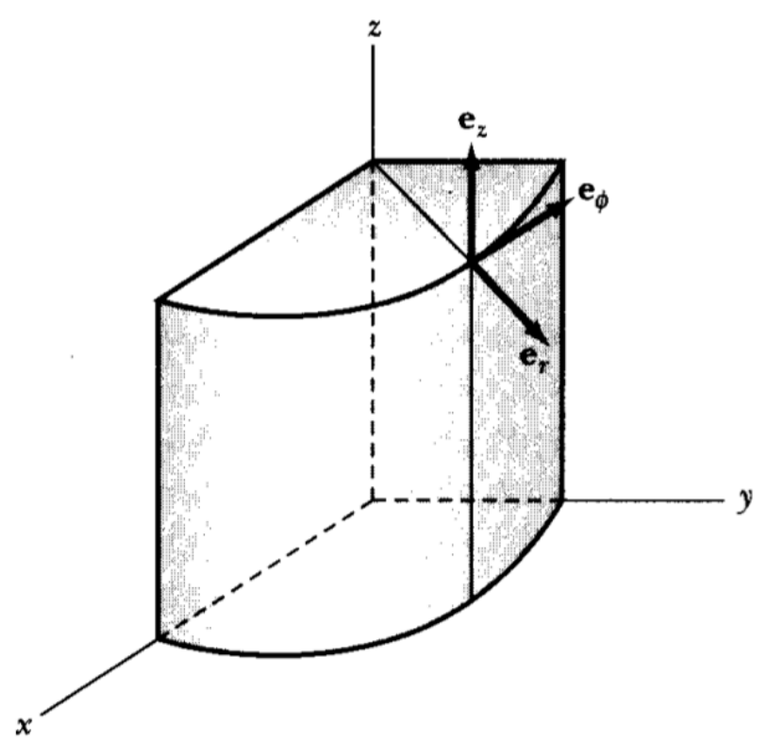


FIGURE 1-17 The cylindrical coordinate system (r, ϕ, z) are shown with respect to the Cartesian system (x, y, z).

The time derivatives of the unit vectors are found by taking the derivatives of the components.

$$\dot{\mathbf{e}}_r = (-\dot{\phi}\sin\phi, \dot{\phi}\cos\phi, 0) = -\dot{\phi}\mathbf{e}_\phi$$
$$\dot{\mathbf{e}}_\phi = (-\dot{\phi}\cos\phi, -\dot{\phi}\sin\phi, 0) = -\dot{\phi}\mathbf{e}_r$$
$$\dot{\mathbf{e}}_z = 0$$

We substitute the unit vector time derivatives into the above expression for **a**.

$$\begin{aligned}\mathbf{a} &= \ddot{r}\mathbf{e}_r + \dot{r}\dot{\phi}\mathbf{e}_\phi + \dot{r}\dot{\phi}\mathbf{e}_\phi + r\ddot{\phi}\mathbf{e}_\phi - r\dot{\phi}^2\mathbf{e}_r + \ddot{z}\mathbf{e}_z \\ &= (\ddot{r} - r\dot{\phi}^2)\mathbf{e}_r + (r\ddot{\phi} + 2\dot{r}\dot{\phi})\mathbf{e}_\phi + \ddot{z}\mathbf{e}_z\end{aligned}$$

1.15 Angular Velocity

A point or a particle moving arbitrarily in space may always be considered, at a given instant, to be moving in a plane, circular path about a certain axis; that is, the path a particle describes during an infinitesimal time interval δt may be represented as an infinitesimal arc of a circle. The line passing through the center of the circle and perpendicular to the instantaneous direction of motion is called the **instantaneous axis of rotation**. As the particle moves in the circular path, the rate of change of the angular position is called the **angular velocity**:

$$\omega = \frac{d\theta}{dt} = \dot{\theta} \tag{1.102}$$

Consider a particle that moves instantaneously in a circle of radius R about an axis perpendicular to the plane of motion, as in Figure 1-18. Let the position vector r of the particle be drawn from an origin located at an arbitrary point O on the axis of rotation. The time rate of change of the position vector is the linear velocity vector of the particle, r = v. For motion in a circle of radius R, the instantaneous *magnitude* of the linear velocity is given by

$$v = R\frac{d\theta}{dt} = R\omega \tag{1.103}$$

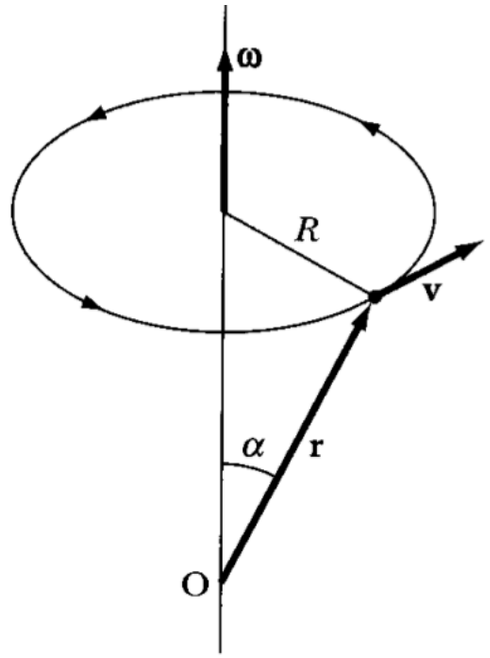


FIGURE 1-18 A particle moving ccw about an axis according to the right-hand rule has an angular velocity $\boldsymbol{\omega} = \mathbf{v} \times \mathbf{r}$ about that axis.

The *direction* of the linear velocity **v** is perpendicular to **r** and in the plane of the circle.

It would be very convenient if we could devise a vector representation of the angular velocity (say, to) so that all the quantities of interest in the motion of the particle could be described on a common basis. We can define a direction for the angular velocity in the following manner. If the particle moves instantaneously in a plane, the normal to that plane defines a precise direction in space—or, rather—two directions. We may choose as positive that direction corresponding to the direction of advance of a right-hand screw when turned in the same sense as the rotation of the particle (see Figure 1-18). We can also write the magnitude of the linear velocity by noting that $R = r \sin\alpha$. Thus

$$v = r\omega \sin\alpha \tag{1.104}$$

Having defined a direction and a magnitude for the angular velocity, we note that if we write

$$\boxed{\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}} \tag{1.105}$$

then both of these definitions are satisfied, and we have the desired vector representation of the angular velocity.

We should note at this point an important distinction between finite and infinitesimal rotations. An **infinitesimal rotation** can be represented by a vector (actually, an *axial* vector), but a **finite rotation** cannot. The impossibility of describing a finite rotation by a vector results from the fact that such rotations do not commute (see the example of Figure 1-9), and therefore, in general, different results will be obtained depending on the order in which the rotations are made. To illustrate this statement, consider the successive application of two finite rotations described by the rotation matrices $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$. Let us associate the vectors **A** and **B** in a one-to-one manner with these rotations. The vector sum is **C** = **A** + **B**, which is equivalent to the matrix $\boldsymbol{\lambda}_3 = \boldsymbol{\lambda}_2\boldsymbol{\lambda}_1$. But because vector addition is commutative, we also have **C** = **B** + **A**, with $\boldsymbol{\lambda}_4 = \boldsymbol{\lambda}_1\boldsymbol{\lambda}_2$. But we know that matrix operations are not commutative, so that in general $\boldsymbol{\lambda}_3 \neq \boldsymbol{\lambda}_4$. Hence, the vector C is not unique, and therefore we cannot associate a vector with a finite rotation.

Infinitesimal rotations do not suffer from this defect of noncommutation. We are therefore led to expect that an infinitesimal rotation can be represented by a vector. Although this expectation is, in fact, fulfilled, the ultimate test of the vector nature of a quantity is contained in its transformation properties. We give only a qualitative argument here.

Refer to Figure 1-19.

If the position vector of a point changes from $\mathbf{r}$ to $\mathbf{r} + \delta\mathbf{r}$, the geometrical situation is correctly represented if we write

$$\delta\mathbf{r} = \delta\boldsymbol{\theta} \times \mathbf{r} \tag{1.106}$$

where $\delta\boldsymbol{\theta}$ is a quantity whose magnitude is equal to the infinitesimal rotation angle and that has a direction along the instantaneous axis of rotation.

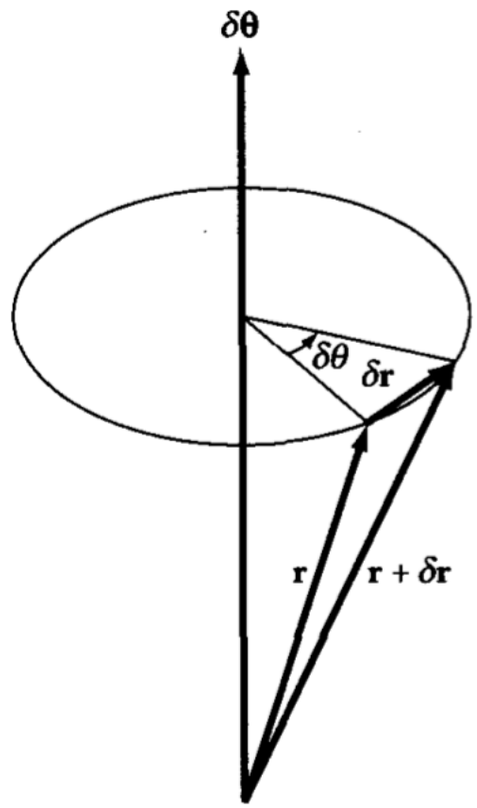


FIGURE 1-19 The position vector **r** changes to **r** + δ**r** by an infinitesimal rotation angle $\delta\theta$.

The mere fact that Equation 1.106 correctly describes the situation illustrated in Figure 1-19 is not sufficient to establish that $\delta\boldsymbol{\theta}$ is a vector. (We reiterate that the true test must be based on the transformation properties of $\delta\boldsymbol{\theta}$.) But if we show that two infinitesimal rotation "vectors"- $\delta\boldsymbol{\theta}_1$ and $\delta\boldsymbol{\theta}_2$ - actually *commute*, the sole objection to representing a finite rotation by a vector will have been removed.

Let us consider that a rotation $\delta\boldsymbol{\theta}_1$ takes $\mathbf{r}$ into $\mathbf{r} + \delta\mathbf{r}_1$ where $\delta\mathbf{r}_1 = \delta\boldsymbol{\theta}_1 \times \mathbf{r}$. If this is followed by a second rotation $\delta\boldsymbol{\theta}_2$ around a different axis, the initial position vector for this rotation is $\mathbf{r} + \delta\mathbf{r}_1$. Thus

$$\boldsymbol{\delta\mathbf{r}_2 = \delta\boldsymbol{\theta}_2 \times (\mathbf{r} + \delta\mathbf{r}_1)}$$

and the final position vector for $\delta\boldsymbol{\theta}_1$ followed by $\delta\boldsymbol{\theta}_2$ is

$$\mathbf{r} + \delta\mathbf{r}_{12} = \mathbf{r} + [\delta\boldsymbol{\theta}_1 \times \mathbf{r} + \delta\boldsymbol{\theta}_2 \times (\mathbf{r} + \delta\mathbf{r}_1)]$$

Neglecting second-order infinitesimals, then,

$$\delta\mathbf{r}_{12} = \delta\boldsymbol{\theta}_1 \times \mathbf{r} + \delta\boldsymbol{\theta}_2 \times \mathbf{r} \tag{1.107}$$

Similarly, if $\delta\boldsymbol{\theta}_2$ is followed by $\delta\boldsymbol{\theta}_1$, we have

$$\mathbf{r} + \delta\mathbf{r}_{21} = \mathbf{r} + [\delta\boldsymbol{\theta}_2 \times \mathbf{r} + \delta\boldsymbol{\theta}_1 \times (\mathbf{r} + \delta\mathbf{r}_2)]$$

or

$$\delta\mathbf{r}_{21} = \delta\boldsymbol{\theta}_2 \times \mathbf{r} + \delta\boldsymbol{\theta}_1 \times \mathbf{r} \tag{1.108}$$

Rotation vectors $\delta\mathbf{r}_{12}$ and $\delta\mathbf{r}_{21}$ are equal, so the rotation "vectors" $\delta\boldsymbol{\theta}_1$ and $\delta\boldsymbol{\theta}_2$ do commute. It therefore seems reasonable that $\delta\boldsymbol{\theta}$ in Equation 1.106 is indeed a vector

It is the fact that $\delta\boldsymbol{\theta}$ is a vector that allows angular velocity to be represented by a vector, because angular velocity is the ratio of an infinitesimal rotation angle to an infinitesimal time:

$$\boldsymbol{\omega} = \frac{\delta\boldsymbol{\theta}}{\delta t}$$

Therefore, dividing Equation 1.106 by δt, we have

$$\frac{\delta\mathbf{r}}{\delta t} = \frac{\delta\boldsymbol{\theta}}{\delta t} \times \mathbf{r}$$

or, in passing to the limit, $\delta t \to 0$

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$$

as before.

1.16 Gradient Operator

We now turn to the most important member of a class called **vector differential operators**—the **gradient operator**.

Consider a scalar ϕ that is an explicit function of the coordinates x_i and, moreover, is a continuous, single-valued function of these coordinates throughout a certain region of space. Under a coordinate transformation that carries the x_i into the x'_i, $\phi'(x'_1, x'_2, x'_3) = \phi(x_1, x_2, x_3)$, and by the chain rule of differentiation, we can write

$$\frac{\partial \phi'}{\partial x'_1} = \sum_j \frac{\partial \phi}{\partial x_j} \frac{\partial x_j}{\partial x'_1} \tag{1.109}$$

The case is similar for $\partial\phi'/\partial x'_2$ and $\partial\phi'/\partial x'_3$, so in general we have

$$\frac{\partial \phi'}{\partial x'_i} = \sum_j \frac{\partial \phi}{\partial x_j} \frac{\partial x_j}{\partial x'_i} \tag{1.110}$$

The inverse coordinate transformation is

$$x_j = \sum_k \lambda_{kj} x'_k \tag{1.111}$$

Differentiating,

$$\frac{\partial x_j}{\partial x'_i} = \frac{\partial}{\partial x'_i}\left(\sum_k \lambda_{kj} x'_k\right) = \sum_k \lambda_{kj}\left(\frac{\partial x'_k}{\partial x'_i}\right) \tag{1.112}$$

But the term in the last parentheses is just δ_{ik}, so

$$\frac{\partial x_j}{\partial x'_i} = \sum_k \lambda_{kj} \delta_{ik} = \lambda_{ij} \tag{1.113}$$

Substituting Equation 1.113 into Equation 1.110, we obtain

$$\frac{\partial \phi'}{\partial x_i'} = \sum_j \lambda_{ij} \frac{\partial \phi}{\partial x_j} \tag{1.114}$$

Because it follows the correct transformation equation of a vector (Equation 1.44), the function $\partial\phi/\partial x_j$ is the *j*th component of a vector termed the **gradient** of the function ϕ. Note that even though ϕ is a scalar, the *gradient* of ϕ is a *vector*. The gradient of ϕ is written either as **grad**ϕ or as $\nabla\phi$ ("del" ϕ)

Because the function ϕ is an arbitrary scalar function, it is convenient to define the differential operator described in the preceding in terms of the gradient operator:

$$(\mathbf{grad})_i = \nabla_i = \frac{\partial}{\partial x_i} \tag{1.115}$$

We can express the complete vector gradient operator as

$$\boxed{\mathbf{grad} = \nabla = \sum_i \mathbf{e}_i \frac{\partial}{\partial x_i}} \qquad \text{Gradient} \tag{1.116}$$

The gradient operator can (a) operate directly on a scalar function, as in $\nabla\phi$; (b) be used in a scalar product with a vector function, as in $\nabla \cdot \mathbf{A}$ (the *divergence* (div) of **A**); or (c) be used in a vector product with a vector function, as in $\nabla \times \mathbf{A}$ (the *curl* of **A**). We present the grad, divergence, and curl:

$$\mathbf{grad}\ \phi = \nabla\phi = \sum_i \mathbf{e}_i \frac{\partial \phi}{\partial x_i} \tag{1.117a}$$

$$\text{div}\ \mathbf{A} = \nabla \cdot \mathbf{A} = \sum_i \frac{\partial A_i}{\partial x_i} \tag{1.117b}$$

$$\mathbf{curl}\ \mathbf{A} = \nabla \times \mathbf{A} = \sum_{i,j,k} \varepsilon_{ijk} \frac{\partial A_k}{\partial x_j} \mathbf{e}_i \tag{1.117c}$$

To see a physical interpretation of the gradient of a scalar function, consider the three-dimensional and topographical maps of Figure 1-20.

The closed loops of part b represent lines of constant height. Let ϕ denote the height at any point $\phi = \phi(x_1, x_2, x_3)$. Then

$$d\phi = \sum_i \frac{\partial \phi}{\partial x_i} dx_i = \sum_i (\nabla \phi)_i \, dx_i$$

The components of the displacement vector ds are the incremental displacements in the direction of the three orthogonal axes:

$$d\mathbf{s} = (dx_1, dx_2, dx_3) \tag{1.118}$$

Therefore

$$\boxed{d\phi = (\nabla \phi) \cdot d\mathbf{s}} \tag{1.119}$$

Let ds be directed tangentially along one of the isolatitude lines (i.e., along a line for which ϕ = const.), as indicated in Figure 1-20.

Because ϕ = const, for this case, $d\phi = 0$.

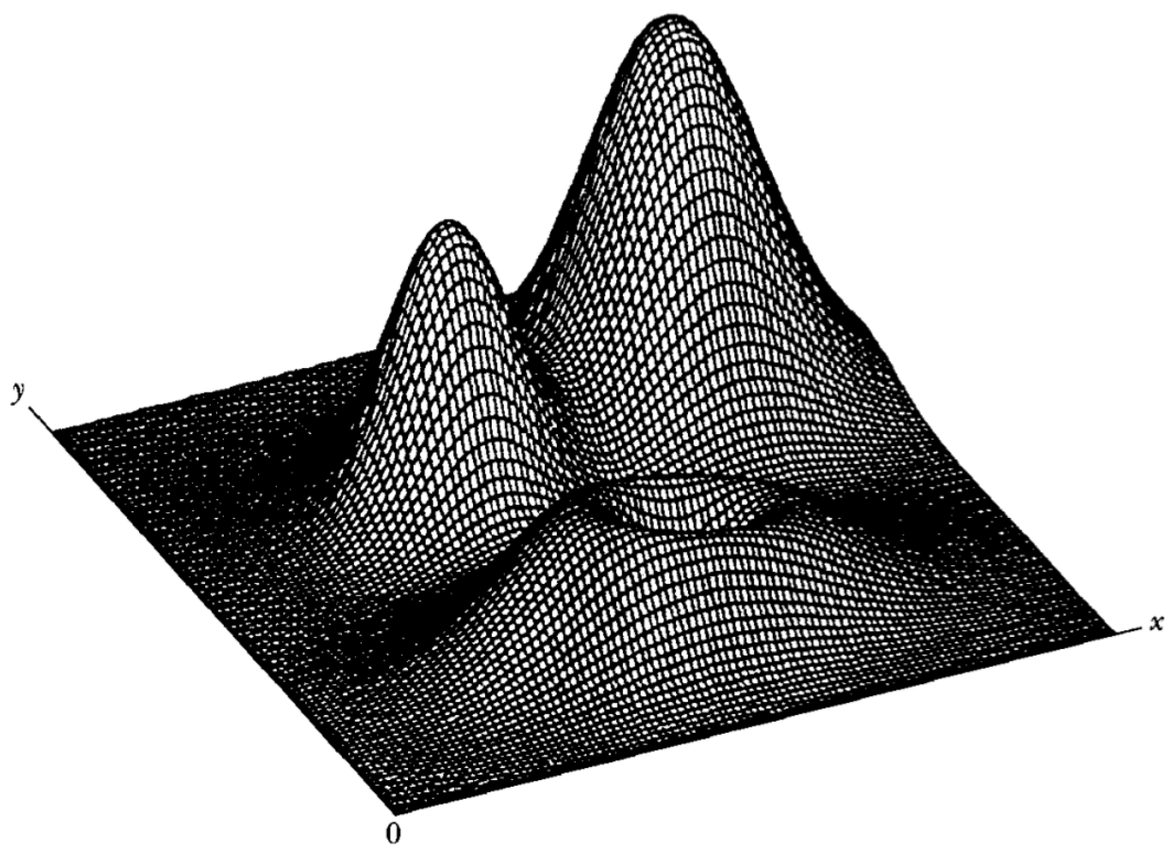


FIGURE 1-20 (a) A three-dimensional contour map can be represented by (b) a topographical map of lines ϕ representing constant height. The gradient $\nabla\phi$ represents the direction perpendicular to the constant ϕ lines.

But, because neither $\nabla\phi$ nor d**s** is in general zero, they must therefore be perpendicular to each other. Thus $\nabla\phi$ is normal to the line (or in three dimensions, to the surface) for which ϕ = const. The maximum value of $d\phi$ results when $\nabla\phi$ and d**s** are in the same direction; then,

$$(d\boldsymbol{\phi})_{\max} = |\nabla\boldsymbol{\phi}|\, ds, \quad \text{for} \quad \nabla\boldsymbol{\phi} \| d\mathbf{s}$$

or

$$|\nabla\boldsymbol{\phi}| = \left(\frac{d\boldsymbol{\phi}}{ds}\right)_{\max} \tag{1.120}$$

Therefore, $\nabla\phi$ is in the direction of the greatest change in ϕ.

We can summarize these results as follows:

1. **The vector $\nabla\phi$ is, at any point, normal to the lines or surfaces for which ϕ = const.**
2. **The vector $\nabla\phi$ has the direction of the maximum change in ϕ.**
3. **Because any direction in space can be specified in terms of the unit vector $\mathbf{n}$ in that direction, the rate of change of ϕ in the direction of $\mathbf{n}$ (the *directional derivative* of ϕ) can be found from $\mathbf{n} \cdot \nabla\phi \equiv \partial\phi/\partial n$.**

The successive operation of the gradient operator produces

$$\nabla \cdot \nabla = \sum_i \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_i} = \sum_i \frac{\partial^2}{\partial x_i^2} \tag{1.121}$$

This important product operator, called the **Laplacian**, is also written

$$\boxed{\nabla^2 = \sum_i \frac{\partial^2}{\partial x_i^2}} \tag{1.122}$$

When the Laplacian operates on a scalar, we have, for example,

$$\nabla^2\psi = \sum_i \frac{\partial^2\psi}{\partial x_i^2} \tag{1.123}$$

1.17 Integration of Vectors

The vector resulting from the volume integration of a vector function $\mathbf{A} = \mathbf{A}(x_i)$ throughout a volume V is given by

$$\int_V \mathbf{A}\, dv = \left(\int_V A_1 dv, \int_V A_2 dv, \int_V A_3 dv \right) \tag{1.124}$$

The symbol $\int_V$ actually represents a triple integral over a certain volume V. Similarly, the symbol $\int_S$ stands for a double integral over a certain surface S.

Thus, we integrate the vector **A** throughout V simply by performing three separate, ordinary integrations.

The integral over surface S of the projection of a vector function $\mathbf{A} = \mathbf{A}(x_i)$ onto the normal to that surface is defined to be

$$\int_S \mathbf{A} \cdot d\mathbf{a}$$

where d**a** is an element of area of the surface (Figure 1-21).

We write d**a** as a vector quantity because we may attribute to it not only a magnitude da but also a direction corresponding to the normal to the surface at the point in question. If the unit normal vector is **n**, then

$$d\mathbf{a} = \mathbf{n}\, da \tag{1.125}$$

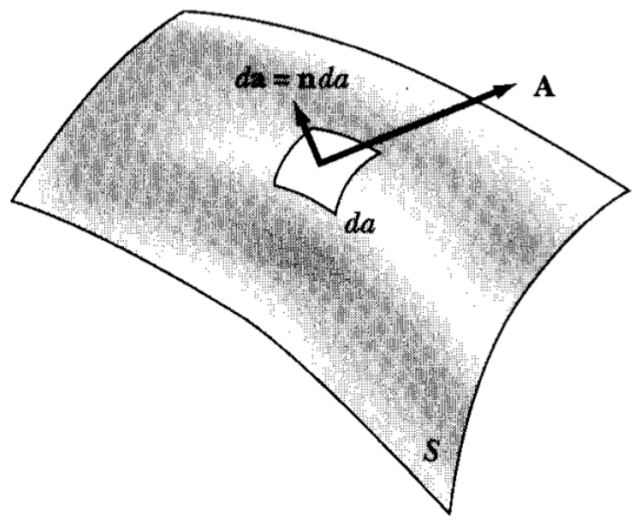


FIGURE 1-21 The differential *d***a** is an element of area of the surface. Its direction is normal to the surface.

Thus, the components of d**a** are the projection of the element of area on the three mutually perpendicular planes defined by the rectangular axes:

$$da_1 = dx_2\,dx_3, \quad \text{etc.} \tag{1.126}$$

Therefore, we have

$$\int_S \mathbf{A}\cdot d\mathbf{a} = \int_S \mathbf{A}\cdot \mathbf{n}da \tag{1.127}$$

or

$$\int_S \mathbf{A}\cdot d\mathbf{a} = \int_S \sum_i A_i da_i \tag{1.128}$$

Equation 1.127 states that the integral of **A** over the surface S is the integral of the normal component of **A** over this surface.

The normal to a surface may be taken to lie in either of two possible directions ("up" or "down"); thus the sign of n is ambiguous. If the surface is dosed, we adopt the convention that the *outward* normal is positive

The **line integral** of a vector function $\mathbf{A} = \mathbf{A}(x_i)$ along a given path extending from the point B to the point C is given by the integral of the component of **A** along the path

$$\int_{BC} \mathbf{A}\cdot d\mathbf{s} = \int_{BC} \sum_i A_i\,dx_i \tag{1.129}$$

The quantity d**s** is an element of length along the given path (Figure 1-22).

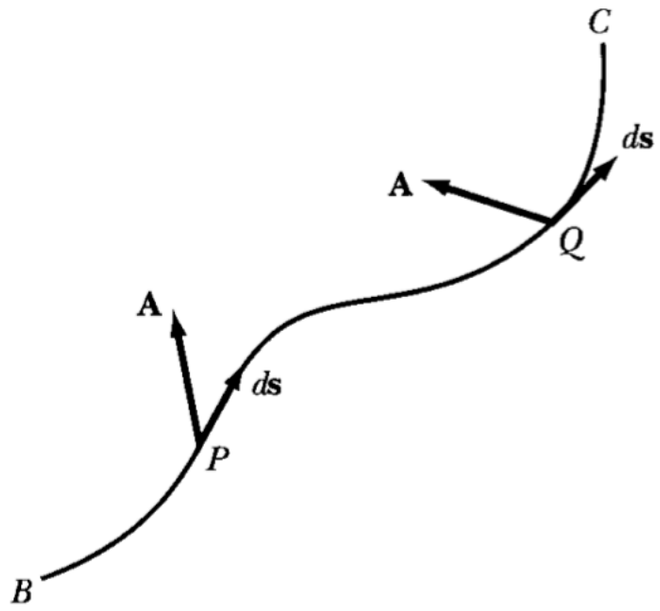


FIGURE 1-22 The element *d***s** is an element of length along the given path from *B* to *C*. Its direction is along the path at a given point.

The direction of d**s** is taken to be positive along the direction the path is traversed. In Figure 1-22 at point P, the angle between d**s** and **A** is less than $\pi/2$, so **A**•d**s** is positive at this point. At point Q, the angle is greater than $\pi/2$, and the contribution to the integral at this point is negative.

It is often useful to relate certain surface integrals to either volume integrals (**Gauss's theorem**) or line integrals (**Stokes's theorem**). Consider Figure 1-23, which shows a closed volume V enclosed by the surface S.

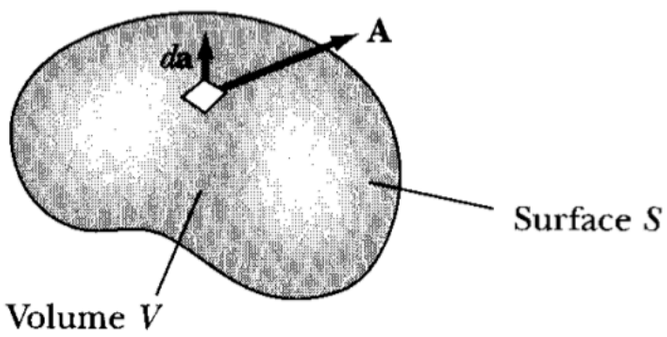


FIGURE 1-23 The differential *d***a** is an element of area on a surface *S* that surrounds a closed volume *V*.

Let the vector **A** and its first derivatives be continuous throughout the volume. Gauss's theorem states that the surface integral of **A** over the closed surface S is equal to the volume integral of the divergence of **A** (∇•A) throughout the volume V enclosed by the surface S. We write this mathematically as

$$\int_S \mathbf{A} \cdot d\mathbf{a} = \int_V \nabla \cdot \mathbf{A}\, dv \tag{1.130}$$

Gauss's theorem is sometimes also called the *divergence theorem*. The theorem is particularly useful in dealing with the mechanics of continuous media.

See Figure 1-24 for the physical description needed for Stokes's theorem, which applies to an open surface S and the contour path C that defines the surface

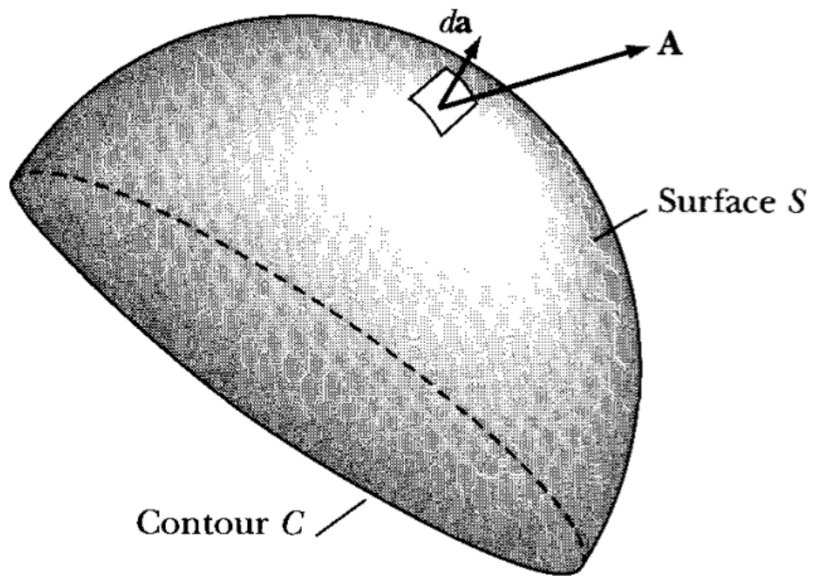


FIGURE 1-24 A contour path *C* defines an open surface *S*. A line integral around the path *C* and a surface integral over the surface *S* is required for Stokes's theorem.

The curl of the vector **A** (∇ x **A**) must exist and be integrable over the entire surface S. Stokes's theorem states that the line integral of the vector **A** around the contour path C is equal to the surface integral of the curl of **A** over the surface defined by C. We write it mathematically as

$$\int_C \mathbf{A} \cdot d\mathbf{s} = \int_S (\nabla \times \mathbf{A}) \cdot d\mathbf{a}$$

where the line integral is around the closed contour path C.

Stokes's theorem is particularly useful in reducing certain surface integrals (two dimensional) to, it is hoped, a simpler line integral (one dimensional). Both Gauss's and Stokes's theorems have wide application in vector calculus. In addition to mechanics, they are also useful in electromagnetic applications and in potential theory.

CHAPTER 2 - Newtonian Mechanics - Single Particle

2.1 Introduction

The science of mechanics seeks to provide a precise and consistent description of the dynamics of particles and systems of particles, that is, a set of physical laws mathematically describing the motions of bodies and aggregates of bodies. For this, we need certain fundamental concepts such as distance and time. The combination of the concepts of distance and time allows us to define the velocity and acceleration of a particle. The third fundamental concept, mass, requires some elaboration, which we give when we discuss Newton's laws.

Physical laws must be based on experimental fact. We cannot expect a priori that the gravitational attraction between two bodies must vary exactly as the inverse square of the distance between them. But experiment indicates that this is so. Once a set of experimental data has been correlated and a postulate has been formulated regarding the phenomena to which the data refer, then various implications can be worked out. If these implications are all verified by experiment, we may believe that the postulate is generally true. The postulate then assumes the status of a physical law. If some experiments disagree with the predictions of the law, the theory must be modified to be consistent with the facts.

Newton provided us with the fundamental laws of mechanics. We state these laws here in modem terms, discuss their meaning, and then derive the implications of the laws in various situations.* But the logical structure of the science of mechanics is not straightforward. Our line of reasoning in interpreting Newton's laws is not the only one possible.^ We do not pursue in any detail the philosophy of mechanics but rather give only sufficient elaboration of Newton's laws to allow us to continue with the discussion of classical dynamics. We devote our attention in this chapter to the motion of a single particle, leaving systems of particles to be discussed in Chapters 9 and 11-13.

2.2 Newton's Laws

We begin by simply stating in conventional form Newton's laws of mechanics:

I. *A body remains at rest or in uniform motion unless acted upon by a force.*

II. *A body acted upon by a force moves in such a manner that the time rate of change of momentum equals the force.*

III. *If two bodies exert forces on each other, these forces are equal in magnitude and opposite in direction.*

These laws are so familiar that we sometimes tend to lose sight of their true significance (or lack of it) as physical laws. The First Law, for example, is meaningless without the concept of "force," a word Newton used in all three laws. In fact, standing alone, the First Law conveys a precise meaning only for zero force; that is, a body remaining at rest or in uniform (i.e., unaccelerated, rectilinear) motion is subject to no force whatsoever A body moving in this manner is termed a **free body** (or **free particle**). The question of the frame of reference with respect to which the "uniform motion" is to be measured is discussed in the following section.

In pointing out the lack of content in Newton's First Law, Sir Arthur Eddington observed, somewhat facetiously, that all the law actually says is that "every particle continues in its state of rest or uniform motion in a straight line except insofar as it doesn't." This is hardly fair to Newton, who meant something very definite by his statement. But it does emphasize that the First Law by itself provides us with only a qualitative notion regarding “force."

The Second Law provides an explicit statement: Force is related to the time rate of change of *momentum*. Newton appropriately defined momentum (although he used the term quantity of motion) to be the product of mass and velocity, such that

$$\mathbf{p} \equiv m\mathbf{v} \tag{2.1}$$

Therefore, Newton's Second Law can be expressed as

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d}{dt}(m\mathbf{v}) \tag{2.2}$$

The definition of force becomes complete and precise only when "mass" is defined. Thus the First and Second Laws are not really "laws" in the usual sense; rather, they may be considered definitions. Because length, time, and mass are concepts normally already understood, we use Newton's First and Second Laws as the operational definition of force. Newton's Third Law, however, is indeed a law. It is a statement concerning the real physical world and contains all of the physics in Newton's laws of motion.

We must hasten to add, however, that the Third Law is not a general law of nature. The law does apply when the force exerted by one (point) object on another (point) object is directed along the line connecting the objects. Such forces are called central forces; the Third Law applies whether a central force is attractive or repulsive. Gravitational and electrostatic forces are central forces, so Newton's laws can be used in problems involving these types of forces. Sometimes, elastic forces (which are actually macroscopic manifestations of microscopic electrostatic forces) are central. For example, two point objects connected by a straight spring or elastic string are subject to forces that obey the Third Law. Any force that depends on the velocities of the interacting bodies is noncentral, and the Third Law may not apply. Velocity-dependent forces are characteristic of interactions that propagate with finite velocity. Thus the force between moving electric charges does not obey the Third Law, because the force propagates with the velocity of light. Even the gravitational force between moving bodies is velocity dependent, but the effect is small and difficult to detect. The only observable effect is the precession of the perihelia of the inner planets (see Section 8.9). We will return to a discussion of Newton's Third Law in Chapter 9.

To demonstrate the significance of Newton's Third Law, let us paraphrase it in the following way, which incorporates the appropriate definition of mass:

> **III′.** *If two bodies constitute an ideal, isolated system, then the accelerations of these bodies are always in opposite directions, and the ratio of the magnitudes of the accelerations is constant. This constant ratio is the inverse ratio of the masses of the bodies.*

With this statement, we can give a practical definition of mass and therefore give precise meaning to the equations summarizing Newtonian dynamics. For two isolated bodies, 1 and 2, the Third Law states that

$$\mathbf{F}_1 = -\mathbf{F}_2 \tag{2.3}$$

Using the definition of force as given by the Second Law, we have

$$\frac{d\mathbf{p}_1}{dt} = -\frac{d\mathbf{p}_2}{dt} \tag{2.4a}$$

or, with constant masses,

$$m_1\left(\frac{d\mathbf{v}_1}{dt}\right) = m_2\left(-\frac{d\mathbf{v}_2}{dt}\right) \tag{2.4b}$$

and, because acceleration is the time derivative of velocity,

$$m_1(\mathbf{a}_1) = m_2(-\mathbf{a}_2) \tag{2.4c}$$

Hence,

$$\frac{m_2}{m_1} = -\frac{a_1}{a_2} \tag{2.5}$$

where the negative sign indicates only that the two acceleration vectors are oppositely directed. Mass is taken to be a positive quantity.

We can always select, say, m_1 as the *unit mass*. Then, by comparing the ratio of accelerations when m_1 is allowed to interact with any other body, we can determine the mass of the other body. To measure the accelerations, we must have appropriate clocks and measuring rods; also, we must choose a suitable coordinate system or reference frame. The question of a "suitable reference frame" is discussed in the next section.

One of the more common methods of determining the mass of an object is by weighing - for example, by comparing its weight to that of a standard by means of a beam balance. This procedure makes use of the fact that in a gravitational field the weight of a body is just the gravitational force acting on the body; that is, Newton's equation $\mathbf{F} = m\mathbf{a}$. becomes $\mathbf{W} = m\mathbf{g}$, where $\mathbf{g}$ is the acceleration due to gravity. The validity of using this procedure rests on a fundamental assumption: that the mass m appearing in Newton's equation and defined according to Statement III' is equal to the mass m that appears in the gravitational force equation. These two masses are called the **inertial mass** and **gravitational mass**, respectively. The definitions may be stated as follows:

Inertial Mass: *That mass determining the acceleration of a body under the action of a given force.*

Gravitational Mass: *That mass determining the gravitational forces between a body and other bodies.*

Galileo was the first to test the equivalence of inertial and gravitational mass in his (perhaps apocryphal) experiment with falling weights at the Tower of Pisa. Newton also considered the problem and measured the periods of pendula of equal lengths but with bobs of different materials. Neither Newton nor Galileo found any difference, but the methods were quite crude. In 1890 Eotvos devised an ingenious method to test the equivalence of inertial and gravitational masses. Using two objects made of different materials, he compared the effect of the Earth's gravitational force (i.e., the weight) with the effect of the inertial force caused by the Earth's rotation. The experiment involved a null method using a sensitive torsion balance and was therefore highly accurate.

More recent experiments (notably those of Dicke), using essentially the same method, have improved the accuracy, and we know now that inertial and gravitational mass are identical to within a few parts in 10^{12}. This result is considerably important in the general theory of relativity. The assertion of the *exact* equality of inertial and gravitational mass is termed the **principle of equivalence**.

Newton's Third Law is stated in terms of two bodies that constitute an isolated system. It is impossible to achieve such an ideal condition; every body in the universe interacts with every other body, although the force of interaction may be far too weak to be of any practical importance if great distances are involved. Newton avoided the question of how to disentangle the desired effects from all the extraneous effects. But this practical difficulty only emphasizes the enormity of Newton's assertion made in the Third Law. It is a tribute to the depth of his perception and physical insight that the conclusion, based on limited observations, has successfully borne the test of experiment for 300 years. Only within the 20th century did measurements of sufficient detail reveal certain discrepancies with the predictions of Newtonian theory. The pursuit of these details led to the development of relativity theory and quantum mechanics."

Another interpretation of Newton's Third Law is based on the concept of momentum. Rearranging Equation 2.4a gives

$$\frac{d}{dt}(\mathbf{p}_1 + \mathbf{p}_2) = 0$$

or

$$\mathbf{p}_1 + \mathbf{p}_2 = \textit{constant} \tag{2.6}$$

The statement that momentum is conserved in the isolated interaction of two particles is a special case of the more general **conservation of linear momentum**. Physicists cherish general conservation laws, and the conservation of linear momentum is believed always to be obeyed. Later we shall modify our definition of momentum from Equation 2.1 for high velocities approaching the speed of light.

2.3 Frames of Reference

Newton realized that, for the laws of motion to have meaning, the motion of bodies must be measured relative to some reference frame. A reference frame is called an **inertial frame** if Newton's laws are indeed valid in that frame; that is, if a body subject to no external force moves in a straight line with constant velocity (or remains at rest), then the coordinate system establishing this fact is an inertial reference frame. This is a clear-cut operational definition and one that also follows from the general theory of relativity.

If Newton's laws are valid in one reference frame, then they are also valid in any reference frame in uniform motion (i.e., not accelerated) with respect to the first system. This is a result of the fact that the equation $\mathbf{F} = m\ddot{\mathbf{r}}$ involves the second time derivative of $\mathbf{r}$: A change of coordinates involving a constant velocity does not influence the equation. This result is called **Galilean invariance** or the **principle of Newtonian relativity**.

Relativity theory has shown us that the concepts of *absolute rest* and an *absolute* inertial reference frame are meaningless. Therefore, even though we conventionally adopt a reference frame described with respect to the "fixed" stars - and, indeed, in such a frame the Newtonian equations are valid to a high degree of accuracy - such a frame is, in fact, not an absolute inertial frame. We may, however, consider the "fixed" stars to define a reference frame that approximates an "absolute" inertial frame to an extent quite sufficient for our present purposes.

Although the fixed-star reference frame is a conveniently definable system and one suitable for many purposes, we must emphasize that the fundamental definition of an inertial frame makes no mention of stars, fixed or otherwise. If a body subject to no force moves with constant velocity in a certain coordinate system, that system is, by definition, an inertial frame. Because precisely describing the motion of a real physical object in the real physical world is normally difficult, we usually resort to idealizations and approximations of varying degree; that is, we ordinarily neglect the lesser forces on a body if these forces do not significantly affect the body's motion.

In Chapter 10, we discuss the modification of Newton's equations that must be made if it is desired to describe the motion of a body with respect to a *noninertial* frame of reference, that is, a frame that is accelerated with respect to an inertial frame.

If we wish to describe the motion of, say, a free particle and if we choose for this purpose some coordinate system in an inertial frame, then we require that the (vector) equation of motion of the particle be independent of the *position* of the origin of the coordinate system and independent of its *orientation* in space. We further require that **time** be homogeneous; that is, a free particle moving with a certain constant velocity in the coordinate system during a certain time interval must not, during a later time interval, be found to move with a different velocity.

We can illustrate the importance of these properties by the following example. Consider, as in Figure 2-1.

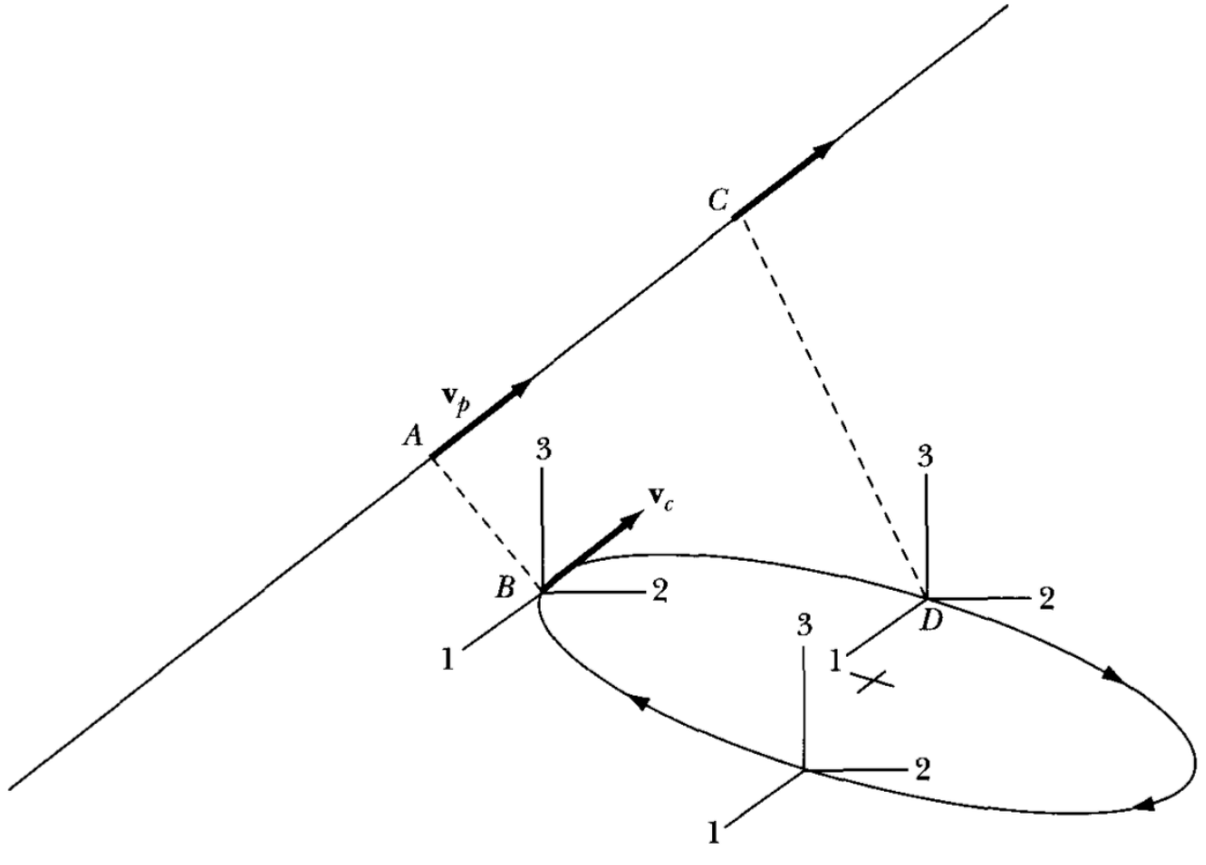


FIGURE 2-1 We choose to describe the path of a free particle moving along the path *AC* in a rectangular coordinate system whose origin moves in a circle. Such a system is not an inertial reference frame.

A free particle moving along a certain path AC. To describe the particle's motion, let us choose a rectangular coordinate system whose origin moves in a circle, as shown. For simplicity, we let the orientation of the axes be fixed in space. The particle moves with a velocity $\mathbf{v}_p$ relative to an inertial reference frame. If the coordinate system moves with a linear velocity $\mathbf{v}_c$ when at the point B, and if $\mathbf{v}_c = \mathbf{v}_p$, then to an observer in the moving coordinate system the particle (at A) will appear to be *at rest*. At some later time, however, when the particle is at C and the coordinate system is at D, the particle will appear to accelerate with respect to the observer. We must, therefore, conclude that the rotating coordinate system does not qualify as an inertial reference frame.

These observation are not sufficient to decide whether time is homogeneous. To reach such a conclusion, repeated measurements must be made in identical situations at various times; identical results would indicate the homogeneity of time

Newton's equations do not describe the motion of bodies in noninertial systems. We can devise a method to describe the motion of a particle by a rotating coordinate system, but, as we shall see in Chapter 10, the resulting equations contain several terms that do not appear in the simple Newtonian equation $\mathbf{F} = m\mathbf{a}$. For the moment, then, we restrict our attention to inertial reference frames to describe the dynamics of particles.

2.4 The Equation of Motion for a Particle

Newton's equation $\mathbf{F} = d\mathbf{p}/dt$ can be expressed alternatively as

$$\mathbf{F} = \frac{d}{dt}(m\mathbf{v}) = m\frac{d\mathbf{v}}{dt} = m\ddot{\mathbf{r}} \tag{2.7}$$

if we assume that the mass m does not vary with time. This is a second-order differential equation that may be integrated to find $\mathbf{r} = \mathbf{r}(t)$ if the function $\mathbf{F}$ is known. Specifying the initial values of $\mathbf{r}$ and $\dot{\mathbf{r}} = \mathbf{v}$ then allows us to evaluate the two arbitrary constants of integration. We then determine the motion of a particle by the force function $\mathbf{F}$ and the initial values of position $\mathbf{r}$ and velocity $\mathbf{v}$.

The force **F** may be a function of any combination of position, velocity, and time and is generally denoted as **F**(**r**, **v**, t). For a given dynamic system, we normally want to know **r** and **v** as a function of time. Solving Equation 2.7 will help us do this by solving for $\ddot{\mathbf{r}}$. Applying Equation 2.7 to physical situations is an important part of mechanics.

In this chapter, we examine several examples in which the force function is known. We begin by looking at simple force functions (either constant or dependent on only one of **r**, **v**, and t) in only one spatial dimension as a refresher of earlier physics courses. It is important to form good habits in problem solving. Here are some useful problem-solving techniques.

1. Make a sketch of the problem, indicating forces, velocities, and so forth.
2. Write down the given quantities.
3. Write down useful equations and what is to be determined.
4. Strategy and the principles of physics must be used to manipulate the equations to find the quantity sought. Algebraic manipulations as well as differentiation or integration is usually required. Sometimes numerical calculations using a computer are the easiest, if not the only, method of solution.
5. Finally, put in the actual values for the assumed variable names to determine the quantity sought.

Let us first consider the problem of a block sliding on an inclined plane. Let the angle of the inclined plane be θ and the mass of the block be 100 g. The sketch of the problem is shown in Figure 2-2a.

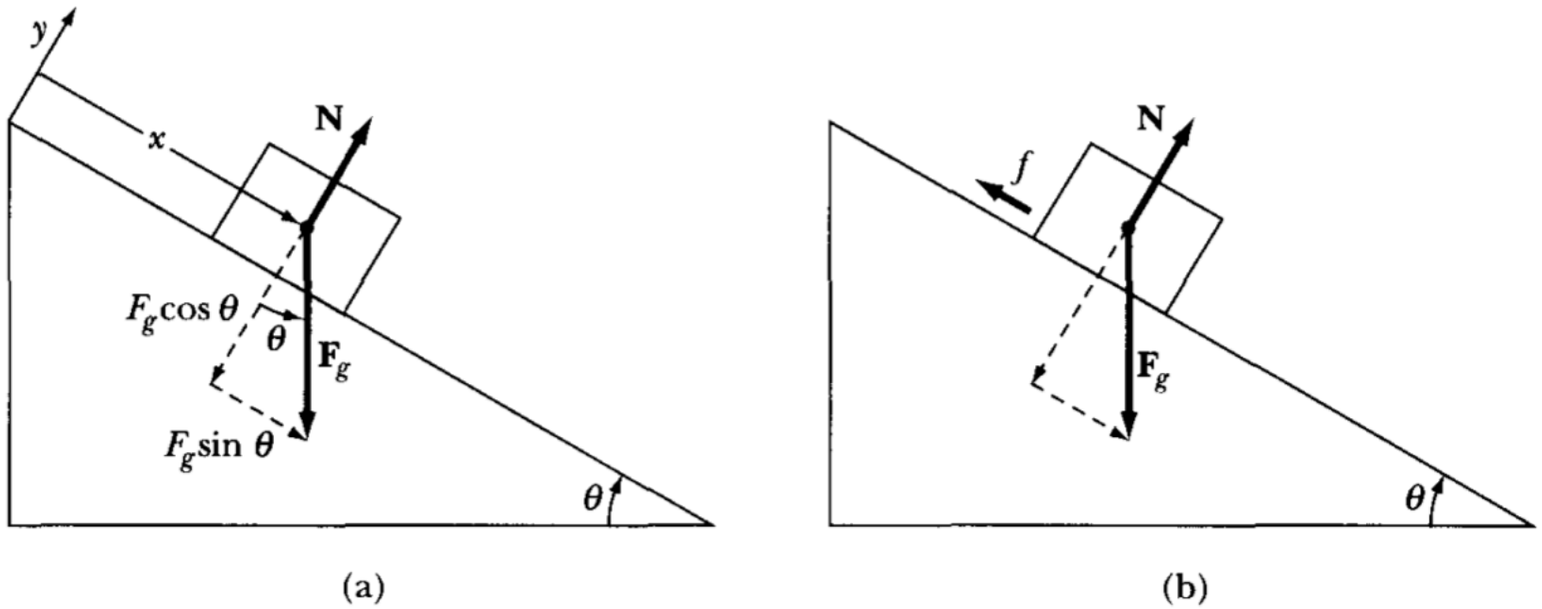


FIGURE 2-2 Examples 2.1 and 2.2.

EXAMPLE 2.1

If a block slides without friction down a fixed, inclined plane with $\theta = 30°$, what is the block's acceleration?

Solution. Two forces act on the block (see Figure 2-2a): the gravitational force $\mathbf{F}_g$ and the plane's normal force **N** pushing upward on the block (no friction in this example). The block is constrained to be on the plane, and the only direction the block can move is the x-direction, up and down the plane. We take the + x-direction to be down the plane. The total force $\mathbf{F}_{net}$ is constant; Equation 2.7 becomes

$$\mathbf{F}_{net} = \mathbf{F}_g + \mathbf{N}$$

and because $\mathbf{F}_{net}$ is the net resultant force acting on the block,

$$\mathbf{F}_{net} = m\ddot{\mathbf{r}}$$

or

$$\mathbf{F}_g + \mathbf{N} = m\ddot{\mathbf{r}} \tag{2.8}$$

This vector must be applied in two directions: x and y (perpendicular to x).

The component of force in the 3)-direction is zero, because no acceleration occurs in this direction. The force $\mathbf{F}_g$ is divided vectorially into its x- and y-components (dashed lines in Figure 2-2a). Equation 2.8 becomes

y-direction

$$-F_g \cos\theta + N = 0 \tag{2.9}$$

x-direction

$$F_g \sin\theta = m\ddot{x} \tag{2.10}$$

with the required result

$$\ddot{x} = \frac{F_g}{m}\sin\theta = \frac{mg\sin\theta}{m} = g\sin\theta$$

$$\ddot{x} = g\sin(30°) = \frac{g}{2} = 4.9\ \text{m/s}^2 \tag{2.11}$$

Therefore the acceleration of the block is a constant.

We can find the velocity of the block after it moves from rest a distance x_0 down the plane by multiplying Equation 2.11 by $2\dot{x}$ and integrating

$$2\dot{x}\ddot{x} = 2\dot{x}g\sin\theta$$

$$\frac{d}{dt}(\dot{x}^2) = 2g\sin\theta\frac{dx}{dt}$$

$$\int_0^{v_0^2} d(\dot{x}^2) = 2g\sin\theta\int_0^{x_0} dx$$

At t=0, both $x = \dot{x} = 0$, and, at $t = t_{final}$, $x = x_0$, and the velocity $\dot{x} = v_0$.

$$v_0^2 = 2g \sin\theta \, x_0$$

$$v_0 = \sqrt{2g \sin\theta \, x_0}$$

EXAMPLE 2.2

If the coefficient of static friction between the block and plane in the previous example is $\mu_s = 0.4$, at what angle θ will the block start sliding if it is initially at rest?

Solution. We need a new sketch to indicate the additional frictional force *f* (see Figure 2-2b). The static frictional force has the approximate maximum value

$$f_{max} = \mu_s N \tag{2.12}$$

and Equation 2.7 becomes, in component form,

y-direction

$$-F_g \cos\theta + N = 0 \tag{2.13}$$

x-direction

$$-f_s + F_g \sin\theta = m\ddot{x} \tag{2.14}$$

The static frictional force f_s will be some value $f_s \leq f_{max}$ required to keep $\ddot{x} = 0$ - that is, to keep the block at rest. However, as the angle θ of the plane increases, eventually the static frictional force will be unable to keep the block at rest. At that angle θ' , f_s , becomes

$$f_s(\theta = \theta') = f_{max} = \mu_s N = \mu_s F_g \cos\theta$$

and

$$m\ddot{x} = F_g \sin\theta - f_{max}$$

$$m\ddot{x} = F_g \sin\theta - \mu_s F_g \cos\theta \tag{2.15}$$

$$\ddot{x} = g(\sin\theta - \mu_s \cos\theta)$$

Just before the block starts to slide, the acceleration $\ddot{x} = 0$, so

$$\sin\theta - \mu_s \cos\theta = 0$$

$$\tan\theta = \mu_s = 0.4$$

$$\theta = \tan^{-1}(0.4) = 22°$$

EXAMPLE 2.3

After the block in the previous example begins to slide, the coefficient of kinetic (sliding) friction becomes $\mu_k = 0.3$. Find the acceleration for the angle $\theta = 30°$.

Solution. Similarly to Example 2.2, the kinetic friction becomes (approximately)

$$f_k = \mu_k N = \mu_k F_g \cos\theta \tag{2.16}$$

and

$$m\ddot{x} = F_g \sin\theta - f_k = mg(\sin\theta - \mu_k \cos\theta) \tag{2.17}$$

$$\ddot{x} = g(\sin\theta - \mu_k \cos\theta) = 0.24\, g \tag{2.18}$$

Generally, the force of static friction ($f_{max} = \mu_s N$) is greater than that of kinetic friction ($f_k = \mu_k N$). This can be observed in a simple experiment. If we lower the angle θ below 16.7°, we find that $\ddot{x} < 0$, and the block eventually stops. If we raise the block back up above $\theta = 16.7°$, we find that the block does not start sliding again until $\theta \geq 22°$ (Example 2.2). The static friction determines when it starts moving again. There is not a discontinuous acceleration as the block starts moving, because of the difference between μ_s and μ_k. For small speeds, the coefficient of friction changes rather quickly from μ_s to μ_k.

The subject of friction is still an interesting and important area of research. There are still surprises. For example, even though we calculate the absolute value of the frictional force as $f = \mu N$, research has shown that the frictional force is directly proportional, not to the load, but to the microscopic area of contact between the two objects (as opposed to the apparent contact area). We use μN an approximation because, as N increases, so does the actual contact area on a microscopic level. For hundreds of years before the 1940s, it was accepted that the load - and not the area - was directly responsible. We also believe that the static frictional force is larger than that of kinetic friction because the bonding of atoms between the two objects does not have as much time to develop in kinetic motion.

Effects of Retarding Forces

We should emphasize that the force F in Equation 2.7 is not necessarily constant, and indeed, it may consist of several distinct parts, as seen in the previous examples. For example, if a particle falls in a constant gravitational field, the gravitational force is $\mathbf{F}_g = m\mathbf{g}$, where $\mathbf{g}$ is the acceleration of gravity. If, in addition, a retarding force $\mathbf{F}_r$ exists that is some function of the instantaneous speed, then the total force is

$$\begin{aligned} \mathbf{F} &= \mathbf{F}_g + \mathbf{F}_r \\ &= m\mathbf{g} + \mathbf{F}_r(v) \end{aligned} \tag{2.19}$$

It is frequently sufficient to consider that $\mathbf{F}_r(v)$ is simply proportional to some power of the speed. In general, real retarding forces are more complicated, but the power-law approximation is useful in many instances in which the speed does not vary greatly. Even more to the point, if $F_r \propto v^n$, then the equation of motion can usually be integrated directly, whereas, if the true velocity dependence were used, numerical integration would probably be necessary. With the power-law approximation, we can then write

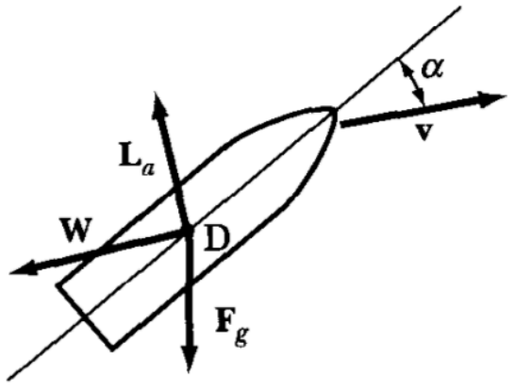
$$\mathbf{F} = m\mathbf{g} - mkv^n \frac{\mathbf{v}}{v} \tag{2.20}$$

where k is a positive constant that specifies the strength of the retarding force and where $\mathbf{v}/v$; is a unit vector in the direction of **v**. Experimentally, we find that, for a relatively small object moving in air, $n \simeq 1$ for velocities less than about 24 m/s (~ 80 ft/s). For higher velocities but below the velocity of sound (~ 330 m/s or 1,100 ft/s), the retarding force is approximately proportional to the square of the velocity. For simplicity, the v^2 dependence is usually taken for speeds up to the speed of sound.

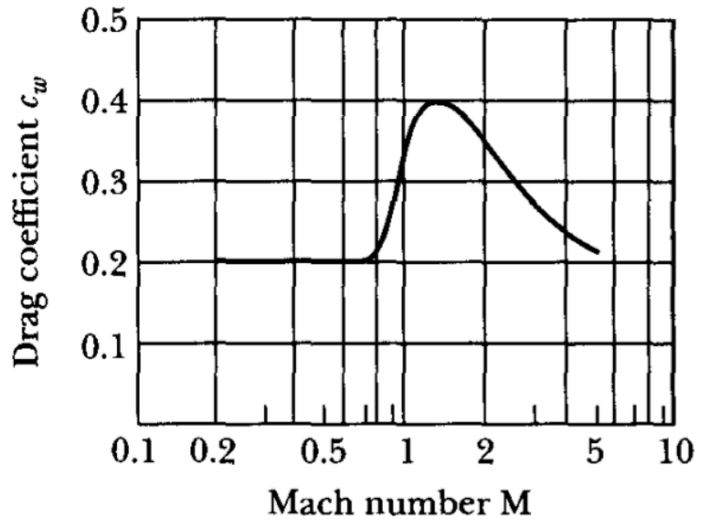
The effect of air resistance is important for a ping-pong ball smashed to an opponent, a high-flying softball hit deep to the outfield, a golfer's chip shot, and a mortar shell lofted against an enemy. Extensive tabulations have been made for military ballistics of projectiles of various sorts for the velocity as a function of flight time. There are several forces on an actual projectile in flight. The air resistance force is called the drag **W** and is opposite to the projectile's velocity as shown in Figure 2-3a. The velocity v is normally not along the symmetry axis of the shell. The component of force acting perpendicular to the drag is called the lift $\mathbf{L}_a$. There may also be various other forces due to the projectile's spin and oscillation, and a calculation of a projectile's ballistic trajectory is quite complex. The Prandtl expression for the air resistance is

$$W = \frac{1}{2} c_W \rho A v^2 \tag{2.21}$$

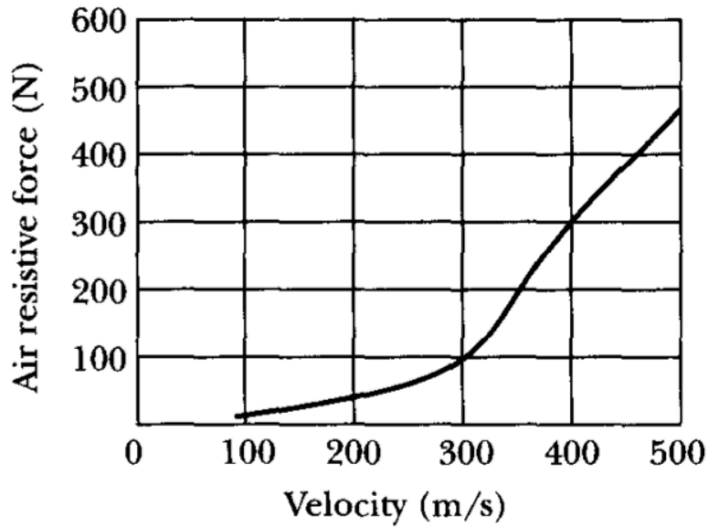
where c_w is the dimensionless drag coefficient, ρ is the air density, v is the velocity, and A is the cross-sectional area of the object (projectile) measured perpendicularly to the velocity. In Figure 2-3b, we plot some typical values for c_w, and in Figures 2-3c and d we display the calculated air resistance W using Equation 2.21 for a projectile diameter of 10 cm and using the values of c_w shown. The air resistance increases dramatically near the speed of sound (Mach number M = speed/speed of sound). Below speeds of about 400 m/s it is evident that an equation of at least second degree is necessary to describe the resistive force. For higher speeds, the retarding force varies approximately linearly with speed.



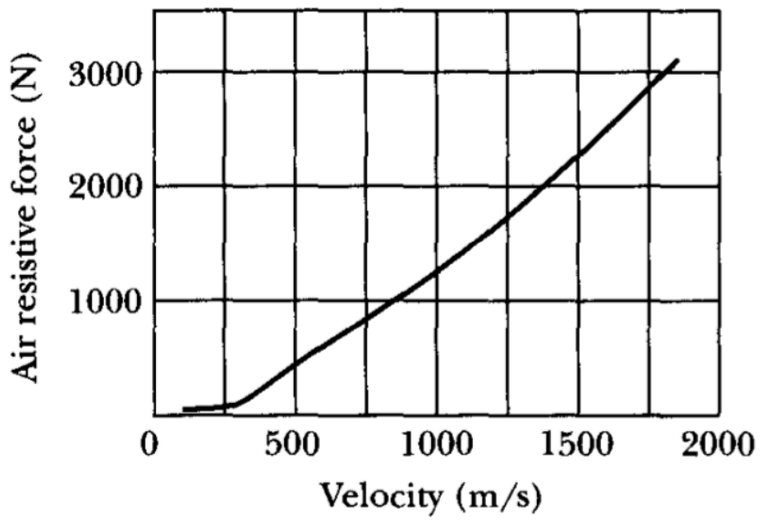
(a)



(b)



(c)



(d)

FIGURE 2-3 (a) Aerodynamic forces acting on projectile. **W** is the drag (air resistive force) and is opposite the velocity of the projectile **v**. Notice that **v** may be at an angle α from the symmetry axis of projectile. The component of force acting perpendicular to the drag is called the lift $\mathbf{L}_a$. The point D is the center of pressure. Finally, the gravitational force $\mathbf{F}_g$ acts down. If the center of pressure is not at the projectile's center of mass, there is also a torque about the center of mass. (b) The drag coefficient c_W, from the Rheinmetall resistance law (Rh82), is plotted versus the Mach number M. Notice the large change near the speed of sound where M = 1. (c) The air resistive force W (drag) is shown as a function of velocity for a projectile diameter of 10 cm. Notice the inflection near the speed of sound. (d) Same as (c) for higher velocities.

Several examples of the motion of a particle subjected to various forces are given below. These examples are particularly good to begin computer calculations using any of the available commercial math programs and spreadsheets or for the students to write their own programs. The computer results, especially the plots, can often be compared with the analytical results presented here. Some of the figures shown in this section were produced using a computer.

EXAMPLE 2.4

As the simplest example of the resisted motion of a particle, find the displacement and velocity of horizontal motion in a medium in which the retarding force is proportional to the velocity.

Solution. A sketch of the problem is shown in Figure 2-4. The Newtonian equation F = ma provides us with the equation of motion:

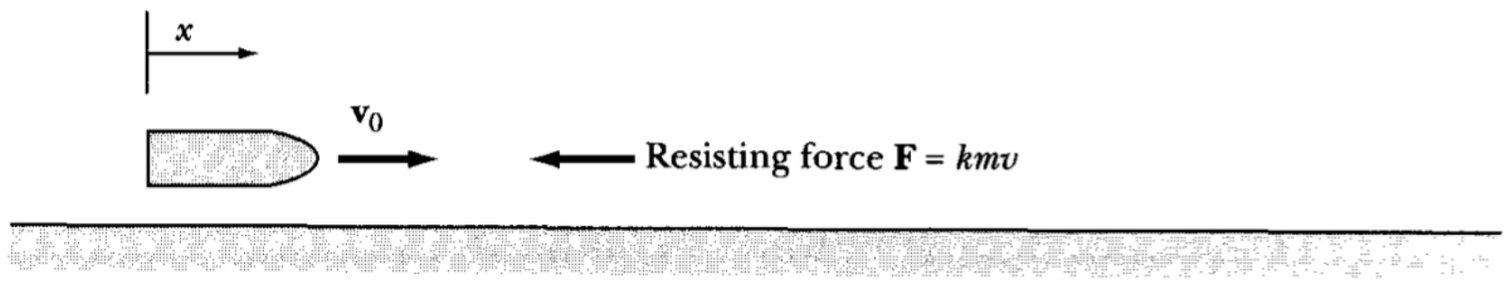


FIGURE 2-4 Example 2.4.

x-direction

$$ma = m\frac{dv}{dt} = -kmv \tag{2.22}$$

where kmv is, the magnitude of the resisting force (k = constant). We are not implying by this form that the retarding force depends on the mass m: this form simply makes the math easier Then

$$\int \frac{dv}{v} = -k\int dt \tag{2.23}$$

$$\ln v = -kt + C_1$$

The integration constant in Equation 2.23 can be evaluated if we prescribe the initial condition $v(t = 0) = v_0$. Then $C_1 = \ln v_0$, and

$$v = v_0 e^{-kt} \tag{2.24}$$

We can integrate this equation to obtain the displacement x as a function of time:

$$v = \frac{dx}{dt} = v_0 e^{-kt}$$

$$x = v_0 \int e^{-kt} dt = -\frac{v_0}{k} e^{-kt} + C_2 \tag{2.25a}$$

The initial condition $x(t = 0) = 0$ implies $C_2 = v_0/k$. Therefore

$$x = \frac{v_0}{k}(1 - e^{-kt}) \tag{2.25b}$$

This result shows that x asymptotically approaches the value v_0/k as $t \to \infty$.

We can also obtain the velocity as a function of displacement by writing

$$\frac{dv}{dx} = \frac{dv}{dt}\frac{dt}{dx} = \frac{dv}{dt} \cdot \frac{1}{v}$$

so that

$$v\frac{dv}{dx} = \frac{dv}{dt} = -kv$$

or

$$\frac{dv}{dx} = -k$$

from which we find, by using the same initial conditions,

$$v = v_0 - kx \tag{2.26}$$

Therefore, the velocity decreases linearly with displacement.

EXAMPLE 2.5

Find the displacement and velocity of a particle undergoing vertical motion in a medium having a retarding force proportional to the velocity.

Solution. Let us consider that the particle is falling downward with an initial velocity v_0 from a height h in a constant gravitational field (Figure 2-5).

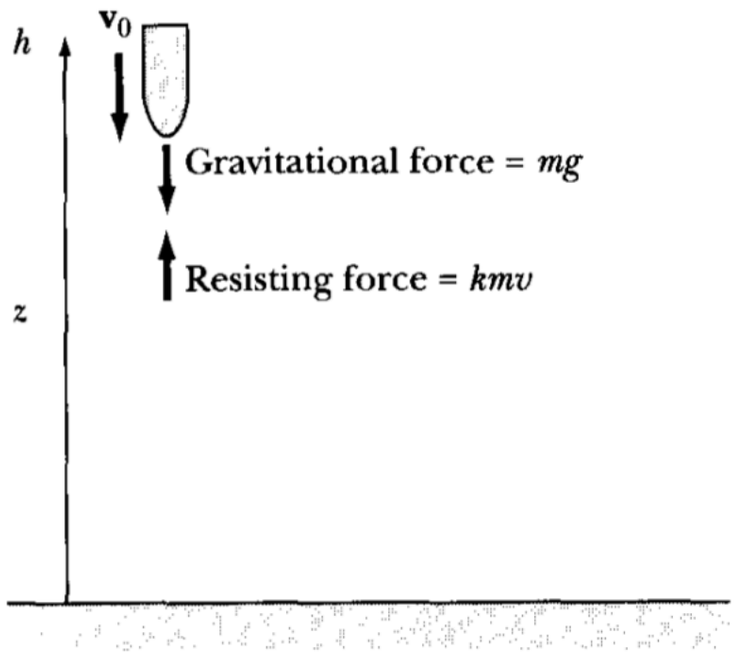


FIGURE 2-5 Example 2.5.

The equation of motion is

z-direction

$$F = m\frac{dv}{dt} = -mg - kmv \tag{2.27}$$

where -kmv represents a positive *upward* force since we take z and $v = \dot{z}$ to be positive upward, and the motion is downward - that is, $v < 0$, so that $-kmv > 0$. From Equation 2.27, we have

$$\frac{dv}{kv + g} = -dt \tag{2.28}$$

Integrating Equation 2.28 and setting $v(t = 0) = v_0$, we have (noting that $v_0 < 0$)

$$\frac{1}{k}\ln(kv + g) = -t + c$$

$$kv + g = e^{-kt+kc}$$

$$v = \frac{dz}{dt} = -\frac{g}{k} + \frac{kv_0 + g}{k}e^{-kt} \tag{2.29}$$

Integrating once more and evaluating the constant by setting $z(t = 0) = h$, we find

$$z = h - \frac{gt}{k} + \frac{kv_0 + g}{k^2}(1 - e^{-kt}) \tag{2.30}$$

Equation 2.29 shows that as the time becomes very long, the velocity approaches the limiting value -g/k; this is called the **terminal velocity**, v_t. Equation 2.27 yields the same result, because the force will vanish—and hence no further acceleration will occur - when $v = -g/k$. If the initial velocity exceeds the terminal velocity in magnitude, then the body immediately begins to slow down and v approaches the terminal speed from the opposite direction. Figure 2-6 illustrates these results for the downward speeds (positive values).

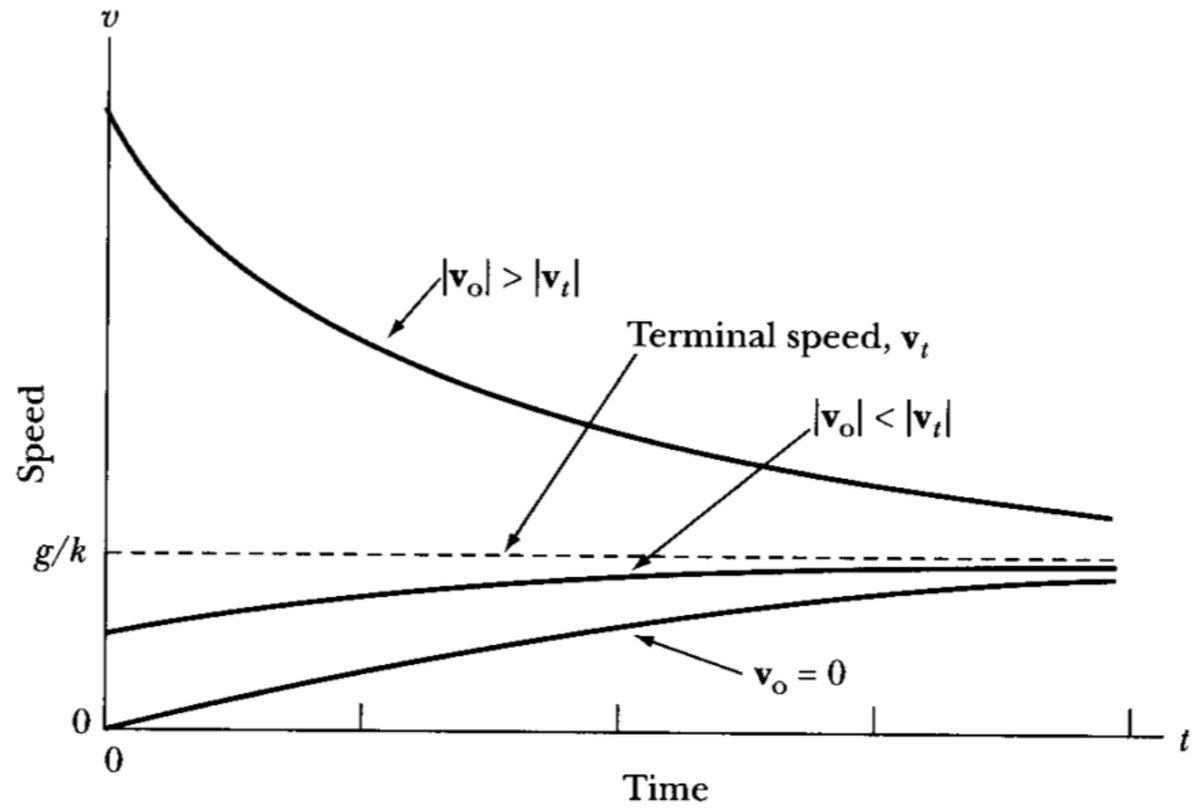


FIGURE 2-6 Results for Example 2.5 indicating the downward speeds for various initial speeds $\mathbf{v}_0$ as they approach the terminal velocity.

EXAMPLE 2.6

Next, we treat projectile motion in two dimensions, first without considering air resistance. Let the muzzle velocity of the projectile be v_0 and the angle of elevation be θ (Figure 2-7).

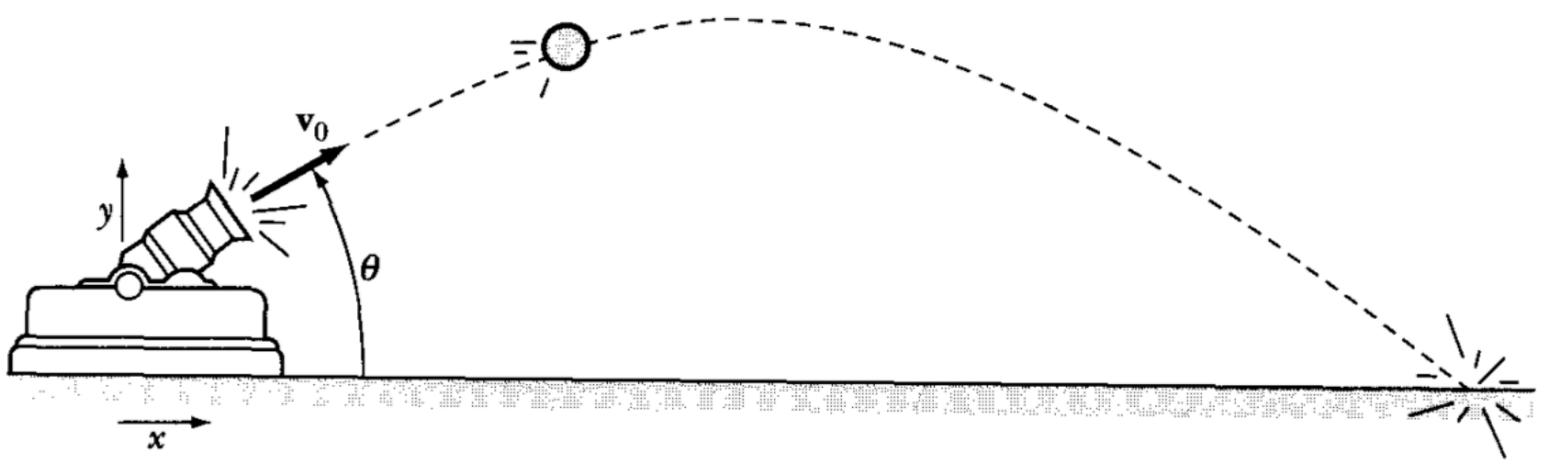


FIGURE 2-7 Example 2.6.

Calculate the projectile's displacement, velocity, and range.

Solution. Using **F** = m**g**, the force components become

x-direction

$$0 = m\ddot{x} \tag{2.31a}$$

y-direction

$$-mg = m\ddot{y} \tag{2.31b}$$

Neglect the height of the gun, and assume x = y = 0 at t = 0. Then

$$\ddot{x} = 0$$

$$\dot{x} = v_0 \cos\theta \tag{2.32}$$

and

$$\ddot{y} = -g$$

$$\dot{y} = -gt + v_0 \sin\theta$$

$$y = \frac{-gt^2}{2} + v_0 t \sin\theta \tag{2.33}$$

The speed and total displacement as functions of time are found to be

$$v = \sqrt{\dot{x}^2 + \dot{y}^2} = (v_0^2 + g^2t^2 - 2v_0 gt \sin\theta)^{1/2} \tag{2.34}$$

and

$$r = \sqrt{x^2 + y^2} = \left(v_0^2 t^2 + \frac{g^2t^2}{4} - v_0 gt^3 \sin\theta\right)^{1/2} \tag{2.35}$$

We can find the range by determining the value of x when the projectile falls back to ground, that is, when y = 0.

$$y = t\left(\frac{-gt}{2} + v_0 \sin\theta\right) = 0 \tag{2.36}$$

One value of y = 0 occurs for t = 0 and the other one for t = T.

$$\frac{-gT}{2} + v_0 \sin\theta = 0$$

$$T = \frac{2v_0 \sin\theta}{g} \tag{2.37}$$

The range R is found from

$$x(t = T) = \text{range} = \frac{2v_0^2}{g} \sin\theta \cos\theta \tag{2.38}$$

$$R = \text{range} = \frac{v_0^2}{g} \sin 2\theta \tag{2.39}$$

Notice that the maximum range occurs for $\theta = 45°$.

Let us use some actual numbers in these calculations. The Germans used a long-range gun named Big Bertha in World War I to bombard Paris. Its muzzle velocity was 1,450 m/s. Find its predicted range, maximum projectile height, and projectile time of flight if $\theta = 55°$. We have v0 = 1450 m/s and $\theta = 55°$, so the range (from Equation 2.39) becomes

$$R = \frac{(1450 \text{ m/s})^2}{9.8 \text{ m/s}^2}[\sin(110°)] = 202 \text{ km}$$

Big Bertha's actual range was 120 km. The difference is a result of the real effect of air resistance.

To find the maximum predicted height, we need to calculated y for the T/2 where T is the projectile time of flight:

$$T = \frac{(2)(1450\ \text{m/s})(\sin 55°)}{9.8\ \text{m/s}^2} = 242\ \text{s}$$

$$y_{\max}\left(t = \frac{T}{2}\right) = \frac{-gT^2}{8} + \frac{v_0 T}{2}\sin\theta$$

$$= \frac{-(9.8\ \text{m/s})(242\ \text{s})^2}{8} + \frac{(1450\ \text{m/s})(242\ \text{s})\ \sin(55°)}{2}$$

$$= 72\ \text{km}$$

EXAMPLE 2.7

Next, we add the effect of air resistance to the motion of the projectile in the previous example. Calculate the decrease in range under the assumption that the force caused by air resistance is directly proportional to the projectile's velocity.

Solution. The initial conditions are the same as in the previous example.

$$\left.\begin{aligned} x(t=0) &= 0 = y(t=0) \\ \dot{x}(t=0) &= v_0\cos\theta \equiv U \\ \dot{y}(t=0) &= v_0\sin\theta \equiv V \end{aligned}\right\} \tag{2.40}$$

However, the equations of motion, Equation 2.31, become

$$m\ddot{x} = -km\dot{x} \tag{2.41}$$

$$m\ddot{y} = -km\dot{y} - mg \tag{2.42}$$

Equation 2.41 is exactly that used in Example 2.4. The solution is therefore

$$x = \frac{U}{k}(1 - e^{-kt}) \tag{2.43}$$

Similarly, Equation 2.42 is the same as the equation of the motion in Example 2.5. We can use the solution found in that example by letting h = 0. (The fact that we considered the particle to be projected downward in Example 2.5 is of no consequence. The sign of the initial velocity automatically takes this into account.) Therefore

$$y = -\frac{gt}{k} + \frac{kV + g}{k^2}(1 - e^{-kt}) \tag{2.44}$$

The trajectory is shown in Figure 2-8 for several values of the retarding force constant k for a given projectile flight.

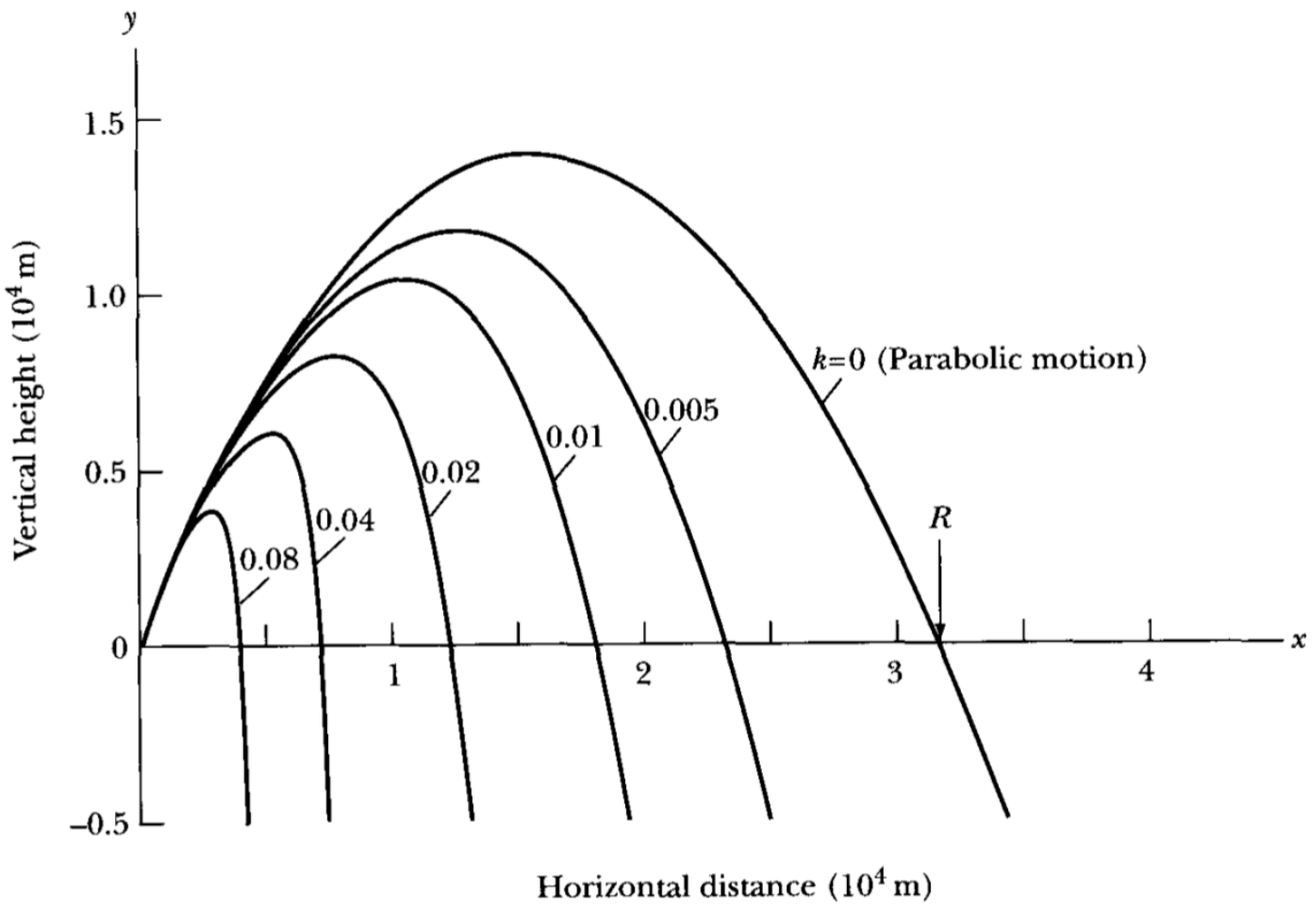


FIGURE 2-8 The calculated trajectories of a particle in air resistance ($F_{res} = -kmv$) for various values of k (in units of s^{-1}). The calculations were performed for values of $\theta = 60°$ and $v_0 = 600$ m/s. The values of y (Equation 2.44) are plotted versus x (Equation 2.43).

The range R', which is the range including air resistance, can be found as previously by calculating the time T required for the entire trajectory and then substituting this value into Equation 2.43 for x. The time T is found as previously by finding t = T when y = 0. From Equation 2.44, we find

$$T = \frac{kV + g}{gk}(1 - e^{-kT}) \tag{2.45}$$

This is a transcendental equation, and therefore we cannot obtain an analytic expression for T. Nonetheless, we still have powerful methods to use to solve such problems. We present two of them here: (1) a *perturbation method* to find an approximate solution, and (2) a numerical method, which can normally be as accurate as desired. We will compare the results.

Perturbation Method To use the perturbation method, we find an expansion parameter or coupling constant that is normally small. In the present case, this parameter is the retarding force constant k, because we have already solved the present problem with k = 0, and now we would like to turn on the retarding force, but let k be small. We therefore expand the exponential term of Equation 2.45 (see Equation D.34 of Appendix D) in a power series with the intention of keeping only the lowest terms of k, where k is our expansion parameter.

$$T = \frac{kV + g}{gk}\left(kT - \frac{1}{2}k^2T^2 + \frac{1}{6}k^3T^3 - \cdots\right) \tag{2.46}$$

If we keep only terms in the expansion through k^2, this equation can be rearranged to yield

$$T = \frac{2V/g}{1 + kV/g} + \frac{1}{3}kT^2 \tag{2.47}$$

We now have the expansion parameter k in the denominator of the first term on the right-hand side of this equation. We need to expand this term in a power series (Taylor series, see Equation D.8 of Appendix D):

$$\frac{1}{1 + kV/g} = 1 - (kV/g) + (kV/g)^2 - \cdots \tag{2.48}$$

where we have kept only terms through k^2, because we only have terms through k in Equation 2.47. If we insert this expansion of Equation 2.48 into the first term on the right-hand side of Equation 2.47 and keep only the terms in k to first order, we have

$$T = \frac{2V}{g} + \left(\frac{T^2}{3} - \frac{2V^2}{g^2}\right)k + O(k^2) \tag{2.49}$$

where we choose to neglect $O(k^2)$, the terms of order k^2 and higher In the limit $k \to 0$ (no air resistance). Equation 2.49 gives us the same result as in the previous example:

$$T(k = 0) = T_0 = \frac{2V}{g} = \frac{2v_0 \sin\theta}{g}$$

Therefore, if k is small (but nonvanishing), the flight time will be *approximately* equal to T_0. If we then use this approximate value for $T = T_0$ in the right-hand side of Equation 2.49, we have

$$T \cong \frac{2V}{g}\left(1 - \frac{kV}{3g}\right) \tag{2.50}$$

which is the desired approximate expression for the flight time.

Next, we write the equation for x (Equation 2.43) in expanded form:

$$x = \frac{U}{k}\left(kt - \frac{1}{2}k^2t^2 + \frac{1}{6}k^3t^3 - \cdots\right) \tag{2.51}$$

Because x(t= T) = R', we have approximately for the range

$$R' \cong U\left(T - \frac{1}{2}kT^2\right) \tag{2.52}$$

where again we keep terms only through the first order of k. We can now evaluate this expression by using the value of T from Equation 2.50. If we retain only terms linear in k, we find

$$R' \cong \frac{2UV}{g}\left(1 - \frac{4kV}{3g}\right) \tag{2.53}$$

The quantity 2UV/g can now be written (using Equations 2.40) as

$$\frac{2UV}{g} = \frac{2v_0^2}{g}\sin\theta\cos\theta = \frac{v_0^2}{g}\sin 2\theta = R \tag{2.54}$$

which will be recognized as the range R of the projectile when air resistance is neglected. Therefore

$$R' \cong R\left(1 - \frac{4kV}{3g}\right) \tag{2.55}$$

Over what range of values for k would we expect our perturbation method to be correct? If we look at the expansion in Equation 2.48, we see that the expansion will not converge unless $kV/g < 1$ or $k < g/V$, and in fact, we would like $k << g/V = g/v_0\sin\theta$).

Numerical Method Equation 2.45 can be solved numerically using a computer by a variety of methods. We set up a loop to solve the equation for T for many values of k up to 0.08 s^{-1}: $T_i(k_i)$. These values of T_i and k_i, are inserted into Equation 2.43 to find the range R'_i, which is displayed in Figure 2-9. The range drops rapidly for increased air resistance, just as one would expect, but it does not display the linear dependence suggested by the perturbation method solution of Equation 2.55.

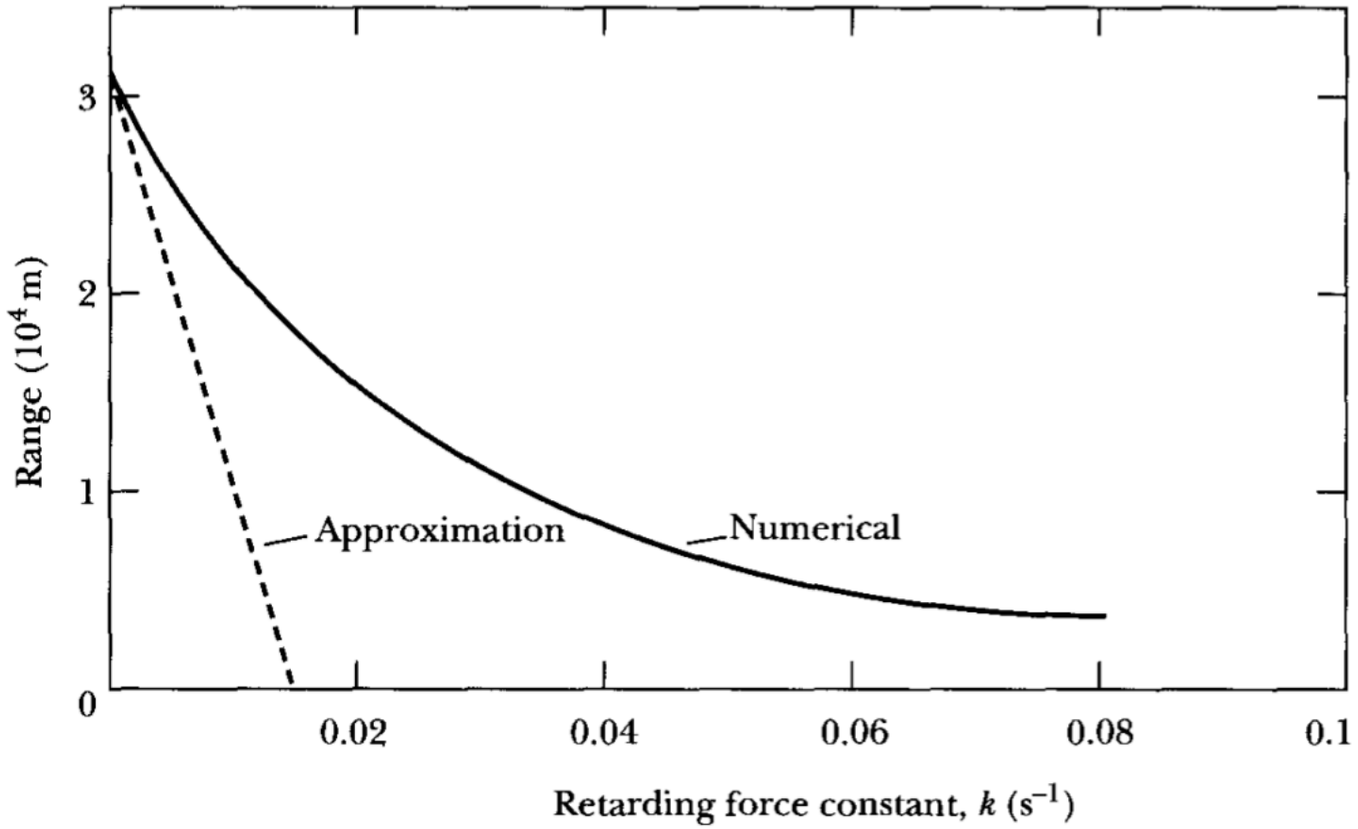


FIGURE 2-9 The range values calculated approximately and numerically for the projectile data given in Figure 2-8 are plotted as a function of the retarding force constant k.

For the projectile motion described in Figures 2-8 and 2-9, the linear approximation is inaccurate for k values as low as 0.01 s^{-1} and incorrectly shows the range is zero for all values of k larger than 0.014 s^{-1} This disagreement with the perturbation method is not surprising because the linear result for the range R' was dependent on $k << g/v_0 \sin\theta) = 0.02$ s^{-1} which is hardly true for even $k = 0.01$ s^{-1} The agreement should be adequate for $k = 0.005$ s^{-1} The results shown in Figure 2-8 indicate that for values of $k > 0.005$ s^{-1} the drag can hardly be considered a perturbation. In fact, for $k > 0.01$ s^{-1} the drag becomes the dominant factor in the projectile motion

The previous example indicates how complicated the real world can be. In that example, we still had to make assumptions that were nonphysical - in assuming, for example, that the retarding force is always linearly proportional to the velocity. Even our numerical calculation is not accurate, because Figure 2-3 shows us that a better assumption would be to include a v^2 retarding term as well.

Adding such a term would not be difficult with the numerical calculation, and we shall do a similar calculation in the next example.

EXAMPLE 2.8

Use the data shown in Figure 2-3 to calculate the trajectory for an actual projectile. Assume a muzzle velocity of 600 m/s, gun elevation of 45°, and a projectile mass of 30 kg. Plot the height y versus the horizontal distance x and plot y, X, and y versus time both with and without air resistance. Include only the air resistance and gravity, and ignore other possible forces such as the lift.

Solution. First, we make a table of retarding force versus velocity by reading Figure 2-3. Read the force every 50 m/s for Figure 2-3c and every 100 m/s for Figure 2-3d. We can then use a straight line interpolation between the tabular values. We use the coordinate system shown in Figure 2-7. The equations of motion become

$$\ddot{x} = -\frac{F_x}{m} \tag{2.56}$$

$$\ddot{y} = -\frac{F_y}{m} - g \tag{2.57}$$

where F_x and F_y are the retarding forces. Assume g is constant. F_x will always be a positive number, but $F_y > 0$ for the projectile going up, and $F_y < 0$ for the projectile coming back down. Let θ be the projectile's elevation angle from the horizontal at any instant.

$$v = \sqrt{\dot{x}^2 + \dot{y}^2} \tag{2.58}$$

$$\tan\theta = \frac{\dot{y}}{\dot{x}} \tag{2.59}$$

$$F_x = F\cos\theta \tag{2.60}$$

$$F_y = F\sin\theta \tag{2.61}$$

We can calculate F_x and F_y at any instant by knowing $\dot{x}$ and $\dot{y}$. Over a small time interval, the next x and y can be calculated.

$$\dot{x} = \int_0^t \ddot{x}\, dt + v_0\cos\theta \tag{2.62}$$

$$\dot{y} = \int_0^t \ddot{y}\, dt + v_0\sin\theta \tag{2.63}$$

$$x = \int_0^t \dot{x}\, dt \tag{2.64}$$

$$y = \int_0^t \dot{y}\, dt \tag{2.65}$$

One can writee a short computer program to contain our table for the retarding forces and to perform the calculations for $\dot{x}$, $\dot{y}$, x, and y) as a function of time. We must perform the integrals by summations over small time intervals, because the forces are time dependent. Figure 2-10 shows the results. Notice the large difference that the air resistance makes. In Figure 2-10a, the horizontal distance (range) that the projectile travels is about 16 km compared to almost 37 km with no air resistance. Our calculation ignored the fact that the air density depends on the altitude. If we take account of the decrease in the air density with altitude, we obtain the third curve with a range of 18 km shown in Figure 2-l0a. If we also included the lift, the range would be still greater Notice that the change in velocities in Figures 2-l0c and 2-l0d mirror the air resistive force of Figure 2-3. The speeds decrease rapidly until the speed reaches the speed of sound, and then the rate of change of the speeds levels off somewhat.

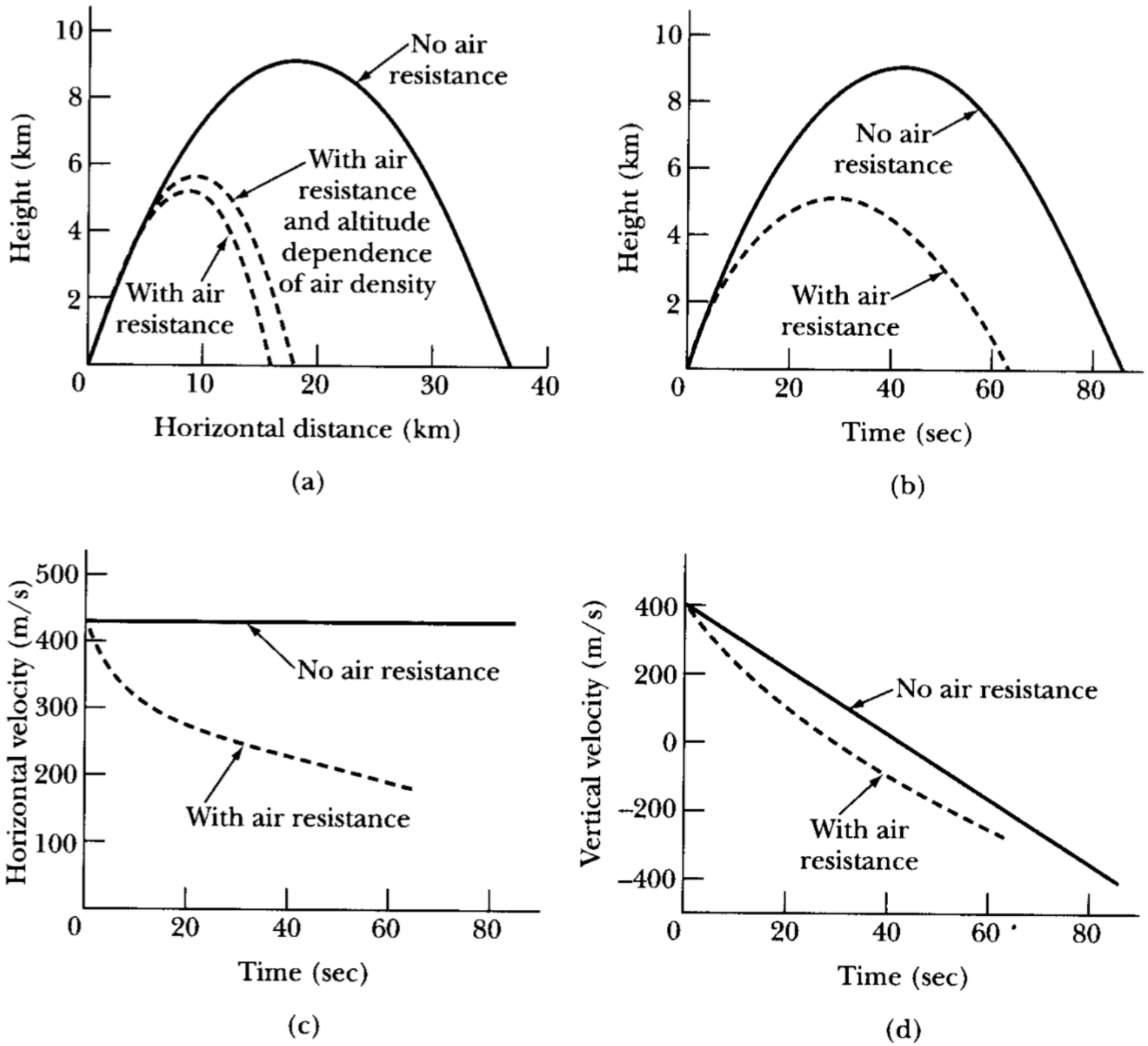


FIGURE 2-10 The results of Example 2.8. The solid lines are the results if no air resistance is included, whereas the dashed lines include the results of adding the air resistive force. In (a) we also include the effect of the air density dependence, which becomes smaller as the projectile rises higher.

This concludes our subsection on the effects of retarding forces. Much more could be done to include realistic effects, but the method is clear Normally, one effect is added at a time, and the results are analyzed before another effect is added.

Other Examples of Dynamics

We conclude this section with two additional standard examples of dynamical particle-like behavior

EXAMPLE 2.9

Atwood's machine consists of a smooth pulley with two masses suspended from a light string at each end (Figure 2-11). Find the acceleration of the masses and the tension of the string (a) when the pulley center is at rest and (b) when the pulley is descending in an elevator with constant acceleration α.

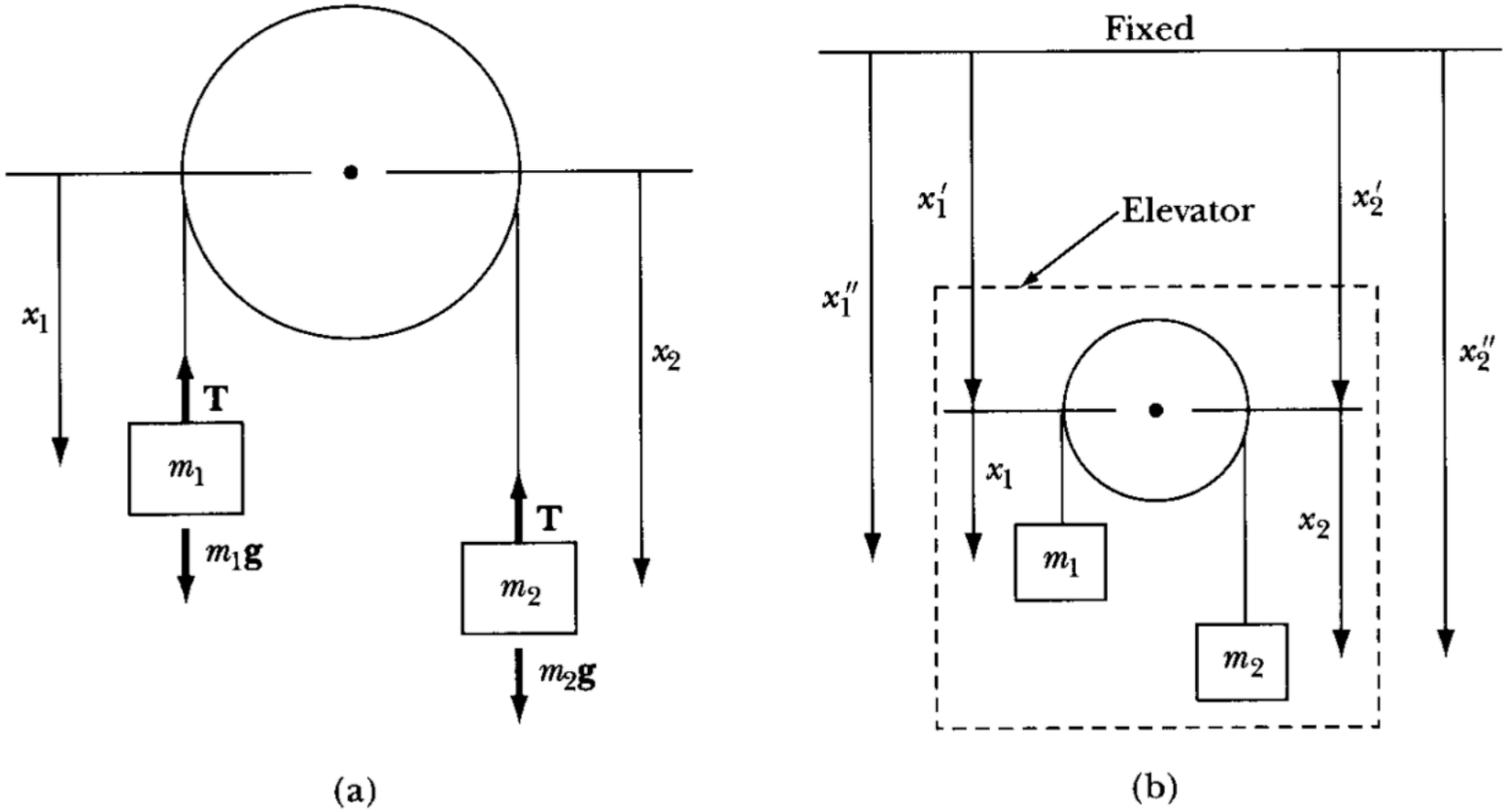


FIGURE 2-11 Example 2.9; Atwood's machine.

Solution. We neglect the mass of the string and assume that the pulley is smooth—that is, no friction on the string. The tension T must be the same throughout the string. The equations of motion become, for each mass, for case (a).

$$m_1 \ddot{x}_1 = m_1 g - T \tag{2.66}$$

$$m_2 \ddot{x} = m_2 g - T \tag{2.67}$$

Notice again the advantage of the force concept: We need only identify the forces acting on each mass. The tension T is the same in both equations. If the string is inextensible, then $\ddot{x}_2 = -\ddot{x}_1$ and Equations 2.66 and 2.67 may be combined

$$\begin{aligned} m_1\ddot{x}_1 &= m_1 g - (m_2 g - m_2\ddot{x}_2) \\ &= m_1 g - (m_2 g + m_2\ddot{x}_1) \end{aligned}$$

Rearranging,

$$\ddot{x}_1 = \frac{g(m_1 - m_2)}{m_1 + m_2} = -\ddot{x}_2 \tag{2.68}$$

If $m_1 > m_2$, then $\ddot{x}_1 > 0$, and $\ddot{x}_2 < 0$. The tension can be obtained from Equations 2.68 and 2.66:

$$T = m_1 g - m_1\ddot{x}_1$$

$$T = m_1 g - m_1 g\frac{(m_1 - m_2)}{m_1 + m_2}$$

$$T = \frac{2m_1 m_2 g}{m_1 + m_2} \tag{2.69}$$

For case (b), in which the pulley is in an elevator, the coordinate system with origins at the pulley center is no longer an inertial system. We need an inertial system with the origin at the top of the elevator shaft (Figure 2-llb). The equations of motion in the inertial system ($x_1'' = x_1' + x_1, x_2'' = x_2' + x_2$) are

$$\begin{aligned} m_1\ddot{x}_1'' &= m_1(\ddot{x}_1' + \ddot{x}_1) = m_1 g - T \\ m_2\ddot{x}_2'' &= m_2(\ddot{x}_2' + \ddot{x}_2) = m_2 g - T \end{aligned}$$

so

$$\left.\begin{aligned} m_1\ddot{x}_1 &= m_1 g - T - m_1\ddot{x}_1' = m_1(g-\alpha) - T \\ m_2\ddot{x}_2 &= m_2 g - T - m_2\ddot{x}_2' = m_2(g-\alpha) - T \end{aligned}\right\} \tag{2.70}$$

where $\ddot{x}_1' = \ddot{x}_2' = \alpha$. We have $\ddot{x}_2' = -\ddot{x}_1'$, so we solve for $\ddot{x}_1$ as before by eliminating T:

$$\ddot{x}_1 = -\ddot{x}_2 = (g-\alpha)\frac{(m_1 - m_2)}{m_1 + m_2} \tag{2.71}$$

and

$$T = \frac{2m_1 m_2(g-\alpha)}{m_1 + m_2} \tag{2.72}$$

Notice that the results for the acceleration and tension are just as if the acceleration of gravity were reduced by the amount of the elevator acceleration α . The change for an ascending elevator should be obvious.

EXAMPLE 2.10

In our last example in this lengthy review of the equations of motion for a particle, let us examine particle motion in an electromagnetic field. Consider a charged particle entering a region of uniform magnetic field **B** - for example, the earth's field - as shown in Figure 2-12. Determine its subsequent motion.

Solution. Choose a Cartesian coordinate system with its y-axis parallel to the magnetic field. If q is the charge on the particle, **v** its velocity, **a** its acceleration, and **B** the earth's magnetic field, then

$$\mathbf{v} = \dot{x}\mathbf{i} + \dot{y}\mathbf{j} + \dot{z}\mathbf{k}$$

$$\mathbf{a} = \ddot{x}\mathbf{i} + \ddot{y}\mathbf{j} + \ddot{z}\mathbf{k}$$

$$\mathbf{B} = B_0\mathbf{j}$$

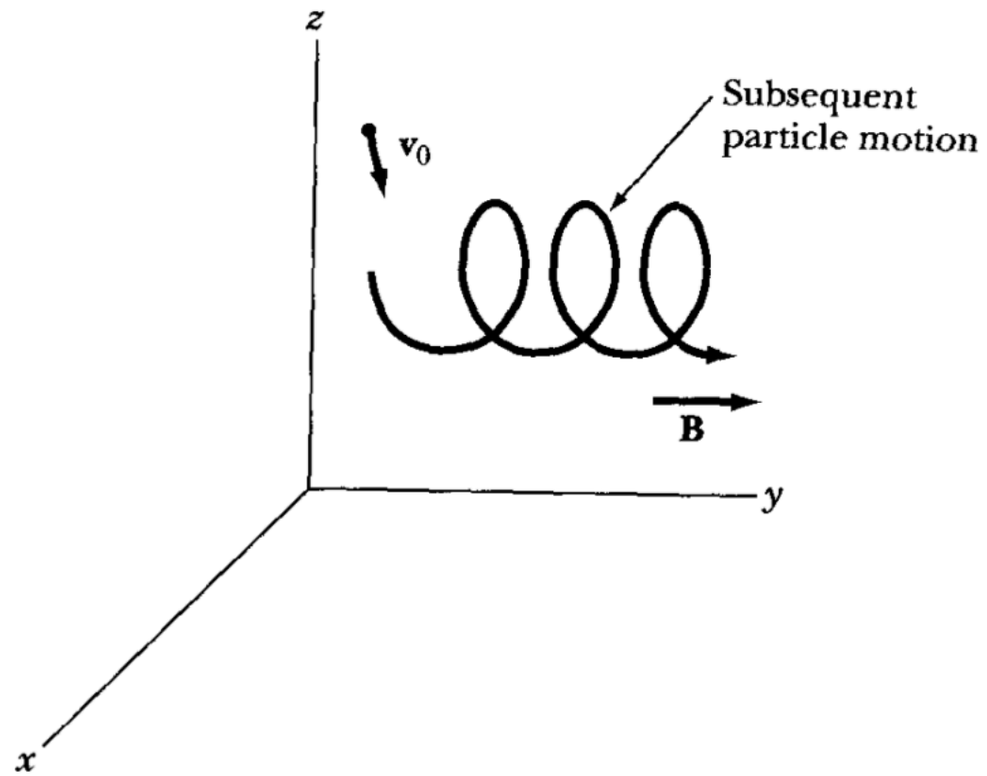


FIGURE 2-12 Example 2.10; a moving particle enters a region of magnetic field.

The magnetic force $\mathbf{F} = q\mathbf{v} \times \mathbf{B} = m\mathbf{a}$, so

$$m(\ddot{x}\mathbf{i} + \ddot{y}\mathbf{j} + \ddot{z}\mathbf{k}) = q(\dot{x}\mathbf{i} + \dot{y}\mathbf{j} + \dot{z}\mathbf{k}) \times B_0\mathbf{j} = qB_0(\dot{x}\mathbf{k} - \dot{z}\mathbf{i})$$

Equating like vector components give

$$\left.\begin{aligned} m\ddot{x} &= -qB_0\dot{z} \\ m\ddot{y} &= 0 \\ m\ddot{z} &= qB_0\dot{x} \end{aligned}\right\} \tag{2.73}$$

Integrating the second of these equations $m\ddot{y} = 0$, yields

$$\dot{y} = \dot{y}_0$$

where $\dot{y}_0$ is a constant and is the initial value of $\dot{y}$. Integrating a second time gives

$$y = \dot{y}_0 t + y_0$$

where y_0 is also a constant.

To integrate the first and last equations of Equation 2.73, let $\alpha = qB_0/m$, so that

$$\left.\begin{aligned} \ddot{x} &= -\alpha \dot{z} \\ \ddot{z} &= \alpha \dot{x} \end{aligned}\right\} \tag{2.74}$$

These coupled, simultaneous differential equations can be easily uncoupled by differentiating one and substituting it into the other, giving

$$\dddot{z} = \alpha \ddot{x} = -\alpha^2 \dot{z}$$

$$\dddot{x} = -\alpha \ddot{z} = -\alpha^2 \dot{x}$$

so that

$$\left.\begin{aligned} \dddot{z} &= -\alpha^2 \dot{z} \\ \dddot{x} &= -\alpha^2 \dot{x} \end{aligned}\right\} \tag{2.75}$$

Both of these differential equations have the same form of solution. Using the technique of Example C.2 of Appendix C, we have

$$x = A \cos \alpha t + B \sin \alpha t + x_0$$

$$z = A' \cos \alpha t + B' \sin \alpha t + z_0$$

where A, A', B, B', x_0, and z_0 are constants of integration that are determined by the particle's initial position and velocity and by the equations of motion, Equation 2.74. These solutions can be rewritten

$$\left.\begin{aligned}(x - x_0) &= A\cos\alpha t + B\sin\alpha t\\ (y - y_0) &= \dot{y}_0 t\\ (z - z_0) &= A'\cos\alpha t + B'\sin\alpha t\end{aligned}\right\} \tag{2.76}$$

The x- and z-coordinates are connected by Equation 2.74, so substituting Equations 2.76 into the first equation of Equation 2.74 gives

$$-\alpha^2 A\cos\alpha t - \alpha^2 B\sin\alpha t = -\alpha(-\alpha A'\sin\alpha t + \alpha B'\cos\alpha t) \tag{2.77}$$

Because Equation 2.77 is valid for all t, in particular t = 0 and t = π/2a, Equation 2.77 yields

$$-\alpha^2 A = -\alpha^2 B'$$

so that

$$A = B'$$

and

$$-\alpha^2 B = \alpha^2 A'$$

gives

$$B = -A'$$

We now have

$$\left.\begin{aligned}(x - x_0) &= A\cos\alpha t + B\sin\alpha t\\ (y - y_0) &= \dot{y}_0 t\\ (z - z_0) &= -B\cos\alpha t + A\sin\alpha t\end{aligned}\right\} \tag{2.78}$$

If at t = 0, $\dot{z} = \dot{z}_0$ and $\dot{x} = 0$, then from Equation 2.78, differentiating and setting t = 0 gives

$$\alpha B = 0$$

and

$$\alpha A = \dot{z}_0$$

so

$$(x - x_0) = \frac{\dot{z}_0}{\alpha} \cos \alpha t$$

$$(y - y_0) = \dot{y}_0 t$$

$$(z - z_0) = \frac{\dot{z}_0}{\alpha} \sin \alpha t$$

Finally,

$$\left.\begin{aligned} x - x_0 &= \left(\frac{\dot{z}_0 m}{qB_0}\right)\cos\left(\frac{qB_0 t}{m}\right) \\ (y - y_0) &= \dot{y}_0 t \\ (z - z_0) &= \left(\frac{\dot{z}_0 m}{qB_0}\right)\sin\left(\frac{qB_0 t}{m}\right) \end{aligned}\right\} \tag{2.79}$$

These are the parametric equations of a circular helix of radius $\dot{z}_0 m/qB_0$. Thus, the faster the particle enters the field or the greater its mass, the larger the radius of the helix. And the greater the charge on the particle or the stronger the magnetic field, the tighter the helix. Notice also how the charged particle is captured by the magnetic field—just drifting along the field direction. In this example, the particle had no initial component of its velocity along the x-axis, but even if it had it would not drift along this axis. Finally, notice that the magnetic force on the particle always acts perpendicular to its velocity and hence cannot speed it up. Equation 2.79 verifies this fact. The earth's magnetic field is not as simple as the uniform field of this example. Nevertheless, this example gives some insight into one of the mechanisms by which the earth's magnetic field traps low-energy cosmic rays and the solar wind to create the Van Allen belts

2.5 Conservation Theorems

We now turn to a detailed discussion of the Newtonian mechanics of a single particle and derive the important theorems regarding conserved quantities. We must emphasize that we are not proving the conservation of the various quantities. We are merely deriving the consequences of Newton's laws of dynamics. These implications must be put to the test of experiment, and their verification then supplies a measure of confirmation of the original dynamical laws. The fact that these conservation theorems have indeed been found to be valid in many instances furnishes an important part of the proof for the correctness of Newton's laws, at least in classical physics.

The first of the conservation theorems concerns the linear momentum of a particle. If the particle is free, that is, if the particle encounters no force, then Equation 2.2 becomes simply $\dot{\mathbf{p}} = 0$. Therefore, **p** is a vector constant in time, and the first conservation theorem becomes

I. *The total linear momentum* **p** *of a particle is conserved when the total force on it is zero*.

Note that this result is derived from a vector equation, $\dot{\mathbf{p}} = 0$, and therefore applies for each component of the linear momentum. To state the result in other terms, we let **s** be some constant vector such that $\mathbf{F} \cdot \mathbf{s} = 0$, independent of time. Then

$$\dot{\mathbf{p}} \cdot \mathbf{s} = \mathbf{F} \cdot \mathbf{s} = 0$$

or, integrating with respect to time,

$$\mathbf{p} \cdot \mathbf{s} = \text{constant} \tag{2.80}$$

which states that the *component of linear momentum in a direction in which the force vanishes is constant in time*.

The angular momentum **L** of a particle with respect to an origin from which the position vector **r** is measured is defined to be

$$\boxed{\mathbf{L} \equiv \mathbf{r} \times \mathbf{p}} \tag{2.81}$$

The torque or moment of force **N** with respect to the same origin is defined to be

$$\boxed{\mathbf{N} \equiv \mathbf{r} \times \mathbf{F}} \tag{2.82}$$

where **r** is the position vector from the origin to the point where the force **F** is applied. Because $\mathbf{F} = m\dot{\mathbf{v}}$ for the particle, the torque becomes

$$\mathbf{N} = \mathbf{r} \times m\dot{\mathbf{v}} = \mathbf{r} \times \dot{\mathbf{p}}$$

Now

$$\dot{\mathbf{L}} = \frac{d}{dt}(\mathbf{r} \times \mathbf{p}) = (\dot{\mathbf{r}} \times \mathbf{p}) + (\mathbf{r} \times \dot{\mathbf{p}})$$

but

$$\dot{\mathbf{r}} \times \mathbf{p} = \dot{\mathbf{r}} \times m\mathbf{v} = m(\dot{\mathbf{r}} \times \dot{\mathbf{r}}) \equiv 0$$

so

$$\boxed{\dot{\mathbf{L}} = \mathbf{r} \times \dot{\mathbf{p}} = \mathbf{N}} \tag{2.83}$$

If no torques act on a particle (i.e., if $\mathbf{N} = 0$), then $\dot{\mathbf{L}} = 0$ and **L** is a vector constant in time. The second important conservation theorem is

II. *The angular momentum of a particle subject to no torque is conserved.*

We remind the student that a judicious choice of the origin of a coordinate system will often allow a problem to be solved much more easily than a poor choice.

For example, the torque will be zero in coordinate systems centered along the resultant line of force. The angular momentum will be conserved in this case.

If work is done on a particle by a force **F** in transforming the particle from Condition 1 to Condition 2, then this **work** is defined to be

$$\boxed{W_{12} \equiv \int_1^2 \mathbf{F} \cdot d\mathbf{r}} \tag{2.84}$$

If **F** is the net resultant force acting on the particle,

$$\mathbf{F} \cdot d\mathbf{r} = m \frac{d\mathbf{v}}{dt} \cdot \frac{d\mathbf{r}}{dt} dt = m \frac{d\mathbf{v}}{dt} \cdot \mathbf{v} dt$$

$$= \frac{m}{2} \frac{d}{dt} (\mathbf{v} \cdot \mathbf{v})\, dt = \frac{m}{2} \frac{d}{dt} (v^2)\, dt = d\left(\frac{1}{2} m v^2\right) \tag{2.85}$$

The integrand in Equation 2.84 is thus an exact differential, and the work done by the total force **F** acting on a particle is equal to its change in kinetic energy:

$$W_{12} = \left(\frac{1}{2} m v^2\right)\Big|_1^2 = \frac{1}{2} m (v_2^2 - v_1^2) = T_2 - T_1 \tag{2.86}$$

where $T = \frac{1}{2} mv^2$ is the kinetic energy of the particle. If $T_1 > T_2$ then $W_{12} < 0$, and the particle has done work with a resulting decrease in kinetic energy. It is important to realize that the force **F** leading to Equation 2.85 is the total (i.e., net resultant) **force** on the particle.

Let us now examine the integral appearing in Equation 2.84 from a different standpoint. In many physical problems, the force **F** has the property that the work required to move a particle from one position to another without any change in kinetic energy depends only on the original and final positions and not on the exact path taken by the particle. For example, assume the work done to move the particle from point 1 in Figure 2-13 to point 2 is independent of the actual paths a, b, or c taken.

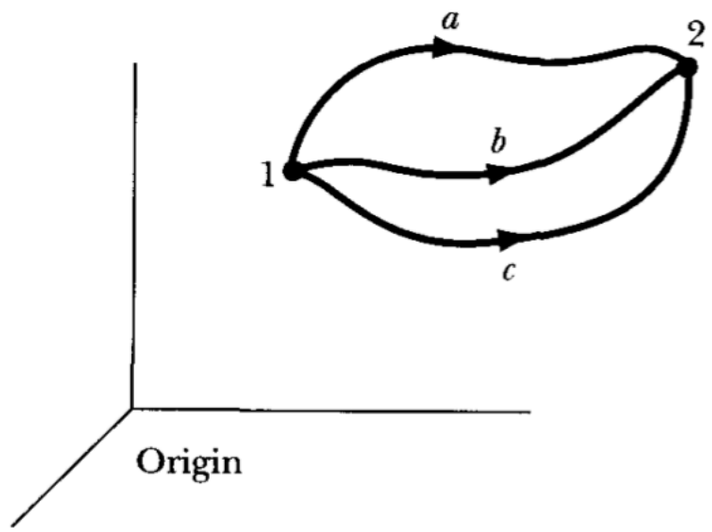


FIGURE 2-13 For some forces (identified later as *conservative*), the work done by the force to move a particle from one position 1 to another position 2 is independent of the path (*a*, *b*, or *c*).

This property is exhibited, for example, by a constant gravitational force field. Thus, if a particle of mass m is raised through a height h (by any path), then an amount of work mgh has been done on the particle, and the particle can do an equal amount of work in returning to its original position. This capacity to do work is called the **potential energy** of the particle.

We may define the potential energy of a particle in terms of the work (done by the force **F**) required to transport the particle from a point 1 to a point 2 (with no net change in kinetic energy):

$$\int_1^2 \mathbf{F} \cdot d\mathbf{r} \equiv U_1 - U_2 \tag{2.87}$$

The work done in moving the particle is thus simply the difference in the potential energy U at the two points. For example, if we lift a suitcase from position 1 on the ground to position 2 in a car trunk, we as the external agent are doing work against the force of gravity. Let the force F in Equation 2.87 be the gravitational force, and in raising the suitcase, $\mathbf{F} \cdot d\mathbf{r}$ becomes negative. The result of the integration in Equation 2.87 is that U_1-U_2 is negative, so that the potential energy at position 2 in the car's trunk is greater than that at position 1 on the ground. The change in potential energy U_2-U_1 is the negative of the work done by the gravitational force, as can be seen by multiplying both sides of Equation 2.87 by -1. As the external agent, we do positive work (against gravity) to raise the potential energy of the suitcase.

Equation 2.87 can be reproduced (The necessary and sufficient condition that permits a vector function to be represented by the gradient of a scalar function is that the curl of the vector function vanishes identically.) if we write **F** as the gradient of the scalar function U:

$$\boxed{\mathbf{F} = -\mathbf{grad}\, U = -\nabla U} \tag{2.88}$$

Then

$$\int_1^2 \mathbf{F} \cdot d\mathbf{r} = -\int_1^2 (\nabla U) \cdot d\mathbf{r} = -\int_1^2 dU = U_1 - U_2 \tag{2.89}$$

In most systems of interest, the potential energy is a function of position and, possibly, time: U= U(r) or U = U(r, t). We do not consider cases in which the potential energy is a function of the velocity.

It is important to realize that the potential energy is defined only to within an additive constant; that is, the force defined by $-\nabla U$ is no different from that defined by $-\nabla(U + \text{ constant})$. Potential energy therefore has no absolute meaning; only differences of potential energy are physically meaningful (as in Equation 2.87).

If we choose a certain inertial frame of reference to describe a mechanical process, the laws of motion are the same as in any other reference frame in uniform motion relative to the original frame. The velocity of a particle is in general different depending on which inertial reference frame we chose as the basis for describing the motion. We therefore find that it is impossible to ascribe an absolute kinetic energy to a particle in much the same way that it is impossible to assign any absolute meaning to potential energy. Both of these limitations are the result of the fact that selecting an origin of the coordinate system used to describe physical processes is always arbitrary. The nineteenth-century Scottish physicist James Clerk Maxwell A831—1879) summarized the situation as follows.

> *We must, therefore, regard the energy of a material system as a quantity of which we may ascertain the increase or diminution as the system passes from one definite condition to another. The absolute value of the energy in the standard condition is unknown to us, and it would be of no value to us if we did know it, as all phenomena depend on the variations of energy and not on its absolute value.*

Next, we define the **total energy** of a particle to be the sum of the kinetic and potential energies:

$$\boxed{E \equiv T + U} \tag{2.90}$$

The total time derivative of E is

$$\frac{dE}{dt} = \frac{dT}{dt} + \frac{dU}{dt} \tag{2.91}$$

To evaluate the time derivatives appearing on the right-hand side of this equation, we first note that Equation 2.85 can be written as

$$\mathbf{F} \cdot d\mathbf{r} = d\left(\frac{1}{2}mv^2\right) = dT \tag{2.92}$$

Dividing through by dt,

$$\frac{dT}{dt} = \mathbf{F} \cdot \frac{d\mathbf{r}}{dt} = \mathbf{F} \cdot \dot{\mathbf{r}} \tag{2.93}$$

We have also

$$\begin{aligned} \frac{dU}{dt} &= \sum_i \frac{\partial U}{\partial x_i} \frac{dx_i}{dt} + \frac{\partial U}{\partial t} \\ &= \sum_i \frac{\partial U}{\partial x_i} \dot{x}_i + \frac{\partial U}{\partial t} \\ &= (\nabla U) \cdot \dot{\mathbf{r}} + \frac{\partial U}{\partial t} \end{aligned} \tag{2.94}$$

Substituting Equations 2.93 and 2.94 into 2.91, we find

$$\begin{aligned} \frac{dE}{dt} &= \mathbf{F} \cdot \dot{\mathbf{r}} + (\nabla U) \cdot \dot{\mathbf{r}} + \frac{\partial U}{\partial t} \\ &= (\mathbf{F} + \nabla U) \cdot \dot{\mathbf{r}} + \frac{\partial U}{\partial t} \end{aligned}$$

$$p = \frac{\partial U}{\partial t} \tag{2.95}$$

because the term $\mathbf{F} + \nabla U$ vanishes in view of the definition of the potential energy (Equation 2.88) if the total force is the conservative force $\mathbf{F} = -\nabla U$.

If U is not an explicit function of the time (i.e., if $\partial U/\partial t = 0$; recall that we do not consider velocity-dependent potentials), the force field represented by **F** is conservative. Under these conditions, we have the third important conservation theorem:

III. *The total energy E of a particle in a conservative force field is a constant in time.*

It must be reiterated that we have not proved the conservation laws of linear momentum, angular momentum, and energy. We have only derived various consequences of Newton's laws; that is, if these laws are valid in a certain situation, then momentum and energy will be conserved. But we have become so enamored with these conservation theorems that we have elevated them to the status of laws and we have come to insist that they be valid in any physical theory, even those that apply to situations in which Newtonian mechanics is not valid, as, for example, in the interaction of moving charges or in quantum-mechanical systems. We do not actually have conservation laws in such situations, but rather conservation *postulates* that we force on the theory. For example, if we have two isolated moving electric charges, the electromagnetic forces between them are not conservative. We therefore endow the electromagnetic field with a certain amount of energy so that energy conservation will be valid. This procedure is satisfactory only if the consequences do not contradict any experimental fact, and this is indeed the case for moving charges. We therefore extend the usual concept of energy to include "electromagnetic energy" to satisfy our preconceived notion that energy must be conserved. This may seem an arbitrary and drastic step to take, but nothing, it is said, succeeds as does success, and these conservation "laws" have been the most successful set of principles in physics. The refusal to relinquish energy and momentum conservation led Wolfgang Pauli to postulate in 1930 the existence of the neutrino to account for the "missing" energy and momentum in radioactive beta decay. This postulate allowed Enrico Fermi to construct a successful theory of beta decay in 1934, but direct observation of the neutrino was not made until 1953 when Reines and Cowan performed their famous experiment. By adhering to the conviction that energy and momentum must be conserved, a new elementary particle was discovered, one that is of great importance in modern theories of nuclear and particle physics. This discovery is only one of the many advances in the understanding of the properties of matter that have resulted directly from the application of the conservation laws.

We shall apply these conservation theorems to several physical situations in the remainder of this book, among them Rutherford scattering and planetary motion. A simple example here indicates the usefulness of the conservation theorems.

EXAMPLE 2.11

A mouse of mass m jumps on the outside edge of a freely turning ceiling fan of rotational inertia I and radius R. By what ratio does the angular velocity change?

Solution. Angular momentum must be conserved during the process. We are using the concept of rotational inertia learned in elementary physics to relate angular momentum L to angular velocity ω . $L = I\omega$. The initial angular momentum $L_0 = \omega_0$ must be equal to the angular momentum L (fan plus mouse) after the mouse jumps on. The velocity of the outside edge is $v = \omega R$.

$$L = I\omega + mvR = \frac{v}{R}(I + mR^2)$$

$$L = L_0 = I\omega_0$$

$$\frac{v}{R}(I + mR^2) = I\frac{v_0}{R}$$

$$\frac{v}{v_0} = \frac{I}{I + mR^2}$$

and

$$\frac{\omega}{\omega_0} = \frac{I}{I + mR^2}$$

2.6 Energy

The concept of energy was not nearly as popular in Newton's time as it is today. Later we shall study two new formulations of dynamics, different from Newton's, based on energy—the Lagrangian and Hamiltonian methods.

Early in the nineteenth century, it became clear that heat was another form of energy and not a form of fluid (called "caloric") that flowed between hot and cold bodies. Count Rumford is generally given credit for realizing that the great amount of heat generated during the boring of a cannon was caused by friction and not the caloric. If frictional energy is just heat energy, interchangeable with mechanical energy, then a total conservation of energy can occur.

Throughout the nineteenth century, scientists performed experiments on the conservation of energy, resulting in the prominence given energy today. Hermann von Helmholtz formulated the general law of conservation of energy in 1847. He based his conclusion largely on the calorimetric experiments of James Prescott Joule begun in 1840.

Consider a point particle under the influence of a conservative force with potential U. The conservation of energy (actually, mechanical energy, to be precise in this case) is reflected in Equation 2.90.

$$E = T + U = \frac{1}{2}mv^2 + U(x) \tag{2.96}$$

where we consider only the one-dimensional case. We can rewrite Equation 2.96 as

$$v(t) = \frac{dx}{dt} = \pm\sqrt{\frac{2}{m}[E - U(x)]} \tag{2.97}$$

and by integrating

$$t - t_0 = \int_{x_0}^{x} \frac{\pm dx}{\sqrt{\frac{2}{m}[E - U(x)]}} \tag{2.98}$$

where $x = x_0$ at $t = t_0$. We have formally solved the one-dimensional case in Equation 2.98; that is, we have found x(t). All that remains is to insert the potential U(x) into Equation 2.98 and integrate, using computer techniques if necessary. We shall study later in some detail the potentials $U = \frac{1}{2} kx^2$ for harmonic oscillations and U= -k/x for the gravitational force.

We can learn a good deal about the motion of a particle simply by examining a plot of an example of U(x) as shown in Figure 2-14.

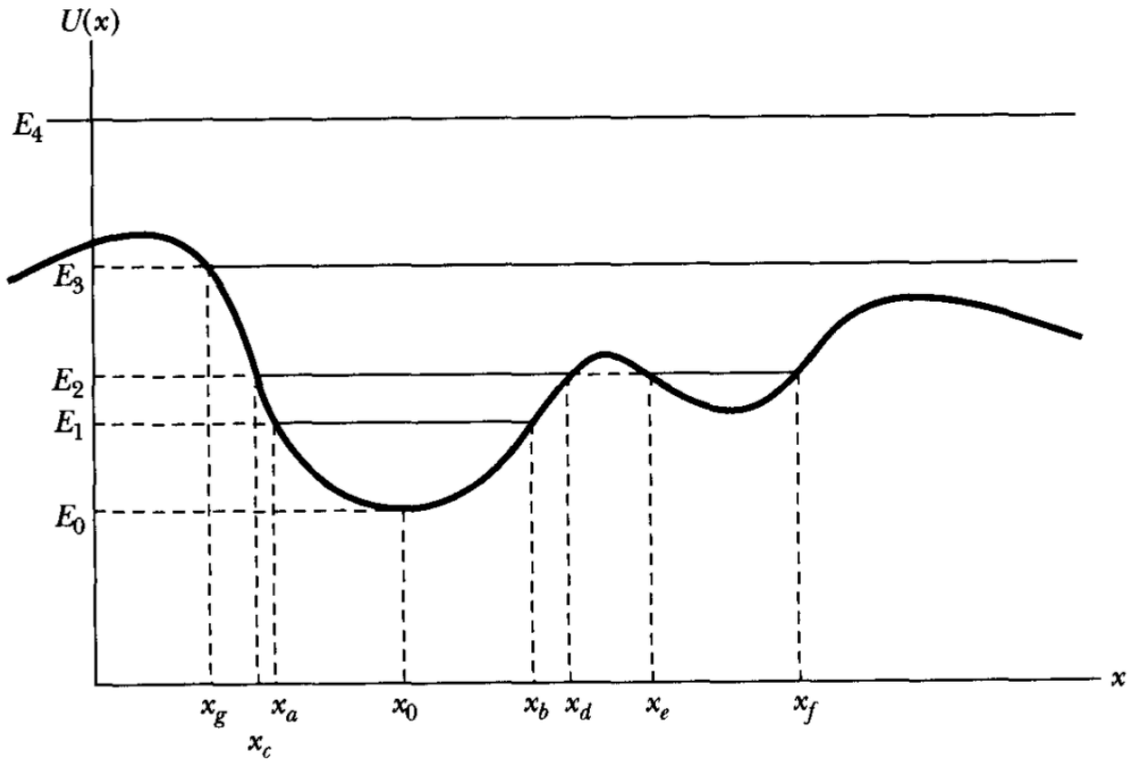


FIGURE 2-14 Potential energy $U(x)$ curve with various energies E indicated. For certain energies, for example E_1 and E_2, the motion is bounded.

First, notice that, because $\frac{1}{2}mv^2 = T > 0$, $E \geq U(x)$ for any real physical motion. We see in Figure 2-14 that the motion is bounded for energies E_1 and E_2. For E_1, the motion is periodic between the turning points x_a and x_b. Similarly, for E_2 the motion is periodic, but there are two possible regions: $x_c \leq x \leq x_d$ and $x_e \leq x \leq x_f$. The particle cannot "jump" from one "pocket" to the other; once in a pocket, it must remain there forever if its energy remains at E_2. The motion for a particle with energy E_0 has only one value, $x = x_0$. The particle is at rest with T= 0 [$E_0 = U(x)$].

The motion for a particle with energy E_3 is simple: The particle comes in from infinity, stops and turns at $x = x_g$ and returns to infinity - much like a tennis ball bouncing against a practice wall. For the energy E_4, the motion is unbounded and the particle may be at any position. Its speed will change because it depends on the difference between E_4 and U(x). If it is moving to the right, it will speed up and slow down but continue to infinity.

The motion of a particle of energy E_1 is similar to that of a mass at the end of a spring. The potential in the region $x_a \leq x \leq x_b$, can be approximated by $U(x) = \frac{1}{2} k(x - x_0)^2$. A particle with energy barely above E_0 will oscillate about the point $x = x_0$. We refer to such a point as an **equilibrium point**, because if the particle is placed at $x = x_0$ it remains there. Equilibrium may be stable, unstable, or neutral. The equilibrium just discussed is *stable* because if the particle were placed on either side of $x = x_0$ it would eventually return there. We can use a hemispherical mixing bowl with a steel ball as an example. With the bowl right side up, the ball can roll around inside the bowl; but it will eventually settle to the bottom - in other words, there is a stable equilibrium. If we turn the bowl upside down and place the ball precisely outside at $x = x_0$, the ball remains there in equilibrium. If we place the ball on either side of $x = x_0$ on the rounded surface, it rolls off; we call this *unstable* equilibrium. *Neutral* equilibrium would apply when the ball rolls on a flat, smooth, horizontal surface.

In general, we can express the potential U(x) in a Taylor series about a certain equilibrium point. For mathematical simplicity, let us assume that the equilibrium point is at $x = 0$ rather than $x = x_0$ (if not, we can always redefine the coordinate system to make it so). Then we have

$$U(x) = U_0 + x\left(\frac{dU}{dx}\right)_0 + \frac{x^2}{2!}\left(\frac{d^2U}{dx^2}\right)_0 + \frac{x^3}{3!}\left(\frac{d^3U}{dx^3}\right)_0 + \cdots \tag{2.99}$$

The zero subscript indicates that the quantity is to be evaluated at $x = 0$. The potential energy U_0 at $x= 0$ is simply a constant that we can define to be zero without any loss of generality. If $x = 0$ is an equilibrium point, then

$$\left(\frac{dU}{dx}\right)_0 = 0 \quad \text{Equilibrium point} \tag{2.100}$$

and Equation 2.99 becomes

$$U(x) = \frac{x^2}{2!}\left(\frac{d^2U}{dx^2}\right)_0 + \frac{x^3}{3!}\left(\frac{d^3U}{dx^3}\right)_0 + \cdots \tag{2.101}$$

Near the equilibrium point x = 0, the value of x is small, and each term in Equation 2.101 is considerably smaller than the previous one. Therefore, we keep only the first term in Equation 2.101:

$$U(x) = \frac{x^2}{2}\left(\frac{d^2U}{dx^2}\right)_0 \tag{2.102}$$

We can determine whether the equilibrium at x = 0 is stable or unstable by examining $(d^2U/dx^2)_0$. If x = 0 is a stable equilibrium, U(x) must be greater (more positive) on either side of x = 0. Because x^2 is always positive, the conditions for the equilibrium are

$$\left(\frac{d^2U}{dx^2}\right)_0 > 0 \quad \text{Stable equilibrium}$$
$$\left(\frac{d^2U}{dx^2}\right)_0 < 0 \quad \text{Unstable equilibrium} \tag{2.103}$$

If $(d^2U/dx^2)_0$ is zero, higher-order terms must be examined.

EXAMPLE 2.12

Consider the system of pulleys, masses, and string shown in Figure 2-15.

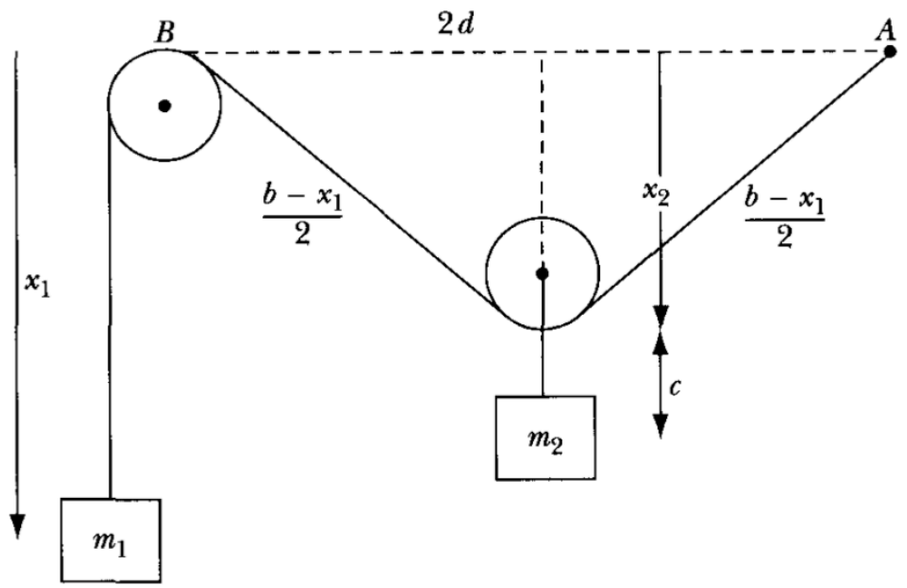


FIGURE 2-15 Example 2.12.

A light string of length b is attached at point A, passes over a pulley at point B located a distance 2d away, and finally attaches to mass m_1. Another pulley with mass m_2 attached passes over the string, pulling it down between A and B. Calculate the distance x_1 when the system is in equilibrium, and determine whether the equilibrium is stable or unstable. The pulleys are massless.

Solution. We can solve this example by either using forces (i.e., when $\ddot{x}_1 = 0 = \dot{x}_1$) or energy. We choose the energy method, because in equilibrium the kinetic energy is zero and we need to deal only with the potential energy when Equation 2.100 applies.

We let U= 0 along the line AB.

$$U = -m_1 g x_1 - m_2 g(x_2 + c) \tag{2.104}$$

We assume that the pulley holding mass m_2 is small, so we can neglect the pulley radius. The distance c in Figure 2-15 is constant.

$$x^2 = \sqrt{[(b - x_1)^2/4]} - d^2$$

$$U = -m_1 g x_1 - m_2 g \sqrt{[(b - x_1)^2/4]} - d^2 - m_2 g c$$

By setting $dU/dx_1 = 0$, we can determine the equilibrium position $(x_1)_0 = x_0$:

$$\left(\frac{dU}{dx_1}\right)_0 = -m_1 g + \frac{m_2 g(b - x_0)}{4\sqrt{[(b - x_0)^2/4]} - d^2} = 0$$

$$4m_1 \sqrt{[(b - x_0)^2/4]} - d^2 = m_2(b - x_0)$$

$$(b - x_0)^2(4m_1^2 - m_2^2) = 16 m_1^2 d^2$$

$$x_0 = b - \frac{4m_1 d}{\sqrt{4m_1^2 - m_2^2}} \qquad (2.105)$$

Notice that a real solution exists only when $4m_1^2 > m_2^2$.

Under what circumstances will the mass m_2 pull the mass m_1 up to the pulley B(i.e., $x_1 = 0$)? We can use Equation 2.103 to determine whether the equilibrium is stable or unstable:

$$\frac{d^2U}{dx_1^2} = \frac{-m_2 g}{4\{[(b - x_1)^2/4] - d^2\}^{1/2}} + \frac{m_2 g(b - x_1)^2}{16\{[(b - x_1)^2/4] - d^2\}^{3/2}}$$

Now insert $x_1 = x_0$.

$$\left(\frac{d^2U}{dx_1^2}\right)_0 = \frac{g(4m_1^2 - m_2^2)^{3/2}}{4m_2^2 d}$$

The condition for the equilibrium (real motion) previously was for $4m_1^2 > m_2^2$, so the equilibrium, when it exists, will be stable, because $(d^2U/dx^2)_0 > 0$.

EXAMPLE 2.13

Consider the one-dimensional potential

$$U(x) = \frac{-Wd^2(x^2 + d^2)}{x^4 + 8d^4} \tag{2.106}$$

Sketch the potential and discuss the motion at various values of x. Is the motion bounded or unbounded? Where are the equilibrium values? Are they stable or unstable? Find the turning points for E= -W/8. The value of W is a positive constant.

Solution. Rewrite the potential as

$$Z(y) = \frac{U(x)}{W} = \frac{-(y^2 + 1)}{y^4 + 8} \qquad \text{where } y = \frac{x}{d} \tag{2.107}$$

First, find the equilibrium points, which will help guide us in sketching the potential.

$$\frac{dZ}{dy} = \frac{-2y}{y^4 + 8} + \frac{4y^3(y^2 + 1)}{(y^4 + 8)^2} = 0$$

This is reduced to

$$y(y^4 + 2y^2 - 8) = 0$$
$$y(y^2 + 4)(y^2 - 2) = 0$$
$$y_0^2 = 2, 0$$

so

$$\left.\begin{aligned} x_{01} &= 0 \\ x_{02} &= \sqrt{2}d \\ x_{03} &= -\sqrt{2}d \end{aligned}\right\} \tag{2.108}$$

There are three equilibrium points. We sketch U(x)/W versus x/d in Figure 2-16.

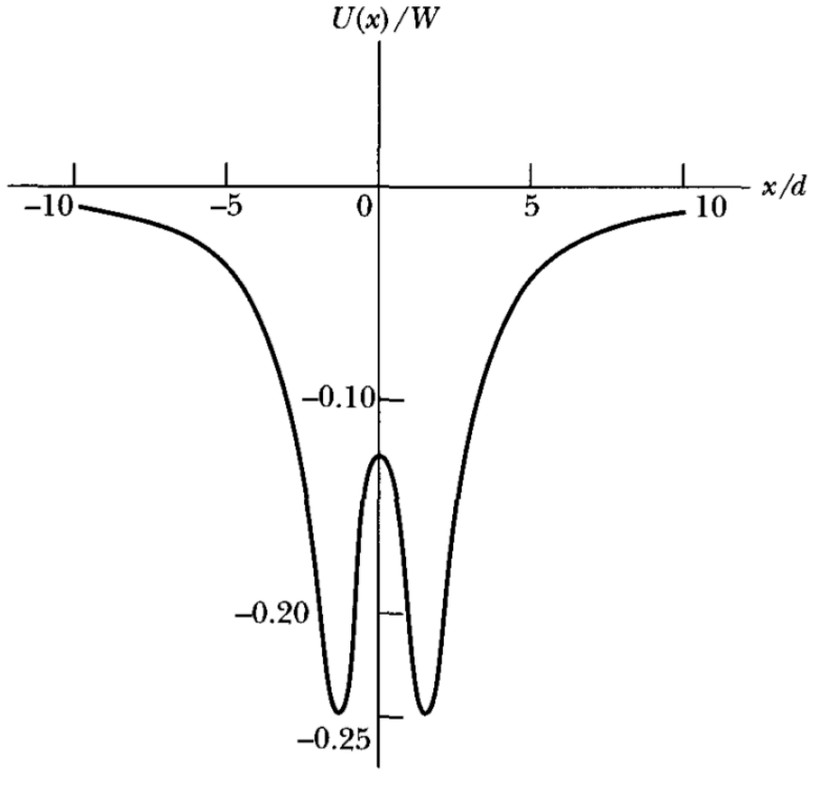


FIGURE 2-16 Example 2.13. Sketch of $U(x)/W$.

The equilibrium is stable at x_{02} and x_{03} but unstable at x_{01}. The motion is bounded for all energies $E < 0$. We can determine turning points for any energy E by setting E = U(x).

$$E = -\frac{W}{8} = U(y) = \frac{-W(y^2 + 1)}{y^4 + 8} \tag{2.109}$$

$$y^4 + 8 = 8y^2 + 8$$

$$y^4 = 8y^2$$

$$y = \pm 2\sqrt{2}, 0 \tag{2.110}$$

The turning points for E = -W/8 are x = - $2\sqrt{2}d$ and + $2\sqrt{2}d$, as well as x = 0 - which is the unstable equilibrium point.

2.7 Limitations of Newtonian Mechanics

In this chapter, we have introduced such concepts as position, time, momentum, and energy. We have implied that these are all measurable quantities and that they can be specified with any desired accuracy, depending only on the degree of sophistication of our measuring instruments. Indeed, this implication appears to be verified by our experience with all macroscopic objects. At any given instant of time, for example, we can measure with great precision the position of, say, a planet in its orbit about the sun. A series of such measurements allows us to determine (also with great precision) the planet's velocity at any given position.

When we attempt to make precise measurements on microscopic objects, however, we find a fundamental limitation in the accuracy of the results. For example, we can conceivably measure the position of an electron by scattering a light photon from the electron. The wave character of the photon precludes an exact measurement, and we can determine the position of the electron only within some uncertainty Δx related to the extent (i.e., the wavelength) of the photon. By the very act of measurement, however, we have induced a change in the state of the electron, because the scattering of the photon imparts momentum to the electron. This momentum is uncertain by an amount Δp. The product Δx Δp is a measure of the precision with which we can simultaneously determine the electron's position and momentum; $\Delta x \to 0, \Delta p \to 0$ implies a measurement with all imaginable precision. It was shown by the German physicist Werner Heisenberg in 1927 that this product must always be larger than a certain minimum value. We cannot, then, simultaneously specify both the position and momentum of the electron with infinite precision, for if $\Delta x \to 0$, then we must have $\Delta p \to \infty$ for **Heisenberg's uncertainty principle** to be satisfied.

The minimum value of $\Delta x \Delta p$ is of the order of 10^{-34} J•s. This is extremely small by macroscopic standards, so for laboratory-scale objects there is no practical difficulty in performing simultaneous measurements of position and momentum.

Newton's laws can therefore be applied as if position and momentum were precisely definable. But because of the uncertainty principle, Newtonian mechanics cannot be applied to microscopic systems. To overcome these fundamental difficulties in the Newtonian system, a new method of dealing with microscopic phenomena was developed, beginning in 1926. The work of Erwin Schrodinger, Heisenberg, Max Born, Paul Dirac, and others subsequently placed this new discipline on a firm foundation. Newtonian mechanics, then, is perfectly adequate for describing large-scale phenomena. But we need the new mechanics (quantum mechanics) to analyze processes in the atomic domain. As the size of the system increases, quantum mechanics goes over into the limiting form of Newtonian mechanics.

In addition to the fundamental limitations of Newtonian mechanics as applied to microscopic objects, there is another inherent difficulty in the Newtonian scheme—one that rests on the concept of time. In the Newtonian view, time is absolute, that is, it is supposed that it is always possible to determine unambiguously whether two events have occurred simultaneously or whether one has preceded the other. To decide on the time sequence of events, the two observers of the events must be in instantaneous communication,- either through some system of signals or by establishing two exactly synchronous clocks at the points of observation. But the setting of two clocks into exact synchronism requires the knowledge of the time of transit of a signal in one direction from one observer to the other (We could accomplish this if we already had two synchronous clocks, but this is a circular argument.) When we actually measure signal velocities, however, we always obtain an average velocity for propagation in opposite directions. And to devise an experiment to measure the velocity in only one direction inevitably leads to the introduction of some new assumption that we
cannot verify before the experiment.

We know that instantaneous communication by signaling is impossible: Interactions between material bodies propagate with finite velocity, and an interaction of some sort must occur for a signal to be transmitted. The maximum velocity with which any signal can be propagated is that of light in free space: $c = 3 \times 10^8$ m/s.

The difficulties in establishing a time scale between separate points lead us to believe that time is, after all, not absolute and that space and time are somehow intimately related. The solution to the dilemma was found during the period 1904-1905 by Hendrik Lorenz, Henri Poincare, and Albert Einstein and is embodied in the special theory of relativity (see Chapter 14).

Newtonian mechanics is therefore subject to fundamental limitations when small distances or high velocities are encountered. Difficulties with Newtonian mechanics may also occur when massive objects or enormous distances are involved. A practical limitation also occurs when the number of bodies constituting the system is large. In Chapter 8, we see that we cannot obtain a general solution in closed form for the motion of a system of more than two interacting bodies even for the relatively simple case of gravitational interaction. To calculate the motion in a three-body system, we must resort to a numerical approximation procedure. Although such a method is in principle capable of any desired accuracy, the labor involved is considerable. The motion in even more complex systems (for example, the system composed of all the major objects in the solar system) can likewise be computed, but the procedure rapidly becomes too unwieldy to be of much use for any larger system. To calculate the motion of the individual molecules in, say, a cubic centimeter of gas containing $\approx 10^{19}$ molecules is clearly out of the question. A successful method of calculating the average properties of such systems was developed in the latter part of the nineteenth century by Boltzmann, Maxwell, Gibbs, Liouville, and others. These procedures allowed the dynamics of systems to be calculated from probability theory, and a statistical mechanics was evolved. Some comments regarding the formulation of statistical concepts in mechanics are found in Section 7.13.

Chapter 3 - Oscillations

3.1 Introduction

We begin by considering the oscillatory motion of a particle constrained to move in one dimension. We assume that a position of stable equilibrium exists for the particle, and we designate this point as the origin (see Section 2.6). If the particle is displaced from the origin (in either direction), a certain force tends to restore the particle to its original position. An example is an atom in a long molecular chain. The restoring force is, in general, some complicated function of the displacement and perhaps of the particle's velocity or even of some higher time derivative of the position coordinate. We consider here only cases in which the restoring force **F** is a function only of the displacement: F= F(x).

We assume that the function F(x) that describes the restoring force possesses continuous derivatives of all orders so that the function can be expanded in a Taylor series:

$$F(x) = F_0 + x\left(\frac{dF}{dx}\right)_0 + \frac{1}{2!}x^2\left(\frac{d^2F}{dx^2}\right)_0 + \frac{1}{3!}x^3\left(\frac{d^3F}{dx^3}\right)_0 + \cdots \tag{3.1}$$

where F_0 is the value of F(x) at the origin (x = 0), and $(d^nF/dx^n)_0$ is the value of the nth derivative at the origin. Because the origin is defined to be the equilibrium point, F_0 must vanish, because otherwise the particle would move away from the equilibrium point and not return. If, then, we confine our attention to displacements of the particle that are sufficiently small, we can normally neglect all terms involving x^2 and higher powers of x. We have, therefore, the approximate relation

$$\boxed{F(x) = -kx} \tag{3.2}$$

where we have substituted $k = -(dF/dx)_0$. Because the restoring force is always directed toward the equilibrium position (the origin), the derivative $(dF/dx)_0$ is negative, and therefore k is a positive constant.

Only the first power of the displacement occurs in F(x), so the restoring force in this approximation is a linear force.

Physical systems described in terms of Equation 3.2 obey Hooke's Law. One of the classes of physical processes that can be treated by applying Hooke's Law is that involving elastic deformations. As long as the displacements are small and the elastic limits are not exceeded, a linear restoring force can be used for problems of stretched springs, elastic springs, bending beams, and the like. But we must emphasize that such calculations are only approximate, because essentially every real restoring force in nature is more complicated than the simple Hooke's Law force. Linear forces are only useful approximations, and their validity is limited to cases in which the amplitudes of the oscillations are small.

Damped oscillations, usually resulting from friction, are almost always the type of oscillations that occur in nature. We learn in this chapter how to design an efficiently damped system. This damping of the oscillations may be counteracted if some mechanism supplies the system with energy from an external source at a rate equal to that absorbed by the damping medium. Motions of this type are called **driven** (or **forced**) **oscillations**. Normally sinusoidal, they have important applications in mechanical vibrations as well as in electrical systems.

The extensive discussion of linear oscillatory systems is warranted by the great importance of oscillatory phenomena in many areas of physics and engineering. It is frequently permissible to use the linear approximation in the analysis of such systems. The usefulness of these analyses is due in large measure to the fact that we can usually use *analytical* methods.

When we look more carefully at physical systems, we find that a large number of them are *nonlinear* in general. We will discuss nonlinear systems in Chapter 4.

3.2 Simple Harmonic Oscillator

The equation of motion for the simple harmonic oscillator may be obtained by substituting the Hooke's Law force into the Newtonian equation F = ma. Thus

$$-kx = m\ddot{x} \tag{3.3}$$

If we define

$$\omega_0^2 \equiv k/m \tag{3.4}$$

Equation 3.3 becomes

$$\boxed{\ddot{x} + \omega_0^2 x = 0} \tag{3.5}$$

According to the results of Appendix C, the solution of this equation can be expressed in either of the forms

$$x(t) = A \sin(\omega_0 t - \delta) \tag{3.6a}$$

$$x(t) = A \cos(\omega_0 t - \phi) \tag{3.6b}$$

where the phases δ and ϕ differ by $\pi/2$. (An alteration of the phase angle corresponds to a change of the instant that we designate t = 0, the origin of the time scale.) Equations 3.6a and b exhibit the well-known sinusoidal behavior of the displacement of the simple harmonic oscillator

We can obtain the relationship between the total energy of the oscillator and the amplitude of its motion as follows. Using Equation 3.6a for x(t), we find for the kinetic energy,

$$\begin{aligned} T = \frac{1}{2} m\dot{x}^2 &= \frac{1}{2} m\omega_0^2 A^2 \cos^2(\omega_0 t - \delta) \\ &= \frac{1}{2} kA^2 \cos^2(\omega_0 t - \delta) \end{aligned} \tag{3.7}$$

The potential energy may be obtained by calculating the work required to displace the particle a distance x. The incremental amount of work dW necessary to move the particle by an amount dx against the restoring force F is

$$dW = -F\,dx = kx\,dx \tag{3.8}$$

Integrating from 0 to x and setting the work done on the particle equal to the potential energy, we have

$$U = \frac{1}{2}kx^2 \tag{3.9}$$

Then

$$U = \frac{1}{2}kA^2 \sin^2(\omega_0 t - \delta) \tag{3.10}$$

Combining the expressions for T and U to find the total energy E, we have

$$E = T + U = \frac{1}{2}kA^2[\cos^2(\omega_0 t - \delta) + \sin^2(\omega_0 t - \delta)]$$

$$\boxed{E = T + U = \frac{1}{2}kA^2} \tag{3.11}$$

so that the total energy is proportional to the square of the amplitude; this is a general result for linear systems. Notice also that E is independent of the time; that is, energy is conserved. (Energy conservation is guaranteed, because we have been considering a system without frictional losses or other external forces.)

The period τ_0 of the motion is defined to be the time interval between successive repetitions of the particle's position and direction of motion. Such an interval occurs when the argument of the sine in Equation 3.6a increases by 2π :

$$\omega_0 \tau_0 = 2\pi \tag{3.12}$$

or

$$\tau_0 = 2\pi\sqrt{\frac{m}{k}} \tag{3.13}$$

From this expression, as well as from Equation 3.6, it should be clear that ω_0 represents the **angular frequency** of the motion, which is related to the **frequency** ω_0 by

$$\boxed{\omega_0 = 2\pi\nu_0 = \sqrt{\frac{k}{m}}} \tag{3.14}$$

$$\boxed{\nu_0 = \frac{1}{\tau_0} = \frac{1}{2\pi}\sqrt{\frac{k}{m}}} \tag{3.15}$$

Note that the period of the simple harmonic oscillator is independent of the amplitude (or total energy); a system exhibiting this property is said to be **isochronous**.

For many problems, of which the simple pendulum is the best example, the equation of motion results in $\ddot{\theta} + \omega_0^2 \sin\theta = 0$, where θ is the displacement angle from equilibrium, and $\omega_0^2 = \sqrt{g/\ell}$, where ℓ is the length of the pendulum arm. We can make this differential equation describe simple harmonic motion by invoking the small oscillation assumption. If the oscillations about the equilibrium are small, we expand $\sin\theta$ and $\cos\theta$ in power series (see Appendix A) and keep only the lowest terms of importance. This often means $\sin\theta \simeq \theta$ and $\cos\theta \simeq 1 - \theta^2/2$, where θ is measured in radians. If we use the small oscillation approximation for the simple pendulum, the equation of motion above becomes $\ddot{\theta} + \omega_0^2\theta = 0$, an equation that does represent simple harmonic motion. We shall often invoke this assumption throughout this text and in its problems.

EXAMPLE 3.1

Find the angular velocity and period of oscillation of a solid sphere of mass m and radius R about a point on its surface. See Figure 3-1.

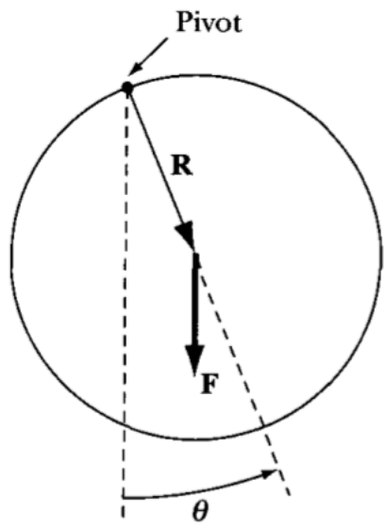


FIGURE 3-1 Example 3.1. The physical pendulum (sphere).

Solution. Let the rotational inertia of the sphere be I about the pivot point. In elementary physics we learn that the value of the rotational inertia about an axis through the sphere's center is $2/5mR^2$. If we use the parallel-axis theorem, the rotational inertia about the pivot point on the surface is $2/5mR^2 + mR^2 = 7/5mR^2$. The equilibrium position of the sphere occurs when the center of mass (center of sphere) is hanging directly below the pivot point. The gravitational force F= mg pulls the sphere back towards the equilibrium position as the sphere swings back and forth with angle θ . The torque on the sphere is $N = I\alpha$, where $\alpha = \ddot{\theta}$ is the angular acceleration. The torque is also **N** = **R** x **F**, with N= RF $\sin\theta$ = Rmg $\sin\theta$. For small oscillations, we have N=Rmg θ . We must have $I\ddot{\theta} = -Rmg\theta$ for the equation of motion in this case, because as $\dot{\theta}$ increases, $\ddot{\theta}$ is negative. We need to solve the equation of motion for θ .

$$\ddot{\theta} + \frac{Rmg}{I}\theta = 0$$

This equation is similar to Equation 3.5 and has solutions for the angular frequency and period from Equations 3.14 and 3.15.

$$\omega = \sqrt{\frac{Rmg}{I}} = \sqrt{\frac{Rmg}{\frac{7}{5}mR^2}} = \sqrt{\frac{5g}{7R}}$$

and

$$\tau = 2\pi\sqrt{\frac{I}{Rmg}} = 2\pi\sqrt{\frac{\frac{7}{5}mR^2}{Rmg}} = 2\pi\sqrt{\frac{7R}{5g}}$$

Note that the mass m does not enter Only the distance R to the center of determines the oscillation frequency.

3.3 Harmonic Oscillations in Two Dimensions

We next consider the motion of a particle that is allowed two degrees of freedom. We take the restoring force to be proportional to the distance of the particle from a force center located at the origin and to be directed toward the origin:

$$\mathbf{F} = -k\mathbf{r} \tag{3.16}$$

which can be resolved in polar coordinates into the components

$$\left.\begin{aligned} F_x &= -kr\cos\theta = -kx \\ F_y &= -kr\sin\theta = -ky \end{aligned}\right\} \tag{3.17}$$

The equations of motion are

$$\left.\begin{aligned} \ddot{x} + \omega_0^2 x &= 0 \\ \ddot{y} + \omega_0^2 y &= 0 \end{aligned}\right\} \tag{3.18}$$

where, as before, $\omega_0^2 = k/m$. The solutions are

$$\left.\begin{aligned} x(t) &= A\cos(\omega_0 t - \alpha) \\ y(t) &= B\cos(\omega_0 t - \beta) \end{aligned}\right\} \tag{3.19}$$

Thus, the motion is one of simple harmonic oscillation in each of the two directions, both oscillations having the same frequency but possibly differing in amplitude and in phase. We can obtain the equation for the path of the particle by eliminating the time t between the two equations (Equation 3.19). First we write

$$\begin{aligned} y(t) &= B\cos[\omega_0 t - \alpha + (\alpha - \beta)] \\ &= B\cos(\omega_0 t - \alpha)\cos(\alpha - \beta) - B\sin(\omega_0 t - \alpha)\sin(\alpha - \beta) \end{aligned} \tag{3.20}$$

Defining $\delta = \alpha - \beta$ and noting that $\cos(\omega_0 t - \alpha) = x/A$, we have

$$y = \frac{B}{A}x\cos\delta - B\sqrt{1 - \left(\frac{x^2}{A^2}\right)}\sin\delta$$

or

$$Ay - Bx\cos\delta = -B\sqrt{A^2 - x^2}\sin\delta \tag{3.21}$$

On squaring, this becomes

$$A^2y^2 - 2ABxy\cos\delta + B^2x^2\cos^2\delta = A^2B^2\sin^2\delta - B^2x^2\sin^2\delta$$

so that

$$B^2x^2 - 2ABxy\cos\delta + A^2y^2 = A^2B^2\sin^2\delta \tag{3.22}$$

If δ is set equal to $\pm\pi/2$, this equation reduces to the easily recognized equation for an ellipse:

$$\frac{x^2}{A^2} + \frac{y^2}{B^2} = 1, \qquad \delta = \pm\pi/2 \tag{3.23}$$

If the amplitudes are equal, A= B, and if $\delta = \pm\pi/2$, we have the special case of circular motion:

$$x^2 + y^2 = A^2, \quad \text{for } A = B \text{ and } \delta = \pm\pi/2 \tag{3.24}$$

Another special case results if the phase δ vanishes; then we have

$$B^2x^2 - 2ABxy + A^2y^2 = 0, \qquad \delta = 0$$

Factoring,

$$(Bx - Ay)^2 = 0$$

which is the equation of a straight line:

$$y = \frac{B}{A}x, \qquad \delta = 0 \tag{3.25}$$

Similarly, the phase $\delta = \pm\pi$ yields the straight line of opposite slope:

$$y = -\frac{B}{A}x, \qquad \delta = \pm\pi \tag{3.26}$$

The curves of Figure 3-2 illustrate Equation 3.22 for the case A= B; δ = 90° or 270° yields a circle, and δ = 180° or 360°(0°) yields a straight line. All other values of δ yield ellipses.

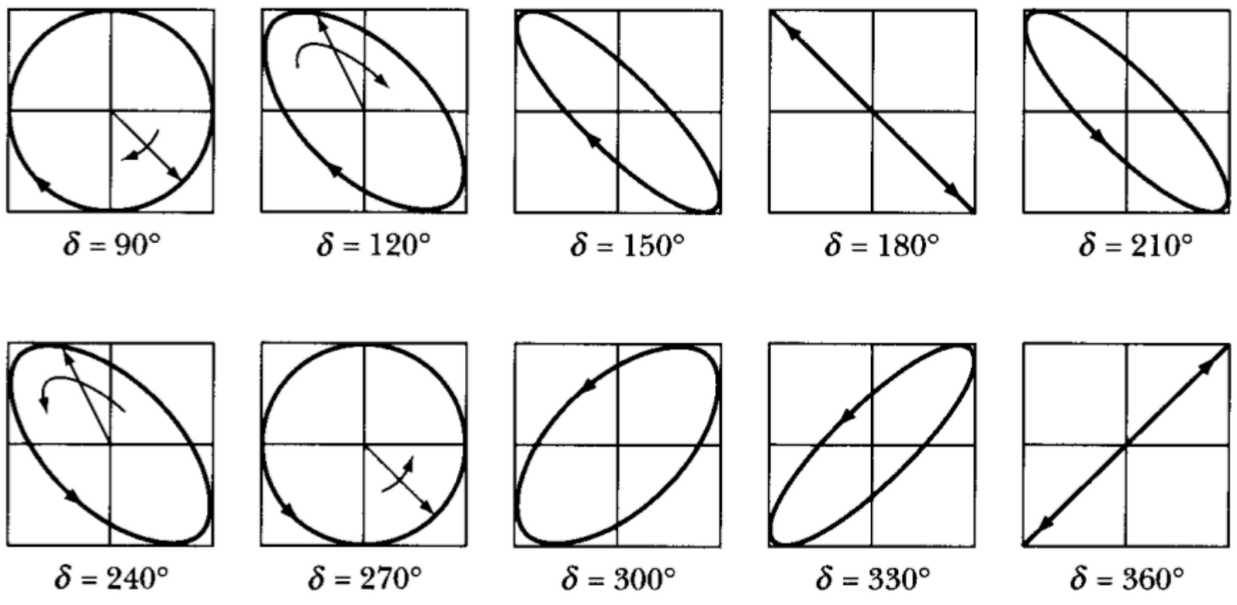


FIGURE 3-2 Two-dimensional harmonic oscillation motion for various phase angles $\delta = \alpha - \beta$.

In the general case of two-dimensional oscillations, the angular frequencies for the motions in the x- and y-directions need not be equal, so that Equation 3.19 becomes

$$\left.\begin{aligned} x(t) &= A\cos(\omega_x t - \alpha) \\ y(t) &= B\cos(\omega_y t - \beta) \end{aligned}\right\} \tag{3.27}$$

The path of the motion is no longer an ellipse but a Lissajous curve. Such a curve will be closed if the motion repeats itself at regular intervals of time. This will be possible only if the angular frequencies ω_x and ω_y are *commensurable*, that is, if ω_x/ω_y is a rational fraction. Such a case is shown in Figure 3-3, in which $\omega_y = \frac{3}{4}\omega_x$, (also $\alpha = \beta$). If the ratio of the angular frequencies is not a rational fraction, the curve will be open; that is, the moving particle will never pass twice through the same point with the same velocity. In such a case, after a sufficiently long time has elapsed, the curve will pass arbitrarily close to any given point lying within the rectangle 2A x 2B and will therefore "fill" the rectangle.

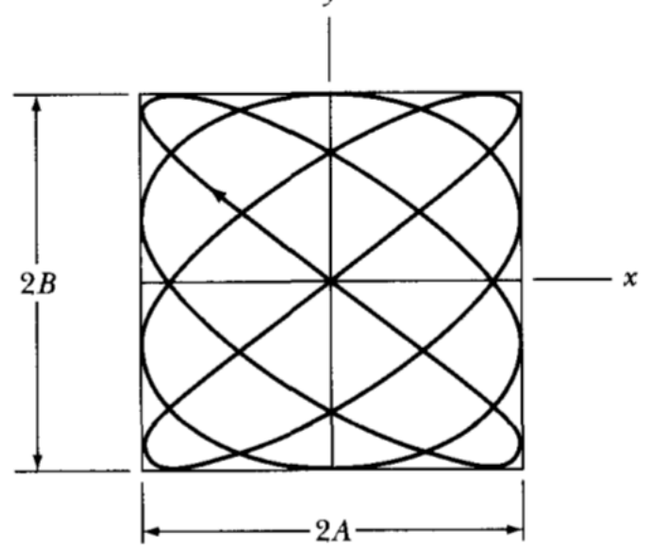


FIGURE 3-3 Closed two-dimensional oscillatory motion (called *Lissajous curves*) occurs under certain conditions for the x and y coordinates.

The two-dimensional oscillator is an example of a system in which an infinitesimal change can result in a qualitatively different type of motion. The motion will be along a closed path if the two angular frequencies are commensurable.

But if the angular frequency ratio deviates from a rational fraction by even an infinitesimal amount, then the path will no longer be closed and it will "fill" the rectangle. For the path to be closed, the angular frequency ratio must be known to be a rational fraction with infinite precision.

If the angular frequencies for the motions in the x- and y-directions are different, the shape of the resulting Lissajous curve strongly depends on the phase difference $\delta = \alpha - \beta$. Figure 3-4 shows the results for the case $\omega_y = 2\omega_x$ for phase differences of 0, $\pi/3$, and $\pi/2$.

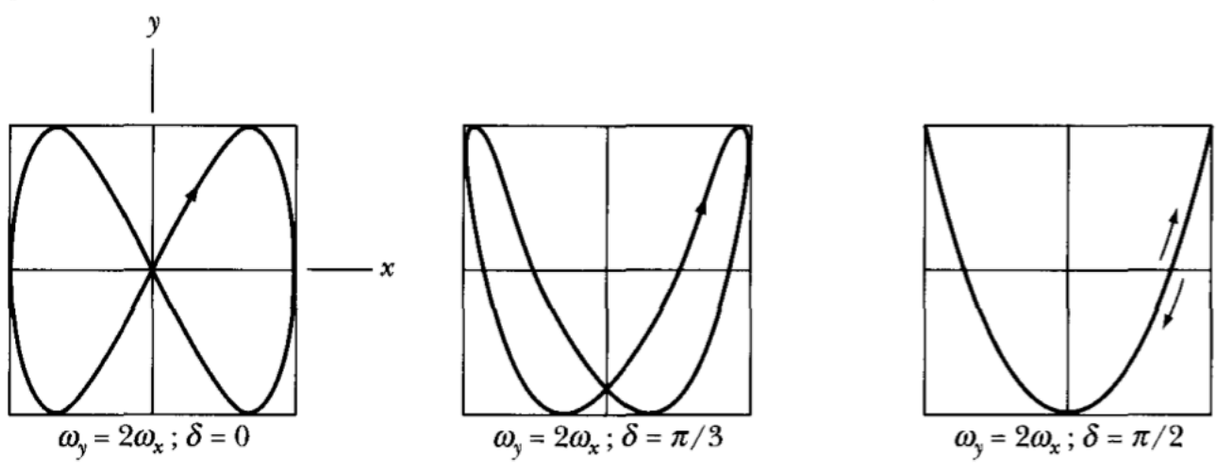


FIGURE 3-4 Lissajous curves depend strongly on the phase differences of the angle δ.

3.4 Phase Diagrams

The state of motion of a one-dimensional oscillator, such as that discussed in Section 3.2, will be completely specified as a function of time if *two* quantities are given at one instant of time, that is, the initial conditions x(t) and $\dot{x}$(t) (Two quantities are needed because the differential equation for the motion is of second order.) We may consider the quantities x(t) and $\dot{x}$(t) to be the coordinates of a point in a two-dimensional space, called **phase space**. (In two dimensions, the phase space is a phase plane. But for a general oscillator with n degrees of freedom, the phase space is a 2n-dimensional space.) As the time varies, the point P(x, $\dot{x}$) describing the state of the oscillating particle will move along a certain phase path in the phase plane. For different initial conditions of the oscillator, the motion will be described by different phase paths. Any given path represents the complete time history of the oscillator for a certain set of initial conditions. The totality of all possible phase paths constitutes the **phase portrait** or the **phase diagram** of the oscillator.

According to the results of the preceding section, we have, for the simple harmonic oscillator.

$$x(t) = A \sin(\omega_0 t - \delta) \tag{3.28a}$$

$$\dot{x}(t) = A\omega_0 \cos(\omega_0 t - \delta) \tag{3.28b}$$

If we eliminate t from these equations, we find for the equation of the path

$$\frac{x^2}{A^2} + \frac{\dot{x}^2}{A^2\omega_0^2} = 1 \tag{3.29}$$

This equation represents a family of ellipses, several of which are shown in Figure 3-5.

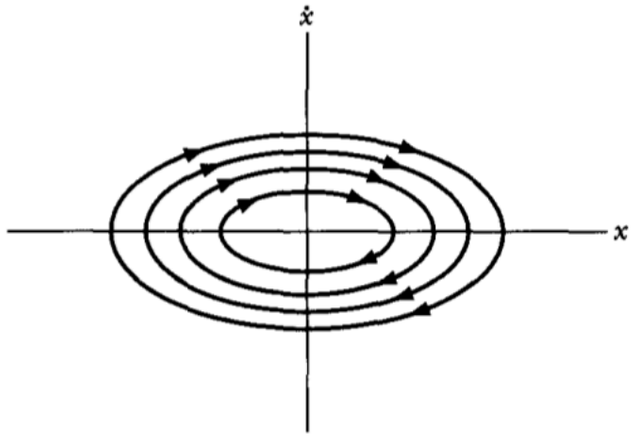


FIGURE 3-5 Phase diagram for a simple harmonic oscillator for a variety of total energies *E*.

We know that the total energy E of the oscillator is $\frac{1}{2}kA^2$ (Equation 3.11), and because ω_0^2 = k/m. Equation 3.29 can be written as

$$\frac{x^2}{2E/k} + \frac{\dot{x}^2}{2E/m} = 1 \tag{3.30}$$

Each phase path, then, corresponds to a definite total energy of the oscillator. This result is expected because the system is conservative (i.e., E = const.).

No two phase paths of the oscillator can cross. If they could cross, this would imply that for a given set of initial conditions $x(t_0)$, $\dot{x}(t_0)$ (i.e., the coordinates of the crossing point), the motion could proceed along different phase paths. But this is impossible because the solution of the differential equation is unique.

If the coordinate axes of the phase plane are chosen as in Figure 3-5, the motion of the **representative point** $P(x, \dot{x})$ will always be in a clockwise direction, because for $x > 0$ the velocity $\dot{x}$ is always decreasing and for $x < 0$ the velocity is always increasing.

To obtain Equations 3.28 for $x(t)$ and $\dot{x}(t)$, we must integrate Equation 3.5, a second-order differential equation:

$$\frac{d^2x}{dt^2} + \omega_0^2 x = 0 \tag{3.31}$$

We can obtain the equation for the phase path, however, by a simpler procedure, because Equation 3.31 can be replaced by the pair of equations

$$\frac{dx}{dt} = \dot{x}, \quad \frac{d\dot{x}}{dt} = -\omega_0^2 x \tag{3.32}$$

If we divide the second of these equations by the first, we obtain

$$\frac{d\dot{x}}{dx} = -\omega_0^2 \frac{x}{\dot{x}} \tag{3.33}$$

This is a first-order differential equation for $\dot{x} = \dot{x}(x)$, the solution to which is just Equation 3.29. For the simple harmonic oscillator, there is no difficulty in obtaining the general solution for the motion by solving the second-order equation. But in more complicated situations, it is sometimes considerably easier to directly find the equation of the phase path $\dot{x} = \dot{x}(x)$ without proceeding through the calculation $x(t)$.

3.5 Damped Oscillations

The motion represented by the simple harmonic oscillator is termed a free oscillation; once set into oscillation, the motion would never cease. This oversimplifies the actual physical case, in which dissipative or frictional forces would eventually damp the motion to the point that the oscillations would no longer occur. We can analyze the motion in such a case by incorporating into the differential equation a term representing the damping force. It does not seem reasonable that the damping force should, in general, depend on the displacement, but it could be a function of the velocity or perhaps of some higher time derivative of the displacement. It is frequently assumed that the damping force is a linear function of the velocity, $\mathbf{F}_d = \alpha\,\mathbf{v}$. We consider here only one-dimensional damped oscillations so that we can represent the damping term by $-b\dot{x}$. The parameter b must be positive in order that the force indeed be resisting. (A force $-b\dot{x}$ with $b < 0$ would act to increase the speed instead of decreasing it as any resisting force must.) Thus, if a particle of mass m moves under the combined influence of a linear restoring force -kx and a resisting force $-b\dot{x}$, the differential equation describing the motion is

$$m\ddot{x} + b\dot{x} + kx = 0 \tag{3.34}$$

which we can write as

$$\boxed{\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = 0} \tag{3.35}$$

Here $\beta = b/2m$ is the damping parameter and $\omega_0 = \sqrt{k/m}$ is the characteristic angular frequency in the absence of damping. The roots of the auxiliary equation are (cf. Equation C.8, Appendix C)

$$\left.\begin{aligned} r_1 &= -\beta + \sqrt{\beta^2 - \omega_0^2} \\ r_2 &= -\beta - \sqrt{\beta^2 - \omega_0^2} \end{aligned}\right\} \tag{3.36}$$

The general solution of Equation 3.35 is therefore

$$\boxed{x(t) = e^{-\beta t}\left[A_1 \exp\left(\sqrt{\beta^2 - \omega_0^2}\,t\right) + A_2 \exp\left(-\sqrt{\beta^2 - \omega_0^2}\,t\right)\right]} \tag{3.37}$$

There are three general cases of interest:

$$\begin{aligned} &\text{Underdamping:} && \omega_0^2 > \beta^2 \\ &\text{Critical damping:} && \omega_0^2 = \beta^2 \\ &\text{Overdamping:} && \omega_0^2 < \beta^2 \end{aligned}$$

The motion of the three cases is shown schematically in Figure 3-6 for specific initial conditions. We shall see that only the case of underdamping results in oscillatory motion. These three cases are discussed separately.

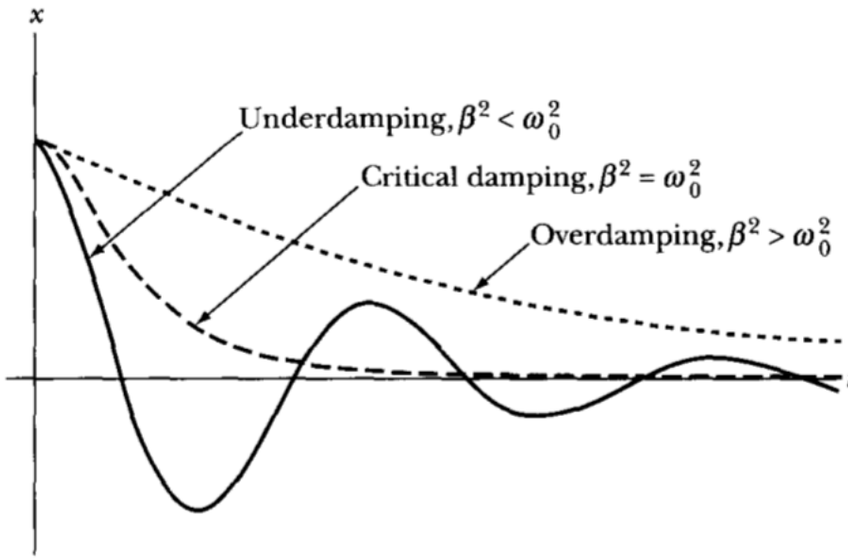


FIGURE 3-6 Damped oscillator motion for three cases of damping.

Underdamped Motion

For the case of underdamped motion, it is convenient to define

$$\omega_1^2 \equiv \omega_0^2 - \beta^2 \tag{3.38}$$

where $\omega_1^2 > 0$; then the exponents in the brackets of Equation 3.37 are imaginary, and the solution becomes

$$x(t) = e^{-\beta t}\left[A_1 e^{i\omega_1 t} + A_2 e^{-i\omega_1 t}\right] \tag{3.39}$$

Equation 3.39 can be rewritten as (See Exercise D-6, Appendix D)

$$x(t) = Ae^{-\beta t}\cos(\omega_1 t - \delta) \tag{3.40}$$

We call the quantity ω_1 the angular frequency of the damped oscillator. Strictly speaking, we cannot define a frequency when damping is present, because the motion is not periodic - that is, the oscillator never passes twice through a given point with the same velocity. However, because $\omega_1 = 2\pi/(2T_1)$, where T_1 is the time between adjacent zero x-axis crossings, the angular frequency ω_1 has meaning for a given time period. Note that 2 T_1 would be the "period" in this case, not T_1. For simplicity, we refer to ω_1 as the "angular frequency" of the damped oscillator, and we note that this quantity is less than the frequency of the oscillator in the absence of damping (i.e., $\omega_1 < \omega_0$). If the damping is small, then

$$\omega_1 = \sqrt{\omega_0^2 - \beta^2} \cong \omega_0$$

so the term angular frequency may be used. But the meaning is not precise unless $\beta = 0$.

The maximum amplitude of the motion of the damped oscillator decreases with time because of the factor $\exp(-\beta t)$, where $\beta > 0$, and the envelope of the displacement versus time curve is given by

$$x_{en} = \pm Ae^{-\beta t} \tag{3.41}$$

This envelope and the displacement curve are shown in Figure 3-7 for the case $\delta = 0$. The sinusoidal curve for undamped motion $\beta = 0$ is also shown in this figure. A close comparison of the two curves indicates that the frequency for the damped case is less (i.e., that the period is *longer*) than that for the undamped case.

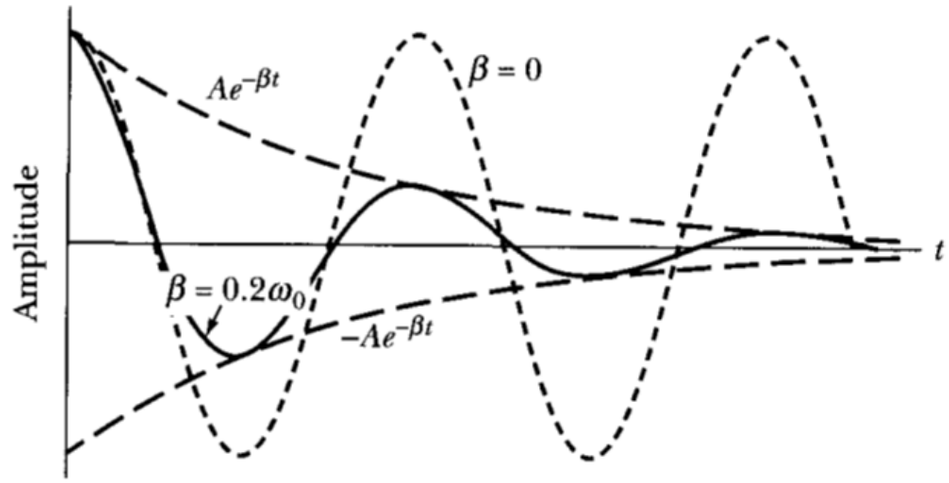


FIGURE 3-7 The underdamped motion (solid line) is an oscillatory motion (short dashes) that decreases within the exponential envelope (long dashes).

The ratio of the amplitudes of the oscillation at two successive maxima is

$$\frac{Ae^{-\beta T}}{Ae^{-\beta(T+\tau_1)}} = e^{\beta\tau_1} \tag{3.42}$$

where the first of any pair of maxima occurs at t = T and where $\tau_1 = 2\pi/\omega_1$. The quantity $\exp(\beta\tau_1)$ is called the **decrement** of the motion; the logarithm of $\exp(\beta\tau_1)$ - that is, $\beta\tau_1$ - is known as the **logarithmic decrement** of the motion.

Unlike the simple harmonic oscillator discussed previously, the energy of the damped oscillator is not constant in time; rather, energy is continually given up to the damping medium and dissipated as heat (or, perhaps, as radiation in the form of fluid waves). The rate of energy loss is proportional to the square of the velocity, so the decrease of energy does not take place uniformly. The loss rate will be a maximum when the particle attains its maximum velocity near (but not exactly at) the equilibrium position, and it will instantaneously vanish when the particle is at maximum amplitude and has zero velocity. Figure 3-8 shows the total energy and the rate of energy loss for the damped oscillator.

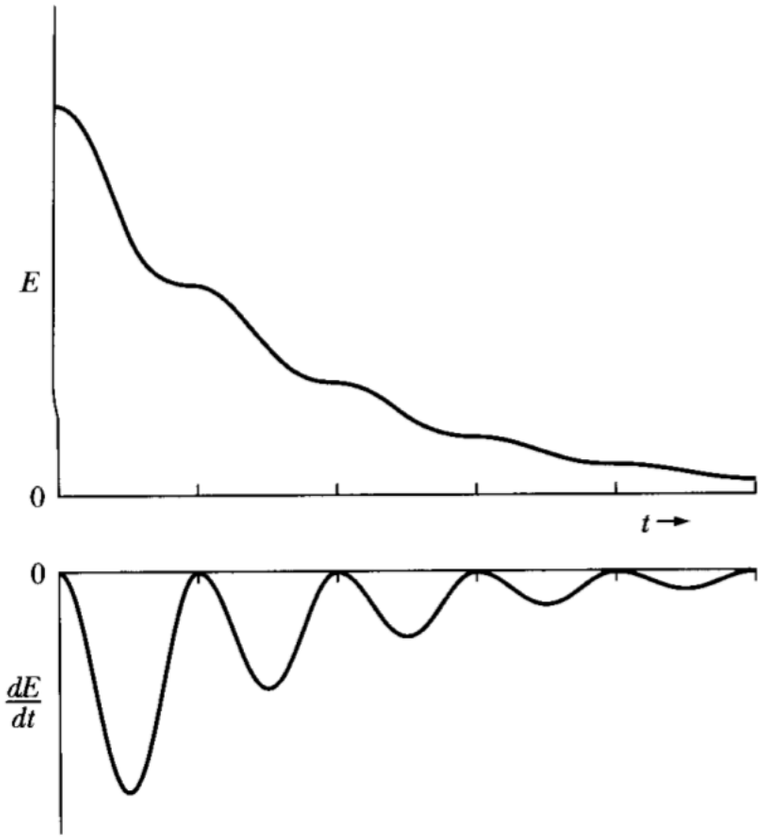


FIGURE 3-8 The total energy and rate of energy loss for the damped oscillator.

EXAMPLE 3.2

Construct a general phase diagram analytically for the damped oscillator Then, using a computer, make a plot for x and $\dot{x}$ versus t and a phase diagram for the following values: A = 1 cm, ω_0 = 1 rad/s, β = 0.2 s^{-1} and $\delta = \pi/2$ rad.

Solution. First, we write the expressions for the displacement and the velocity:

$$x(t) = Ae^{-\beta t}\cos(\omega_1 t - \delta)$$

$$\dot{x}(t) = -Ae^{-\beta t}[\beta\cos(\omega_1 t - \delta) + \omega_1 \sin(\omega_1 t - \delta)]$$

These equations can be converted into a more easily recognized form by introducing a change of variables according to the following linear transformations:

$$u = \omega_1 x, \quad w = \beta x + \dot{x}$$

Then

$$u = \omega_1 A e^{-\beta t} \cos(\omega_1 t - \delta)$$

$$w = -\omega_1 A e^{-\beta t} \sin(\omega_1 t - \delta)$$

If we represent *u* and *w* in polar coordinates (Figure 3-9), then

$$\rho = \sqrt{u^2 + w^2}, \qquad \phi = \omega_1 t$$

Thus

$$\rho = \omega_1 A e^{-(\beta/\omega_1)\phi}$$

which is the equation of a logarithmic spiral.

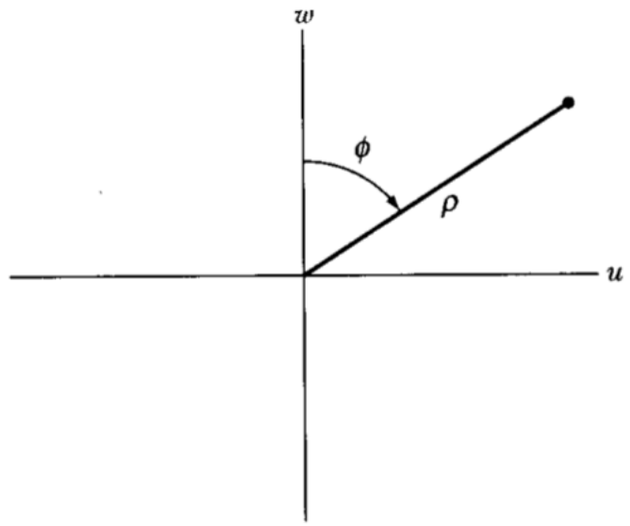
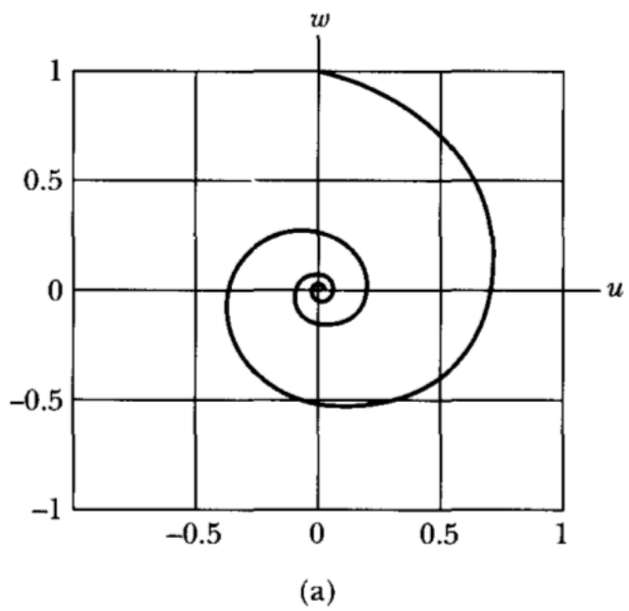


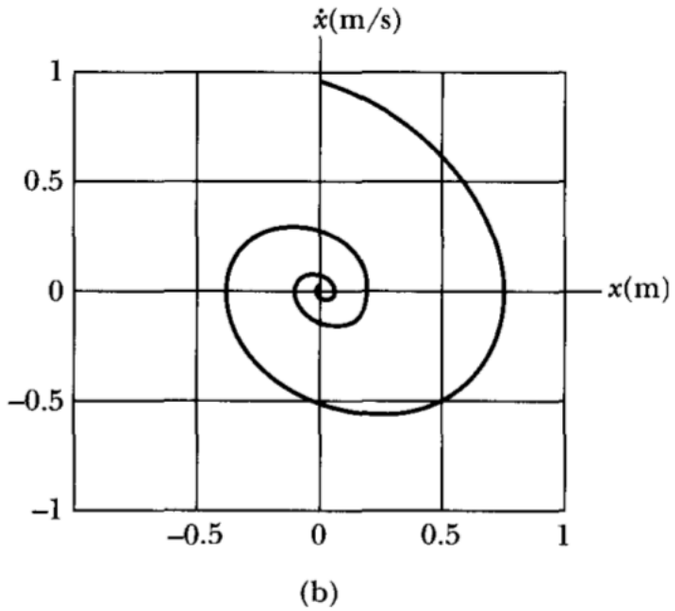
FIGURE 3-9 Example 3.2.

Because the transformation from x, $\dot{x}$ to u, w is linear, the phase path has basically the same shape in the u-w plane (Figure 3-10a) and x-$\dot{x}$ plane (Figure 3-10b). They both show a spiral phase path of the underdamped oscillator The continually decreasing magnitude of the radius vector for a representative point in the phase plane always indicates damped motion of the oscillator.

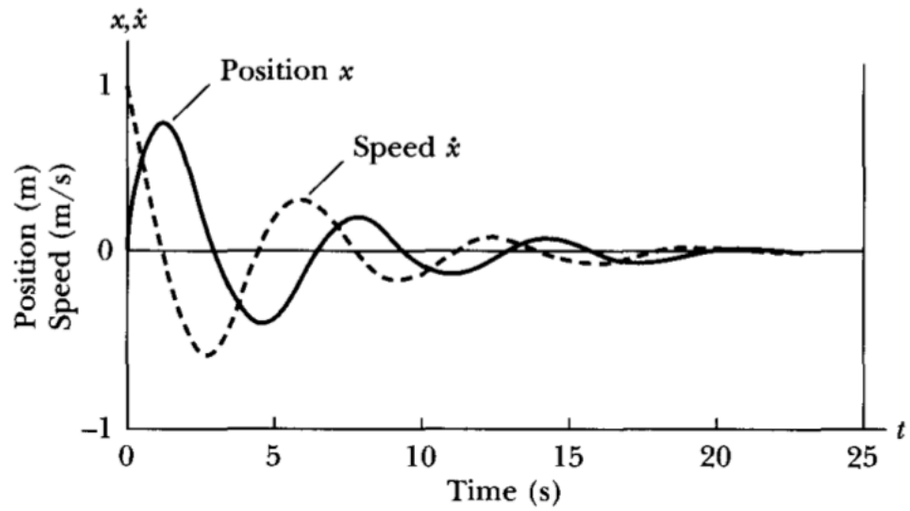
The actual calculation using numbers can be done by various means with a computer. We chose to use one of the commercially available numerical programs that has good graphics output. We chose the values A = 1, β = 0.2, k= 1, m= 1, and $\delta = \pi/2$ in the appropriate units to produce Figure 3-10. For the particular value of δ chosen, the amplitude has x = 0 at t = 0, but $\dot{x}$ has a large positive value, which causes x to rise to a maximum value of about 0.7 m at 2 s (Figure 3-10c). The weak damping parameter β allows the system to oscillate about zero several times (Figure 3-10c) before the system finally spirals down to zero. The system crosses the x = 0 line eleven times before x decreases finally to less than 10^{-3} of its maximum amplitude. The phase diagram of Figure 3-10b displays the actual path.



(a)



(b)



(c)

FIGURE 3-10 Results for Example 3.2. The phase path (a) of the w, u coordinates and (b) of the $\dot{x}$, x coordinates, and (c) a numerical calculation of position and speed versus time. The spiral path is characteristic of the underdamped oscillator.

Critically Damped Motion

If the damping force is sufficiently large (i.e., if $\beta^2 > \omega_0^2$), the system is prevented from undergoing oscillatory motion. If zero initial velocity occurs, the displacement decreases monotonically from its initial value to the equilibrium position (x= 0). The case of **critical damping** occurs when $\beta^2 = \omega_0^2$. The roots of the auxiliary equation are then equal, and the function x must be written as (cf. Equation C.ll, Appendix C)

$$x(t) = (A + Bt)\, e^{-\beta t} \tag{3.43}$$

This displacement curve for critical damping is shown in Figure 3-6 for the case in which the initial velocity is zero. For a given set of initial conditions, a critically damped oscillator will approach equilibrium at a rate more rapid than that for either an overdamped or an underdamped oscillator. This is important in designing certain practical oscillatory systems (e.g., galvanometers) when the system must return to equilibrium as rapidly as possible. A pneumatic-tube screen-door closure system is a good example of a device that should be critically damped. If the closure were underdamped, the door would slam shut as other doors with springs always seem to do. If it were overdamped, it might take an unreasonably long time to close.

Overdamped Motion

If the damping parameter β is even larger than ω_0, then overdamping results. Because $\beta^2 > \omega_0^2$, the exponents in the brackets of Equation 3.37 become real quantities:

$$x(t) = e^{-\beta t}[A_1 e^{\omega_2 t} + A_2 e^{-\omega_2 t}] \tag{3.44}$$

where

$$\omega_2 = \sqrt{\beta^2 - \omega_0^2} \tag{3.45}$$

Note that ω_2 does not represent an angular frequency, because the motion is not periodic. The displacement asymptotically approaches the equilibrium position (Figure 3-6).

Overdamping results in a decrease of the amplitude to zero that may have some strange behavior as shown in the phase space diagram of Figure 3-11.

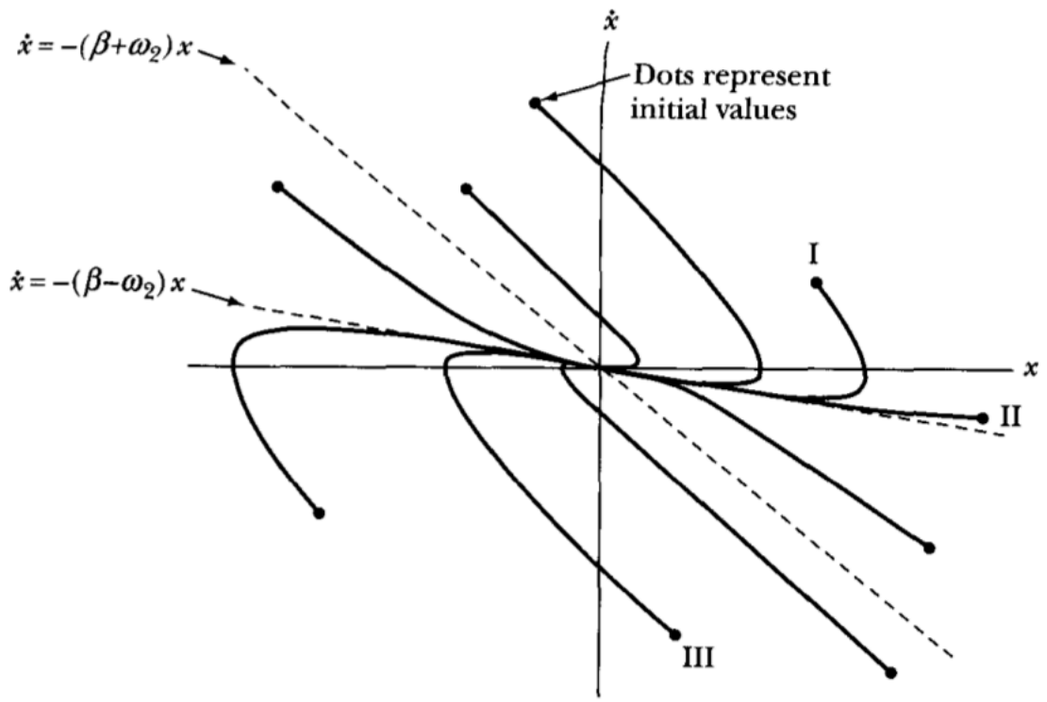


FIGURE 3-11 The phase paths for overdamped motion are shown for several initial values of $(x, \dot{x})$. We examine more closely the paths labeled I, II, and III.

Notice that for all the phase paths of the initial positions shown, the asymptotic paths at longer times are along the dashed curve $\dot{x} = -(\beta - \omega_2)x$. Only a special case has a phase path along the other dashed curve. Depending on the initial values of the position and the velocity, a change in sign of both x and $\dot{x}$ may occur; for example, see the phase path labeled III in Figure 3-11. Figure 3-12 displays X and it as a function of time for the three phase paths labeled I, II, and III in Figure 3-11. All three cases have initial positive displacements, $x(0) = x_0 > 0$. Each of the three phase paths has interesting behavior depending on the initial value, $\dot{x}(0) = \dot{x}_0$, of the velocity:

I. $\dot{x}_0 > 0$, so that x(t) reaches a maximum at some t > 0 before approaching zero. The velocity $\dot{x}$ decreases, becomes negative, and then approaches zero.

II. $\dot{x}_0 < 0$, with x(t) and $\dot{x}$(t) monotonically approaching zero

III. $\dot{x}_0 < 0$, but below the curve $\dot{x} = -(\beta - \omega_2)x$, so that x(t) goes negative before approaching zero, and $\dot{x}$(t) goes positive before approaching zero. The motion in this case could be considered oscillatory.

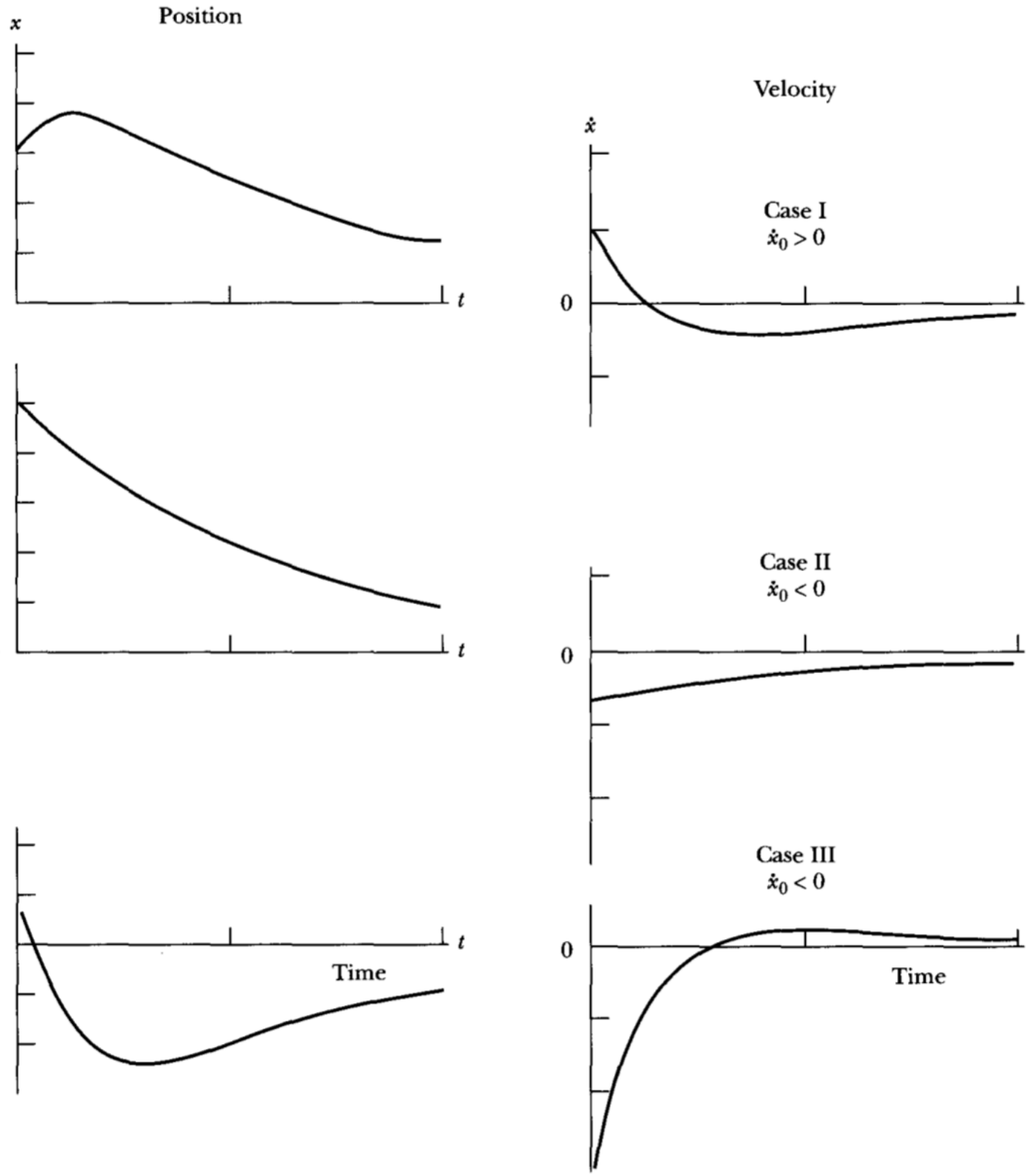


FIGURE 3-12 The position and velocity as functions of time for the three phase paths labeled I, II, and III shown in Figure 3-11.

The initial points lying between the two dashed curves in Figure 3-11 seem to The initial points lying between the two dashed curves in Figure 3-11 seem to have phase paths decreasing monotonically to zero, whereas those lying outside those two lines do not. Critical damping has phase paths similar to the overdamping curves shown in Figure 3-11, rather than the spiral paths of Figure 3-l0b.

EXAMPLE 3.3

Consider a pendulum of length ℓ and a bob of mass m at its end (Figure 3-13) moving through oil with θ decreasing. The massive bob undergoes small oscillations, but the oil retards the bob's motion with a resistive force proportional to the speed with $F_{res} = 2m\sqrt{g/\ell}(\ell\dot{\theta})$. The bob is initially pulled back at t = 0 with $\theta = \alpha$ and $\dot{\theta} = 0$. Find the angular displacement θ and velocity $\dot{\theta}$ as a function of time. Sketch the phase diagram if $\sqrt{g/\ell}$ = 10 s^{-1} and α = 10^{-2} rad.

Solution. Gravity produces the restoring force, and the component pulling the bob back to equilibrium is mgsin 6. Newton's Second Law becomes

$$\text{Force} = m(\ell\ddot{\theta}) = \text{Restoring force} + \text{Resistive force}$$

$$m\ell\ddot{\theta} = -mg\sin\theta - 2m\sqrt{g/\ell}(\ell\dot{\theta}) \tag{3.46}$$

Check that the force direction is correct, depending on the signs of θ and $\dot{\theta}$ For small oscillations $\sin\theta \approx \theta$, and Equation 3.46 becomes

$$\ddot{\theta} + 2\sqrt{g/\ell}\,\dot{\theta} + \frac{g}{\ell}\theta = 0 \tag{3.47}$$

Comparing this equation with Equation 3.35 reveals that $\omega_0^2 = g/\ell$, and $\beta^2 = g/\ell$. Therefore, $\omega_0^2 = \beta^2$ and the pendulum is critically damped. After being initially pulled back and released, the pendulum accelerates and then decelerates as θ goes to zero. The pendulum moves only in one direction as it returns to its equilibrium position.

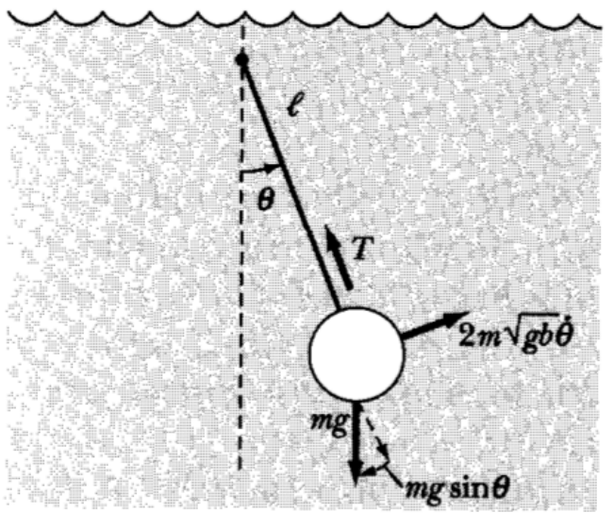


FIGURE 3-13 Example 3.3. The bob is moving with decreasing θ.

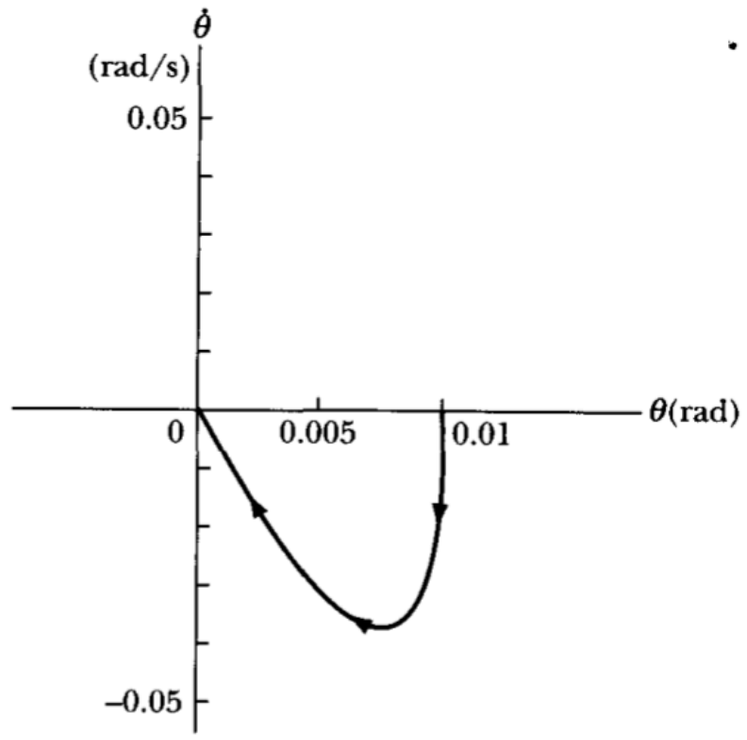


FIGURE 3-14 Phase diagram for Example 3.3.

The solution of Equation 3.47 is Equation 3.43. We can determine the values of A and B by substituting Equation 3.43 into Equation 3.47 using the initial conditions.

$$\begin{aligned} \theta(t) &= (A + Bt)\,e^{-\beta t} \\ \theta(t=0) &= \alpha = A \\ \dot{\theta}(t) &= Be^{-\beta t} - \beta(A + Bt)\,e^{-\beta t} \end{aligned} \tag{3.43}$$

$$\dot{\theta}(t=0) = 0 = B - \beta A$$

$$B = \beta A = \beta\alpha \tag{3.48}$$

$$\theta(t) = \alpha(1 + \sqrt{g/\ell}\; t)e^{-\sqrt{g/\ell}\, t} \tag{3.49}$$

$$\dot{\theta}(t) = \frac{-\alpha g}{\ell}\, te^{-\sqrt{g/\ell}\, t} \tag{3.50}$$

If we calculate $\theta(t)$ and $\dot{\theta}(t)$ for several values of time up to about 0.5 s, we can sketch the phase diagram of Figure 3-14. Notice that Figure 3-14 is consistent with the typical paths shown in Figure 3-11. The angular velocity is always negative after the bob starts until it returns to equilibrium. The bob speeds up quickly and then slows down.

3.6 Sinusoidal Driving Forces

The simplest case of driven oscillation is that in which an external driving force varying harmonically with time is applied to the oscillator. The total force on the particle is then

$$F = -kx - b\dot{x} + F_0 \cos \omega t \tag{3.51}$$

where we consider a linear restoring force and a viscous damping force in addition to the driving force. The equation of motion becomes

$$m\ddot{x} + b\dot{x} + kx = F_0 \cos \omega t \tag{3.52}$$

or, using our previous notation,

$$\boxed{\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = A \cos \omega t} \tag{3.53}$$

where $A = F_0/m$ and where ω is the angular frequency of the driving force. The solution of Equation 3.53 consists of two parts, a **complementary function** $x_c(t)$, which is the solution of Equation 3.53 with the right-hand side set equal to zero, and a **particular solution** $x_p(t)$, which reproduces the right-hand side. The complementary solution is the same as that given in Equation 3.37 (see Appendix C):

$$x_c(t) = e^{-\beta t}\left[A_1 \exp\left(\sqrt{\beta^2 - \omega_0^2}\, t\right) + A_2 \exp\left(-\sqrt{\beta^2 - \omega_0^2}\, t\right)\right] \tag{3.54}$$

For the particular solution, we trv

$$x_p(t) = D \cos(\omega t - \delta) \tag{3.55}$$

Substituting $x_p(t)$ in Equation 3.53 and expanding $\cos(\omega t - \delta)$ and $\sin(\omega t - \delta)$, we obtain

$$\begin{aligned}\{A - D[(\omega_0^2 - \omega^2)\cos\delta + 2\omega\beta\sin\delta]\}\cos\omega t \\ -\{D[(\omega_0^2 - \omega^2)\sin\delta - 2\omega\beta\cos\delta]\}\sin\omega t = 0\end{aligned} \tag{3.56}$$

Because $\sin\omega t$ and $\cos\omega t$ are linearly independent functions, this equation can be satisfied in general only if the coefficient of each term vanishes identically. From the $\sin\omega t$ term, we have

$$\tan\delta = \frac{2\omega\beta}{\omega_0^2 - \omega^2} \tag{3.57}$$

so we can write

$$\left.\begin{aligned}\sin\delta &= \frac{2\omega\beta}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\omega^2\beta^2}} \\ \cos\delta &= \frac{\omega_0^2 - \omega^2}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\omega^2\beta^2}}\end{aligned}\right\} \tag{3.58}$$

And from the coefficient of the $\cos\omega t$ term, we have

$$\begin{aligned}D &= \frac{A}{(\omega_0^2 - \omega^2)\cos\delta + 2\omega\beta\sin\delta} \\ &= \frac{A}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\omega^2\beta^2}}\end{aligned} \tag{3.59}$$

Thus, the particular integral is

$$\boxed{x_p(t) = \frac{A}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\omega^2\beta^2}}\cos(\omega t - \delta)} \tag{3.60}$$

with

$$\boxed{\delta = \tan^{-1}\left(\frac{2\omega\beta}{\omega_0^2 - \omega^2}\right)} \tag{3.61}$$

The quantity δ represents the phase difference between the driving force and the resultant motion; a real delay occurs between the action of the driving force and the response of the system. For a fixed ω_0, as ω increases from 0, the phase increases from $\delta = 0$ at $\omega = 0$ to $\delta = \pi/2$ at $\omega = \omega_0$ and to π as $\omega \to \infty$. The variation of δ with ω is shown later in Figure 3-16.

The general solution is

$$x(t) = x_c(t) + x_p(t) \tag{3.62}$$

But $x_c(t)$ here represents transient effects (i.e., effects that die out), and the terms contained in this solution damp out with time because of the factor exp $(-\beta t)$. The term $x_p(t)$ represents the steady-state effects and contains all the information for t large compared with $1/\beta$. Thus,

$$x(t \gg 1/\beta) = x_p(t)$$

The steady-state solution is important in many applications and problems (see Section 3.7).

The details of the motion during the period before the transient effects have disappeared (i.e., $t \leq 1/\beta$) strongly depend on the oscillator's conditions at the time that the driving force is first applied and also on the relative magnitudes of the driving frequency ω and the damping frequency $\sqrt{\omega_0^2 - \beta^2}$ in the case of underdamped, undriven oscillations. This can be shown by numerically calculating $x_p(t)$, $x_c(t)$, and the sum x(t) (see Equation 3.62) for different values of β and ω as we have done for Figure 3-15.

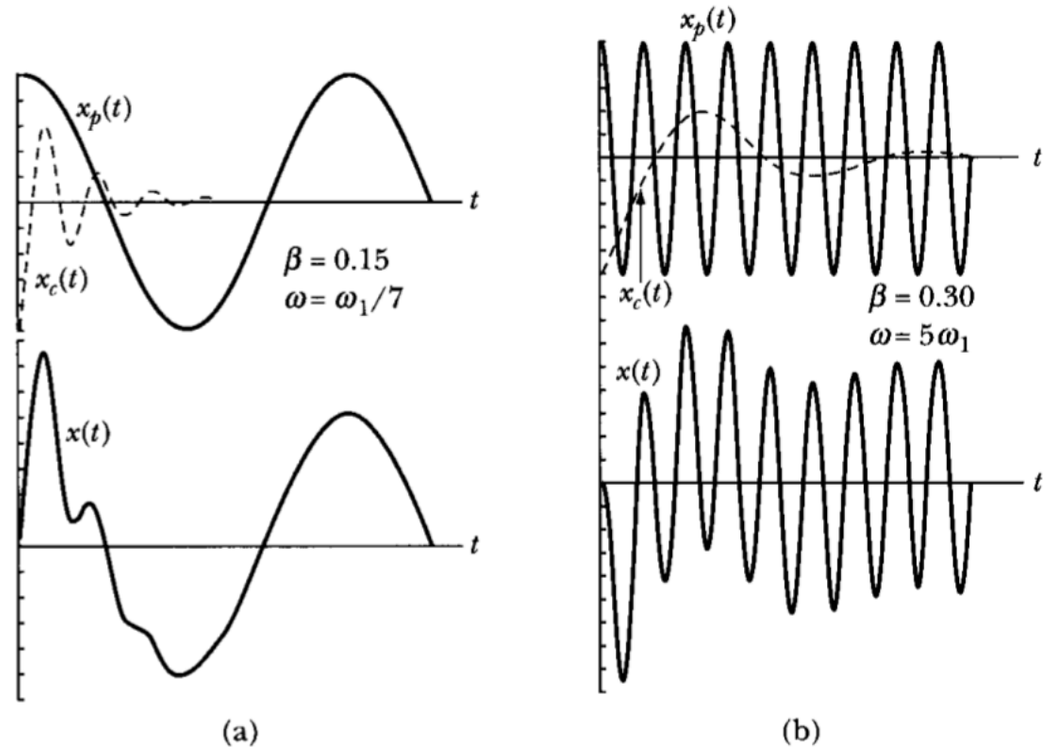


FIGURE 3-15 Examples of sinusoidal driven oscillatory motion with damping. The steady-state solution x_p, transient solution x_c, and sum x are shown in (a) for driving frequency ω greater than the damping frequency $\omega_1 (\omega > \omega_1)$ and in (b) for $\omega < \omega_1$.

Figure 3-15 illustrates the transient motion of an underdamped oscillator when driving frequencies less than and greater than $\omega_1 = \sqrt{\omega_0^2 - \beta^2}$ are applied. If $\omega < \omega_1$ (Figure 3-15a), the transient response of the oscillator greatly distorts the sinusoidal shape of the forcing function during the time interval immediately after the application of the driving force, whereas if $\omega > \omega_1$ (Figure 3-15b), the effect is a modulation of the forcing function with little distortion of the high-frequency sinusoidal oscillations.

The steady-state solution (x_p) is widely studied in many applications and problems (see Section 3.7). The transient effects (x_c), although perhaps not as important overall, must be understood and accounted for in many cases, especially in certain types of electrical circuits.

Resonance Phenomena

To find the angular frequency ω_R at which the amplitude D (Equation 3.59) is a maximum (i.e., the **amplitude resonance frequency**), we set

$$\left.\frac{dD}{d\omega}\right|_{\omega=\omega_R} = 0$$

Performing the differentiation, we find

$$\omega_R = \sqrt{\omega_0^2 - 2\beta^2} \tag{3.63}$$

Thus, the resonance frequency ω_R is lowered as the damping coefficient β is increased. No resonance occurs if $\beta > \omega_0/2$, for then ω_R is imaginary and D decreases monotonically with increasing ω.

We may now compare the oscillation frequencies for the various cases we have considered:

1. **Free oscillations, no damping (Equation 3.4):**

$$\omega_0^2 = \frac{k}{m}$$

2. **Free oscillations, damping (Equation 3.38):**

$$\omega_1^2 = \omega_0^2 - \beta^2$$

3. **Driven oscillations, damping (Equation 3.63):**

$$\omega_R^2 = \omega_0^2 - 2\beta^2$$

and we note that $\omega_0 > \omega_1 > \omega_R$.

We customarily describe the degree of damping in an oscillating system in terms of the "quality factor" Q of the system:

$$Q \equiv \frac{\omega_R}{2\beta} \tag{3.64}$$

If little damping occurs, then Q is very large and the shape of the resonance curve approaches that for an undamped oscillator But the resonance can be completely destroyed if the damping is large and Q is very small. Figure 3-16 shows the resonance and phase curves for several different values of Q.

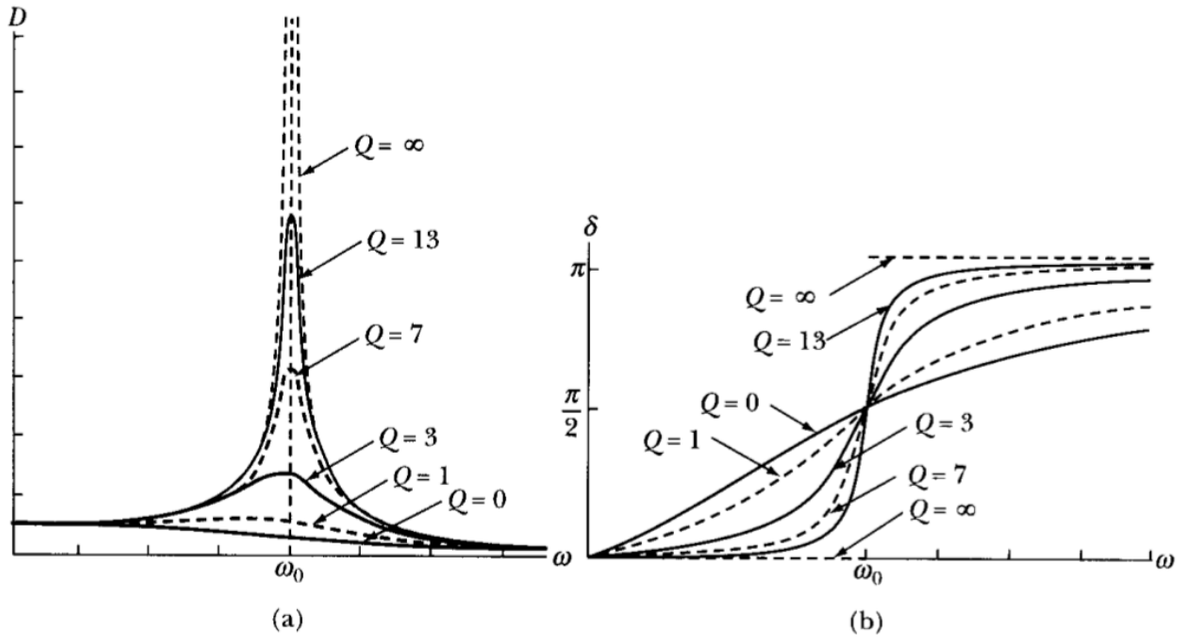


FIGURE 3-16 (a) The amplitude D is displayed as a function of the driving frequency ω for various values of the quality factor Q. Also shown is (b) the phase angle δ, which is the phase angle between the driving force and the resultant motion.

These curves indicate the lowering of the resonance frequency with a decrease in Q (i.e., with an increase of the damping coefficient β . The effect is not large, however; the frequency shift is less than 3% even for Q as small as 3 and is about 18% for Q= 1.

For a lightly damped oscillator, we can show that

$$Q \cong \frac{\omega_0}{\Delta\omega} \tag{3.65}$$

where $\Delta\omega$ represents the frequency interval between the points on the amplitude resonance curve that are $1/\sqrt{2}$ $= 0.707$ of the maximum amplitude.

The values of Q found in real physical situations vary greatly. In rather ordinary mechanical systems (e.g., loudspeakers), the values may be in the range from a few to 100 or so. Quartz crystal oscillators or tuning forks may have Qs of 10^4. Highly tuned electrical circuits, including resonant cavities, may have values of 10^4 to 10^5. We may also define Qs for some atomic systems. According to the classical picture, the oscillation of electrons within atoms leads to optical radiation. The sharpness of spectral lines is limited by the damping due to the loss of energy by radiation (**radiation damping**). The minimum width of a line can be calculated classically and is $\Delta\omega = 2 \times 10^{-8}\ \omega$. The Q of such an oscillator is therefore approximately 5×10^7. Resonances with the largest known Qs occur in the radiation from gas lasers. Measurements with such devices have yielded Qs of approximately 10^{14}.

Equation 3.63 gives the frequency for *amplitude* resonance. We now calculate the frequency for *kinetic energy* resonance—that is, the value of ω for which T is a maximum. The kinetic energy is given by $T = \frac{1}{2}m\dot{x}^2$, and computing $\dot{x}$ from Equation 3.60, we have

$$\dot{x} = \frac{-A\omega}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\omega^2\beta^2}} \sin(\omega t - \delta) \tag{3.66}$$

so that the kinetic energy becomes

$$T = \frac{mA^2}{2} \cdot \frac{\omega^2}{(\omega_0^2 - \omega^2)^2 + 4\omega^2\beta^2} \sin^2(\omega t - \delta) \tag{3.67}$$

To obtain a value of T independent of the time, we compute the average of T over one complete period of oscillation:

$$\langle T \rangle = \frac{mA^2}{2} \cdot \frac{\omega^2}{(\omega_0^2 - \omega^2)^2 + 4\omega^2\beta^2} \langle \sin^2(\omega t - \delta) \rangle \tag{3.68}$$

The average value of the square of the sine function taken over one period is

$$\langle \sin^2(\omega t - \delta) \rangle = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} \sin^2(\omega t - \delta)\, dt = \frac{1}{2} \tag{3.69}$$

Therefore,

$$\langle T \rangle = \frac{mA^2}{4} \cdot \frac{\omega^2}{(\omega_0^2 - \omega^2)^2 + 4\omega^2\beta^2} \tag{3.70}$$

The value of ω for $\langle T \rangle$ a maximum is labeled ω_E and is obtained from

$$\left.\frac{d\langle T \rangle}{d\omega}\right|_{\omega = \omega_E} = 0 \tag{3.71}$$

Differentiating Equation 3.70 and equating the result to zero, we find

$$\omega_E = \omega_0 \tag{3.72}$$

so the kinetic energy resonance occurs at the natural frequency of the system for undamped oscillations.

We see therefore that the amplitude resonance occurs at a frequency $\sqrt{\omega_0^2 - 2\beta^2}$, whereas the kinetic energy resonance occurs at ω_0. Because the potential energy is proportional to the square of the amplitude, the potential energy resonance must also occur at $\sqrt{\omega_0^2 - 2\beta^2}$. That the kinetic and potential energies resonate at different frequencies is a result of the fact that the damped oscillator is not a conservative system. Energy is continually exchanged with the driving mechanism, and energy is being transferred to the damping medium.

3.7 Physical Systems

We stated in the introduction to this chapter that linear oscillations apply to more systems than just the small oscillations of the mass-spring and the simple pendulum. The same mathematical formulation applies to a whole host of physical systems. Mechanical systems include the torsion pendulum, vibrating string or membrane, and elastic vibrations of bars or plates. These systems may have overtones, and each overtone can be treated much the same as we did in the previous discussion.

We can apply our mechanical system analog to acoustic systems. In this case, the air molecules vibrate. We can have resonances that depend on the properties and dimensions of the medium. Several factors cause the damping, including friction and sound-wave radiation. The driving force can be a tuning fork or vibrating string, among many sources of sound.

Atomic systems can also be represented classically as linear oscillators. When light (consisting of electromagnetic radiation of high frequency) falls on matter, it causes the atoms and molecules to vibrate. When light having one of the resonant frequencies of the atomic or molecular system falls on the material, electromagnetic energy is absorbed, causing the atoms or molecules to oscillate with large amplitude. Large electromagnetic fields of the same frequency are produced by the oscillating electric charges. Wave mechanics (or quantum mechanics) uses linear oscillator theory to explain many of the phenomena associated with light absorption, dispersion, and radiation.

Even to describe nuclei, linear oscillator theory is used. One of the modes of excitation of nuclei is collective excitation. Neutrons and protons vibrate in various collective motions. Resonances occur, and damping exists. The classical mechanical analog is very useful in describing the motion.

Electrical circuits are, however, the most noted examples of nonmechanical oscillations. Indeed, because of its great practical importance, the electrical example has been so thoroughly investigated that the situation is frequently reversed, and mechanical vibrations are analyzed in terms of the "equivalent electrical circuit." We devote two examples to electrical circuits.

EXAMPLE 3.4

Find the equivalent electrical circuit for the hanging mass-spring shown in Figure 3-17a and determine the time dependence of the charge q in the system.

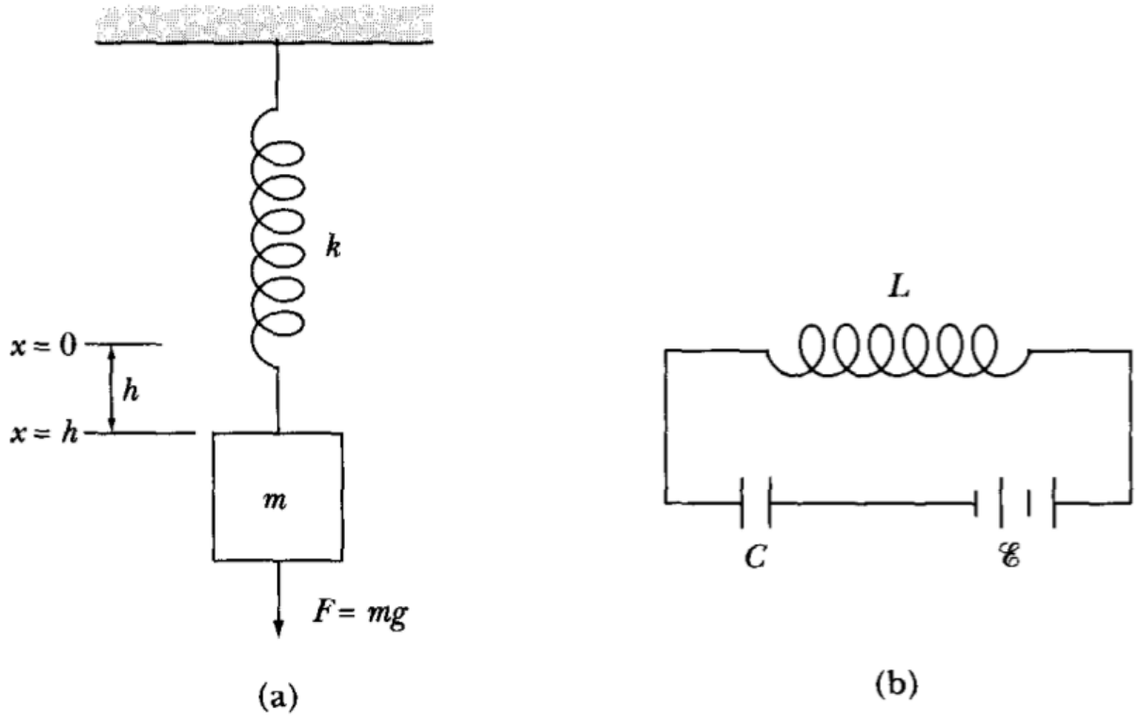


FIGURE 3-17 Example 3.4 (a) hanging mass–spring system; (b) equivalent electrical circuit.

Solution. Let us first consider the analogous quantities in mechanical and electrical systems. The force F{= mg the mechanical case) is analogous to the emf $\mathcal{E}$. The damping parameter b has the electrical analog resistance R, which is not present in this case. The displacement x has the electrical analog charge q. We show other quantities in Table 3-1.

TABLE 3-1 Analogous Mechanical and Electrical Quantities

Mechanical		Electrical	
x	Displacement	q	Charge
$\dot{x}$	Velocity	$\dot{q} = I$	Current
m	Mass	L	Inductance
b	Damping resistance	R	Resistance
$1/k$	Mechanical compliance	C	Capacitance
F	Amplitude of impressed force	$\mathcal{E}$	Amplitude of impressed emf

If we examine Figure 3-17a, we have 1/k —> C, m -> L, F —> $\mathcal{E}$, x —> q, and $\dot{x}$ —> I. Without the weight of the mass, the equilibrium position would be at x = 0; the addition of the gravitational force extends the spring by an amount h = mg/k and displaces the equilibrium position to x = h. The equation of motion becomes

$$m\ddot{x} + k(x - h) = 0 \tag{3.73}$$

or

$$m\ddot{x} + kx = kh$$

with solution

$$x(t) = h + A\cos\omega_0 t \tag{3.74}$$

where we have chosen the initial conditions x(t = 0) = h + A and $\dot{x}$(t = 0) =0.

We draw the equivalent electrical circuit in Figure 3-17b. Kirchoff's equation around the circuit becomes

$$L\frac{dI}{dt} + \frac{1}{C}\int I\,dt = \mathcal{E} = \frac{q_1}{C} \tag{3.75}$$

where q_1 represents the charge that must be applied to C to produce a voltage $\mathcal{E}$.

If we use I = $\dot{q}$, we have

$$L\ddot{q} + \frac{q}{C} = \frac{q_1}{C} \tag{3.76}$$

If q = q_0 and I = 0 at t = 0, the solution is

$$q(t) = q_1 + (q_0 - q_1)\cos\omega_0 t \tag{3.77}$$

which is the exact electrical analog of Equation 3.74.

EXAMPLE 3.5

Consider the series RLC circuit shown in Figure 3-18 driven by an alternating emf of value $E_0 \sin \omega t$. Find the current, the voltage V_L across the inductor, and the angular frequency ω at which V_L is a maximum.

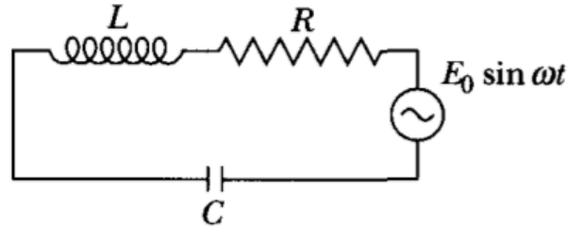


FIGURE 3-18 Example 3.5. RLC circuit with an alternating emf.

Solution. The voltage across each of the circuit elements in Figure 3-18 are

$$V_L = L\frac{dI}{dt} = L\ddot{q}$$

$$V_R = LI = L\frac{dq}{dt} = L\dot{q}$$

$$V_C = \frac{q}{C}$$

so the voltage drops around the circuit become

$$L\ddot{q} + R\dot{q} + \frac{q}{C} = E_0 \sin \omega t$$

We identify this equation as similar to Equation 3.53, which we have already solved. In addition to the relationships in Table 3-1, we also have $\beta = b/2m \rightarrow R/2L$, $\omega_0 = \sqrt{k/m} \rightarrow 1/\sqrt{LC}$, and $A = F_0/m \rightarrow E_0/L$. The solution for the charge q is given by transcribing Equation 3.60, and the equation for the current I is given by transcribing Equation 3.66, which allows us to write

$$I = \frac{-E_0}{\sqrt{R^2 + \left(\frac{1}{\omega C} - \omega L\right)^2}} \sin(\omega t - \delta)$$

where δ can be found by transcribing Equation 3.61.

The voltage across the inductor is found from the time derivative of the current.

$$V_L = L\frac{dI}{dt} = \frac{-\omega L E_0}{\sqrt{R^2 + \left(\frac{1}{\omega C} - \omega L\right)^2}} \cos(\omega t - \delta)$$

$$= V(\omega)\cos(\omega t - \delta)$$

To find the driving frequency ω_{max} which makes V_L a maximum, we must take the derivative of V_L with respect to ω and set the result equal to zero. We only need to consider the amplitude V(ω) and not the time dependence.

$$\frac{dV(\omega)}{d\omega} = \frac{LE_0\left(R^2 - \frac{2L}{C} + \frac{2}{\omega^2 C^2}\right)}{\left[R^2 + \left(\frac{1}{\omega C} - \omega L\right)^2\right]^{3/2}}$$

We have skipped a few intermediate steps to arrive at this result. We determine the value ω_{max} sought by setting the term in parentheses in the numerator equal to zero. By doing so and solving for ω_{max} gives

$$\omega_{max} = \frac{1}{\sqrt{LC - \frac{R^2C^2}{2}}}$$

which is the result we need. Note the difference between this frequency and those given by the natural frequency, $\omega_0 = 1/\sqrt{LC}$, and the charge resonance frequency (given by transcribing Equation 3.63), $\omega_R = \sqrt{1/LC - 2R^2/L^2}$.

3.8 Principle of Superposition - Fourier Series

The oscillations we have been discussing obey a differential equation of the form

$$\left(\frac{d^2}{dt^2} + a\frac{d}{dt} + b\right)x(t) = A\cos\omega t \tag{3.78}$$

The quantity in parentheses on the left-hand side is a **linear operator**, which we may represent by **L**. If we generalize the time-dependent forcing function on the right-hand side, we can write the equation of motion as

$$\mathbf{L}x(t) = F(t) \tag{3.79}$$

An important property of linear operators is that they obey the **principle of superposition**. This property results from the fact that linear operators are distributive, that is,

$$\mathbf{L}(x_1 + x_2) = \mathbf{L}(x_1) + \mathbf{L}(x_2) \tag{3.80}$$

Therefore, if we have two solutions, $x_1(t)$ and $x_2(t)$, for two different forcing functions, $F_i(t)$ and $F_2(t)$,

$$\mathbf{L}x_1 = F_1(t), \qquad \mathbf{L}x_2 = F_2(t) \tag{3.81}$$

we can add these equations (multiplied by arbitrary constants α_1 and α_2 and obtain

$$\mathbf{L}(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 F_1(t) + \alpha_2 F_2(t) \tag{3.82}$$

We can extend this argument to a set of solutions $x_n(t)$, each of which is appropriate for a given $F_n(t)$:

$$\mathbf{L}\left(\sum_{n=1}^{N}\alpha_n x_n(t)\right) = \sum_{n=1}^{N}\alpha_n F_n(t) \tag{3.83}$$

This equation is just Equation 3.79 if we identify the linear combinations as

$$x(t) = \sum_{n=1}^{N} \alpha_n x_n(t)$$
$$F(t) = \sum_{n=1}^{N} \alpha_n F_n(t) \tag{3.84}$$

If each of the individual functions $F_n(t)$ has a simple harmonic dependence on time, such as cos ω_n t, we know that the corresponding solution $x_n(t)$ is given by Equation 3.60. Thus, if F(t) has the form

$$F(t) = \sum_n \alpha_n \cos(\omega_n t - \phi_n) \tag{3.85}$$

the steady-state solution is

$$x(t) = \frac{1}{m} \sum_n \frac{\alpha_n}{\sqrt{(\omega_0^2 - \omega_n^2)^2 + 4\omega_n^2 \beta^2}} \cos(\omega_n t - \phi_n - \delta_n) \tag{3.86}$$

where

$$\delta_n = \tan^{-1}\left(\frac{2\omega_n \beta}{\omega_0^2 - \omega_n^2}\right) \tag{3.87}$$

We can write down similar solutions where F(t) is represented by a series of terms, sin($\omega_n t - \phi_n$). We therefore arrive at the important conclusion that if some arbitrary forcing function F(t) can be expressed as a series (finite or infinite) of harmonic terms, the complete solution can also be written as a similar series of harmonic terms.

This is an extremely useful result, because, according to Fourier's theorem, any arbitrary periodic function (subject to certain conditions that are not very restrictive) can be represented by a series of harmonic terms. Thus, in the usual physical case in which F(t) is periodic with period $\tau = 2\pi/\omega$,

$$F(t + \tau) = F(t) \tag{3.88}$$

we then have

$$F(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos n\omega t + b_n \sin n\omega t) \tag{3.89}$$

where

$$\left.\begin{aligned} a_n &= \frac{2}{\tau}\int_0^{\tau} F(t')\cos n\omega t'\, dt' \\ b_n &= \frac{2}{\tau}\int_0^{\tau} F(t')\sin n\omega t'\, dt' \end{aligned}\right\} \tag{3.90}$$

or, because F{t) has a period t, we can replace the integral limits 0 and t by the limits ~"9 ^ ~ —tt/co and +-^t = + tt/co:

$$\left.\begin{aligned} a_n &= \frac{\omega}{\pi}\int_{-\pi/\omega}^{+\pi/\omega} F(t')\cos n\omega t'\, dt' \\ b_n &= \frac{\omega}{\pi}\int_{-\pi/\omega}^{+\pi/\omega} F(t')\sin n\omega t'\, dt' \end{aligned}\right\} \tag{3.91}$$

Before we discuss the response of damped systems to arbitrary forcing functions (in the following section), we give an example of the Fourier representation of periodic functions.

EXAMPLE 3.6

A sawtooth driving force function is shown in Figure 3-19.

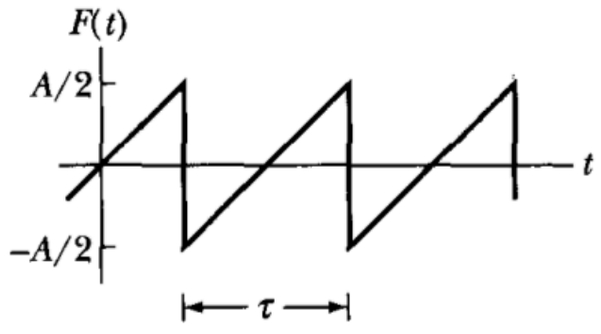


FIGURE 3-19 Example 3.6. A sawtooth driving force function.

Find the coefficients a_n and b_n, and express F(t) as a Fourier series.

Solution. In this case, F(t) is an odd function, and is expressed by

$$F(t) = A \cdot \frac{t}{\tau} = \frac{\omega A}{2\pi} t, \quad -\tau/2 < t < \tau/2 \tag{3.92}$$

Because F(t) is odd, the coefficients a_n all vanish identically. The b_n are given by

$$b_n = \frac{\omega^2 A}{2\pi^2} \int_{-\pi/\omega}^{+\pi/\omega} t' \sin n\omega t' \, dt'$$

$$= \frac{\omega^2 A}{2\pi^2} \left[- \frac{t' \cos n\omega t'}{n\omega} + \frac{\sin n\omega t'}{n^2\omega^2} \right] \Bigg|_{-\pi/\omega}^{+\pi/\omega}$$

$$= \frac{\omega^2 A}{2\pi^2} \cdot \frac{2\pi}{n\omega^2} \cdot (-1)^{n+1} = \frac{A}{n\pi}(-1)^{n+1} \tag{3.93}$$

where the term $(-1)^{n+1}$ takes account of the fact that

$$-\cos n\pi = \begin{cases} +1, & n \text{ odd} \\ -1, & n \text{ even} \end{cases} \tag{3.94}$$

Therefore we have

$$F(t) = \frac{A}{\pi} \left[\sin \omega t - \frac{1}{2} \sin 2\omega t + \frac{1}{3} \sin 3\omega t - \cdots \right] \tag{3.95}$$

Figure 3-20 shows the results for two terms, five terms, and eight terms of this expansion. The convergence toward the sawtooth function is none too rapid.

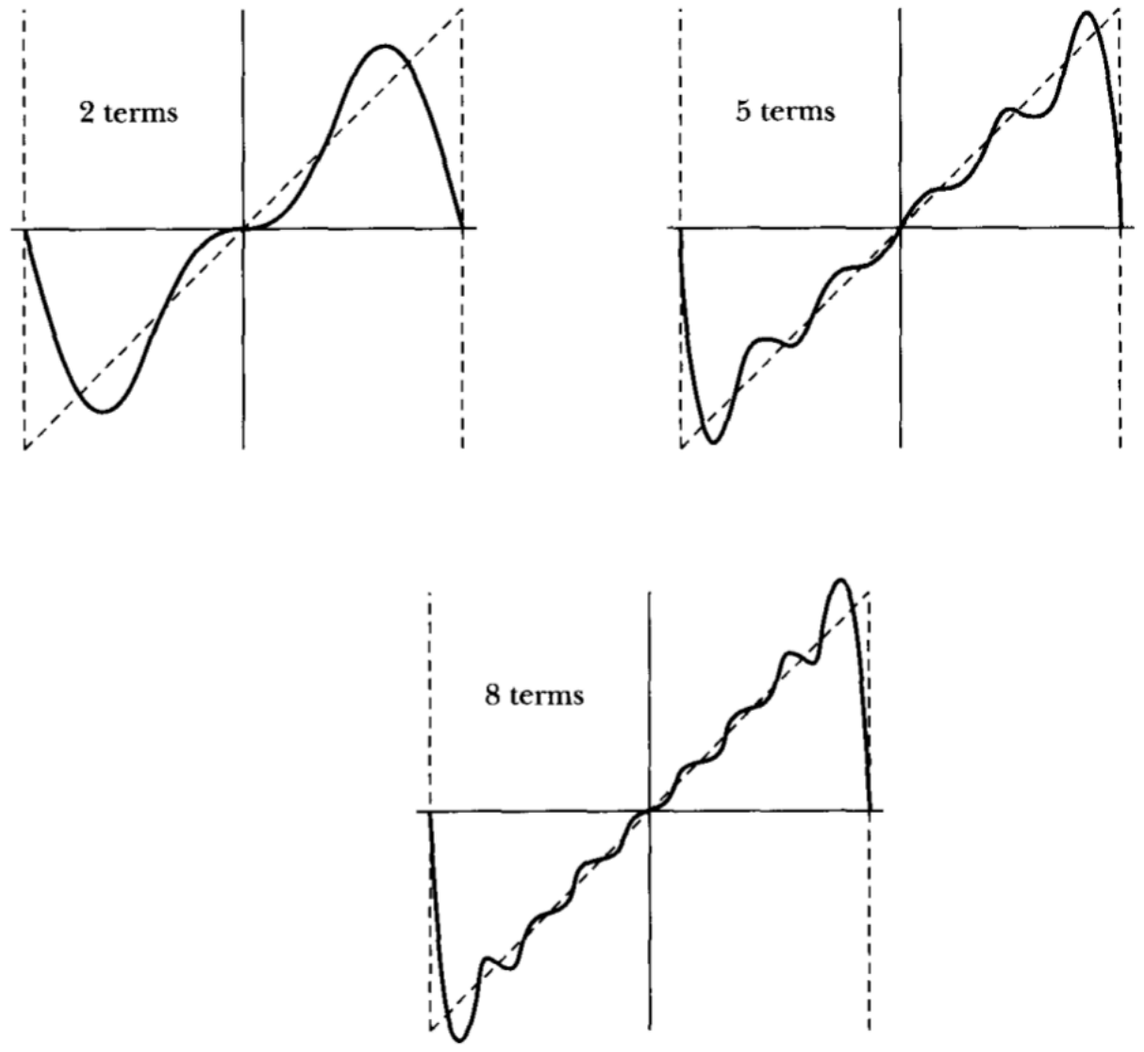


FIGURE 3-20 Results of Example 3.6. Fourier series representation of sawtooth driving force function.

We should note two features of the expansion. At the points of discontinuity (t = $\pm\tau/2$) the series yields the mean value (zero), and in the region immediately adjacent to the points of discontinuity, the expansion "overshoots" the original function. This latter effect, known as the **Gibbs phenomenon**, occurs in all orders of approximation. The Gibbs overshoot amounts to about 9% on each side of *any* discontinuity, even in the limit of an infinite series.

3.9 The Response of Linear Oscillators to Impulsive Forcing Functions (Optional)

In the previous discussions, we have mainly considered steady-state oscillations. or many types of physical problems (particularly those involving oscillating electrical circuits), the transient effects are quite important. Indeed, the transient solution may be of dominating interest in such cases. In this section, we investigate the transient behavior of a linear oscillator subjected to a driving force that acts discontinuously. Of course, a "discontinuous" force is an idealization, because it always takes a finite time to apply a force. But if the application time is small compared with the natural period of the oscillator, the result of the ideal case is a close approximation to the actual physical situation.

The differential equation describing the motion of a damped oscillator is

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = \frac{F(t)}{m} \tag{3.96}$$

The general solution is composed of the complementary and particular solutions:

$$x(t) = x_c(t) + x_p(t) \tag{3.97}$$

We can write the complementary solution as

$$x_c(t) = e^{-\beta t}(A_1 \cos \omega_1 t + A_2 \sin \omega_1 t) \tag{3.98}$$

where

$$\omega_1 \equiv \sqrt{\omega_0^2 - \beta^2} \tag{3.99}$$

The particular solution $x_p(t)$ depends on the nature of the forcing function F(t).

Two types of idealized discontinuous forcing functions are of considerable interest. These are the **step function** (or **Heaviside function**) and the **impulse function**, shown in Figures 3-21a and b, respectively.

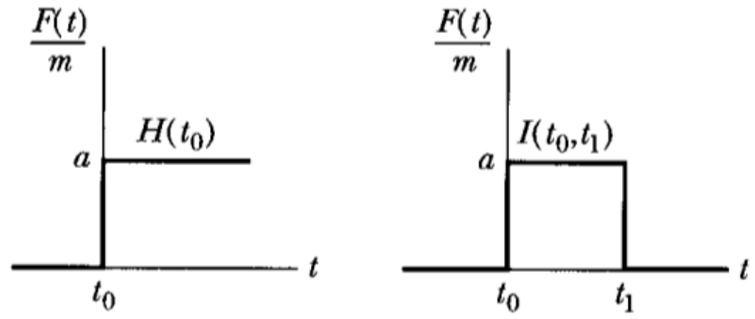


FIGURE 3-21 (a) Step function; (b) impulse function.

The step function H is given by

$$H(t_0) = \begin{cases} 0, & t < t_0 \\ a, & t > t_0 \end{cases} \tag{3.100}$$

where a is a constant with the dimensions of acceleration and where the argument t_0 indicates that the time of application of the force is $t = t_0$.

The impulse function I is a positive step function applied at $t = t_0$, followed by a negative step function applied at some later time t_1. Thus

$$I(t_0, t_1) = H(t_0) - H(t_1)$$

$$I(t_0, t_1) = \begin{cases} 0, & t < t_0 \\ a, & t_0 < t < t_1 \\ 0, & t > t_1 \end{cases} \tag{3.101}$$

Although we write the Heaviside and impulse functions as $H(t_0)$ and $I(t_0, t_1)$ for simplicity, these functions depend on the time t and are more properly written as $H(t; t_0)$ and $I(t; t_0, t_1)$.

Response to a Step Function

For step functions, the differential equation that describes the motion for $t > t_0$ is

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = a, \quad t > t_0 \tag{3.102}$$

We consider the initial conditions to be $x(t_0) = 0$ and $\dot{x}(t_0) = 0$. The particular solution is just a constant, and examination of Equation 3.102 shows that it must be a/ω_0^2 . Thus, the general solution for $t > t_0$ is

$$x(t) = e^{-\beta(t-t_0)}[A_1 \cos \omega_1(t - t_0) + A_2 \sin \omega_1(t - t_0)] + \frac{a}{\omega_0^2} \tag{3.103}$$

Applying the initial conditions yields

$$A_1 = -\frac{a}{\omega_0^2}, \quad A_2 = -\frac{\beta a}{\omega_1 \omega_0^2} \tag{3.104}$$

Therefore, for $t > t_0$, we have

$$x(t) = \frac{a}{\omega_0^2}\left[1 - e^{-\beta(t-t_0)} \cos \omega_1(t - t_0) - \frac{\beta e^{-\beta(t-t_0)}}{\omega_1} \sin \omega_1(t - t_0)\right] \tag{3.105}$$

and $x(t) = 0$ for $t < t_0$.

If, for simplicity, we take $t_0 = 0$, the solution can be expressed as

$$x(t) = \frac{H(0)}{\omega_0^2}\left[1 - e^{-\beta t} \cos \omega_1 t - \frac{\beta e^{-\beta t}}{\omega_1} \sin \omega_1 t\right] \tag{3.106}$$

This response function is shown in Figure 3-22 for the case $\beta = 0.2\omega_0$.

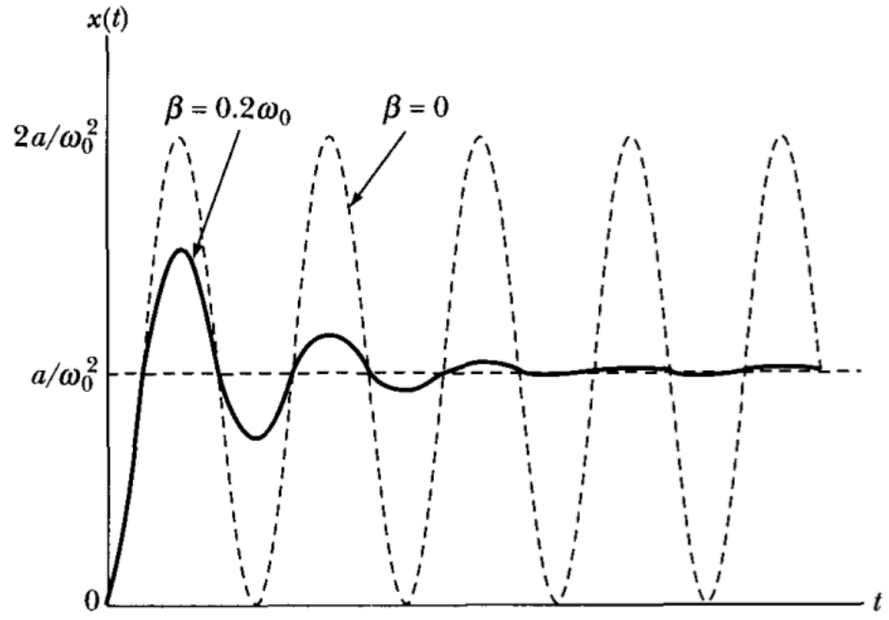


FIGURE 3-22 Response function solution to the step force function.

It should be clear that the ultimate condition of the oscillator (i.e., the steady-state condition) is simply a displacement by an amount a/ω_0^2 .

If no damping occurs, $\beta = 0$ and $\omega_1 = \omega_0$. Then, for $t_0 = 0$, we have

$$x(t) = \frac{H(0)}{\omega_0^2}[1 - \cos \omega_0 t], \quad \beta = 0 \tag{3.107}$$

The oscillation is thus sinusoidal with amplitude extremes x = 0 and $x = 2\,a/\omega_0^2$ (see Figure 3.22).

Response to an Impulse Function

If we consider the impulse function as the difference between two step functions separated by a time $t_1 = t_0 = \tau$, then, because the system is linear, the general solution for $t > t_1$ is given by the superposition of the solutions (Equation 3.105) for the two step functions taken individually:

$$x(t) = \frac{a}{\omega_0^2}\left[1 - e^{-\beta(t-t_0)}\cos\omega_1(t-t_0) - \frac{\beta e^{-\beta(t-t_0)}}{\omega_1}\sin\omega_1(t-t_0)\right]$$

$$-\frac{a}{\omega_0^2}\left[1 - e^{-\beta(t-t_0-\tau)}\cos\omega_1(t-t_0-\tau) - \frac{\beta e^{-\beta(t-t_0-\tau)}}{\omega_1}\sin\omega_1(t-t_0-\tau)\right]$$

$$= \frac{ae^{-\beta(t-t_0)}}{\omega_0^2}\left[e^{\beta\tau}\cos\omega_1(t-t_0-\tau) - \cos\omega_1(t-t_0)\right.$$

$$\left. + \frac{\beta e^{\beta\tau}}{\omega_1}\sin\omega_1(t-t_0-\tau) - \frac{\beta}{\omega_1}\sin\omega_1(t-t_0)\right], \qquad t > t_1 \tag{3.108}$$

The *total* response (i.e., Equations 3.105 and 3.108) to an impulse function of duration $\tau = 5 \times 2\pi/\omega_1$ applied at t = t_0 is shown in Figure 3-23 for $\beta = 0.2\omega_0$.

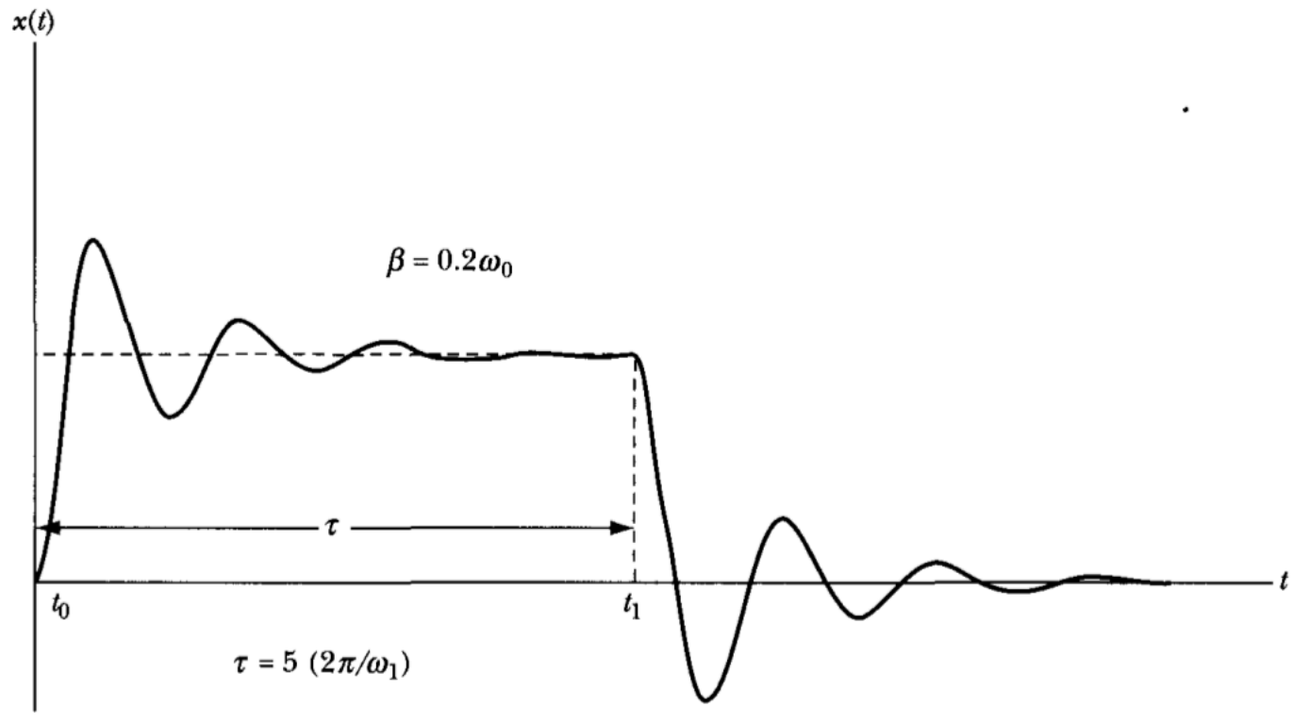


FIGURE 3-23 Response function solution to the impulse force function.

If we allow the duration τ of the impulse function to approach zero, the response function will become vanishingly small. But if we allow $a \to \infty$ as $\tau \to 0$ so that the product $a\tau$ is constant, then the response will be finite. This particular limiting case is considerably important, because it approximates the application of a driving force that is a "spike" at t = t_0 (i.e., $\tau \ll 2\pi/\omega_1$). We want to expand Equation 3.108 by letting $\tau \to 0$, but with $b = a\tau$ = constant. Let A = t - t_0 and B= t, then use Equations D.ll and D.12 (from Appendix D) to obtain

$$x(t) = \frac{ae^{-\beta(t-t_0)}}{\omega_0^2}\left\{e^{\beta\tau}[\cos\omega_1(t-t_0)\cos\omega_1\tau + \sin\omega_1(t-t_0)\sin\omega_1\tau]\right.$$

$$-\cos\omega_1(t-t_0) + \frac{\beta e^{\beta\tau}}{\omega_1}[\sin\omega_1(t-t_0)\cos\omega_1\tau - \cos\omega_1(t-t_0)\sin\omega_1\tau] \qquad (3.109)$$

$$\left. - \frac{\beta}{\omega_1}\sin\omega_1(t-t_0)\right\}, \qquad t > t_0$$

Because τ is small, we can expand $e^{\beta\tau}$, $\cos\omega_1\tau$, and $\sin\omega_1\tau$ using Equations D.34, D.29, and D.28, keeping only the first two terms in each. After multiplying out all the terms containing τ , we keep only the lowest-order term of τ .

$$x(t) = \frac{ae^{-\beta(t-t_0)}}{\omega_0^2}\sin\omega_1(t-t_0)\left[\omega_1\tau + \frac{\beta^2\tau}{\omega_1}\right], \qquad t > t_0$$

Using Equation 3.99 for ω_0^2 and $\tau = b/a$ gives us, finally,

$$x(t) = \frac{b}{\omega_1}e^{-\beta(t-t_0)}\sin\omega_1(t-t_0), \qquad t > t_0 \qquad (3.110)$$

This response function is shown in Figure 3-24 for the case $\beta = 0.2\omega_0$. Notice that, as t becomes large, the oscillator returns to its original position of equilibrium.

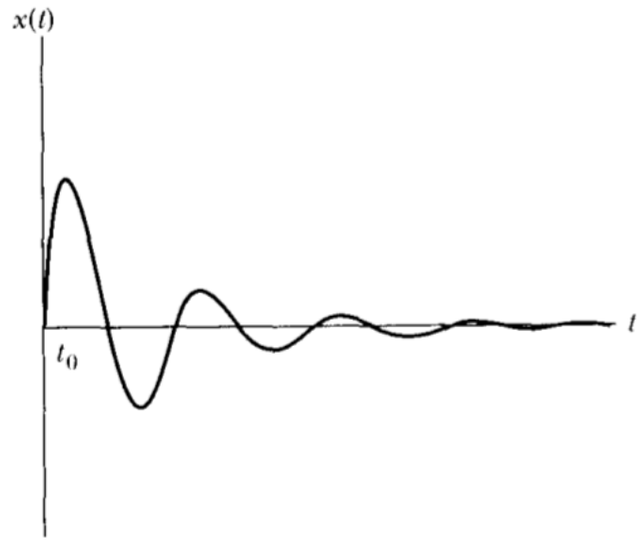


FIGURE 3-24 Response function solution to a spike (or delta function) force function.

A "spike" of this type is usually termed a **delta function** and is written $\delta(t - t_0)$. The delta function has the property that $\delta(t) = 0$ for t t ≠ 0 and $\delta(0) = \infty$, but

$$\int_{-\infty}^{+\infty} \delta(t - t_0)\, dt = 1$$

This is therefore not a proper function in the mathematical sense, but it can be defined as the limit of a well-behaved and highly local function (such as a Gaussian function) as the width parameter approaches zero.

The fact that the response of a linear oscillator to an impulsive driving force can be represented in the simple manner of Equation 3.110 leads to a powerful technique for dealing with general forcing functions, which was developed by Green. Green's method is based on representing an arbitrary forcing function as a series of impulses, shown schematically in Figure 3-25.

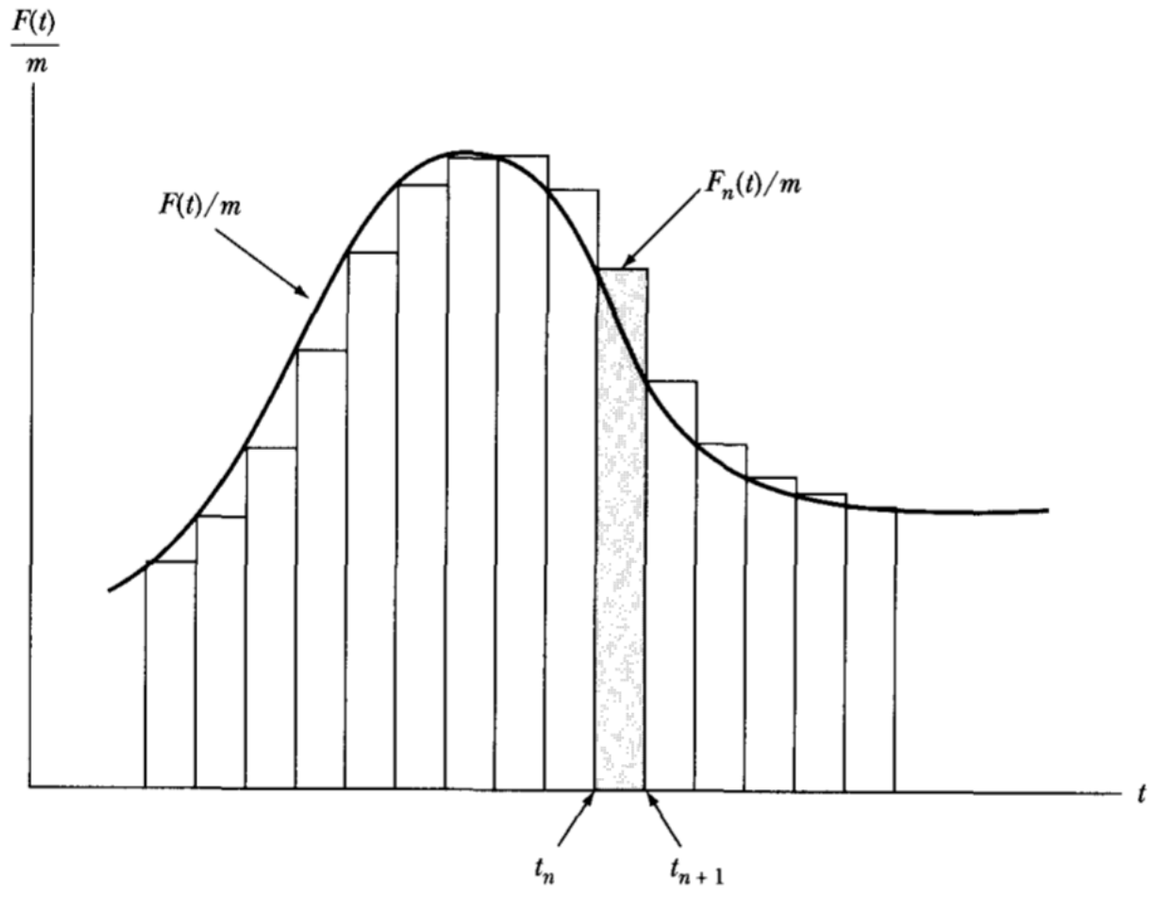


FIGURE 3-25 An arbitrary force function can be represented as a series of impulses, a method known as Green's methods.

If the driven system is linear, the principle of superposition is valid, and we can express the inhomogeneous part of the differential equation as the sum of individual forcing functions $F_n(t)/m$, which in Green's method are impulse functions:

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = \sum_{n=-\infty}^{\infty} \frac{F_n(t)}{m} = \sum_{n=-\infty}^{\infty} I_n(t) \tag{3.111}$$

where

$$\begin{aligned} I_n(t) &= I(t_n, t_{n+1}) \\ &= \begin{cases} a_n(t_n), & t_n < t < t_{n+1} \\ 0, & \text{Otherwise} \end{cases} \end{aligned} \tag{3.112}$$

The interval of time over which I_n acts is $t_{n+1} - t_n = \tau$, and $\tau \ll 2\pi/\omega_1$. The solution for the *n*th impulse is, according to Equation 3.110,

$$x_n(t) = \frac{a_n(t_n)\tau}{\omega_1} e^{-\beta(t-t_n)} \sin \omega_1(t - t_n), \qquad t > t_n + \tau \tag{3.113}$$

and the solution for all the impulses up to and including the *N*th impulse is

$$x(t) = \sum_{n=-\infty}^{N} \frac{a_n(t_n)\tau}{\omega_1} e^{-\beta(t-t_n)} \sin \omega_1(t - t_n), \qquad t_N < t < t_{N+1} \tag{3.114}$$

If we allow the interval t to approach zero and write t_n as t' if, then the sum becomes an integral:

$$x(t) = \int_{-\infty}^{t} \frac{a(t')}{\omega_1} e^{-\beta(t-t')} \sin \omega_1(t - t')\, dt' \tag{3.115}$$

We define

$$G(t, t') \equiv \begin{cases} \dfrac{1}{m\omega_1} e^{-\beta(t-t')} \sin \omega_1(t - t'), & t \geq t' \\ 0, & t < t' \end{cases} \tag{3.116}$$

Then, because

$$ma(t') = F(t') \tag{3.117}$$

we have

$$\boxed{x(t) = \int_{-\infty}^{t} F(t')\, G(t, t')\, dt'} \tag{3.118}$$

The function G(t, t') is known as the **Green's function** for the linear oscillator equation (Equation 3.96).

The solution expressed by Equation 3.118 is valid only for an oscillator initially at rest in its equilibrium position, because the solution we used for a single impulse (Equation 3.110) was obtained for just such an initial condition.

For other initial conditions, the general solution may be obtained in an analogous manner.

Green's method is generally useful for solving linear, inhomogeneous differential equations. The main advantage of the method lies in the fact that the Green's function G(t, t'), which is the solution of the equation for an infinitesimal element of the inhomogeneous part, *already contains the initial conditions* - so the general solution, expressed by the integral of F(t') G(t, t'), automatically also contains the initial conditions.

EXAMPLE 3.7

Find x(t) for an exponentially decaying forcing function beginning at t = 0 and having the following form for t > 0:

$$F(t) = F_0 e^{-\gamma t}, \qquad t > 0 \tag{3.119}$$

Solution. The solution for x(t) according to Green's method is

$$x(t) = \frac{F_0}{m\omega_1}\int_0^t e^{-\gamma t'} e^{-\beta(t-t')} \sin \omega_1 (t - t')\, dt' \tag{3.120}$$

Making a change of variable to z = ω_1(t - t'), we find

$$\begin{aligned} x(t) &= -\frac{F_0}{m\omega_1^2}\int_{\omega_1 t}^0 e^{-\gamma t} e^{[(\gamma-\beta)/\omega_1]z} \sin z\, dz \\ &= \frac{F_0/m}{(\gamma-\beta)^2 + \omega_1^2}\left[e^{-\gamma t} - e^{-\beta t}\left(\cos \omega_1 t - \frac{\gamma-\beta}{\omega_1}\sin \omega_1 t\right)\right] \end{aligned} \tag{3.121}$$

This response function is illustrated in Figure 3-26 for three different combinations of the damping parameters β and γ .

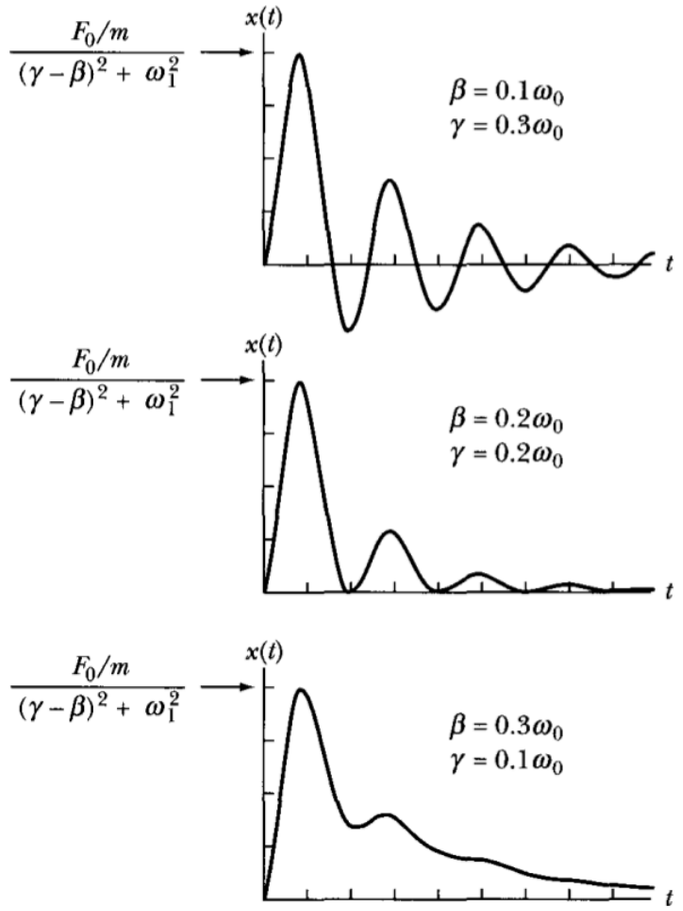


FIGURE 3-26 Response function for Example 3.7.

When β is large compared with γ, and if both are small compared with ω_0, then the response approaches that for a "spike"; compare Figure 3-24 with the upper curve in Figure 3-26. When β is small compared with γ, the response approaches the shape of the forcing function itself - that is, an initial increase followed by an exponential decay. The lower curve in Figure 3-26 shows a decaying amplitude on which is superimposed a residual oscillation. When β and γ are equal, Equation 3.121 becomes

$$x(t) = \frac{F_0}{m\omega_1^2} e^{-\beta t}(1 - \cos \omega_1 t), \qquad \beta = \gamma \tag{3.122}$$

Thus, the response is oscillatory with a "period" equal to $2\pi/\omega_1$ but with an exponentially decaying amplitude, as shown in the middle curve of Figure 3-26.

A response of the type given by Equation 3.121 could result, for example, if a quiescent but intrinsically oscillatory electronic circuit were suddenly driven by the decaying voltage on a capacitor.

CHAPTER 4 - Nonlinear Oscillations and Chaos

4.1 Introduction

The discussion of oscillators in Chapter 3 was limited to linear systems. When pressed to divulge greater detail, however, nature insists of being nonlinear; examples are the flapping of a flag in the wind, the dripping of a leaky water faucet, and the oscillations of a double pendulum. The techniques learned thus far for linear systems may not be useful for nonlinear systems, but a large number of techniques have been developed for nonlinear systems, some of which we address in this chapter. We use numerical techniques to solve some of the nonlinear equations in this chapter.

The equation of motion for the damped and driven oscillator of Chapter 3 moving in only one dimension can be written as

$$m\ddot{x} + f(\dot{x}) + g(x) = h(t) \tag{4.1}$$

If f($\dot{x}$) or g(x) contains powers of $\dot{x}$ or x, respectively, higher than linear, then the physical system is nonlinear. Complete solutions are not always available for Equation 4.1, and sometimes special treatment is needed to solve such equations. For example, we can learn much about a physical system by considering the deviation of the forces from linearity and by examining phase diagrams. Such a system is the simple plane pendulum, a system that is linear only when small oscillations are assumed.

In the beginning of the nineteenth century, the famous French mathematician Pierre Simon de Laplace espoused the view that if we knew the position and velocities of all the particles in the universe, then we would know the future for all time. This is the *deterministic* view of nature. In recent years, researchers in many disciplines have come to realize that knowing the laws of nature is not enough. Much of nature seems to be chaotic. In this case, we refer to deterministic chaos, as opposed to randomness, to be the motion of a system whose time evolution has a sensitive dependence on initial conditions.

The deterministic development refers to the way a system develops from one moment to the next, where the present system depends on the one just past in a well-determined way through physical laws. We are not referring to a random process in which the present system has no causal connection to the previous one (e.g., the flipping of a coin).

Measurements made on the state of a system at a given time may not allow us to predict the future situation even moderately far ahead, despite the fact that the governing equations are known exactly. Deterministic chaos is always associated with a nonlinear system; nonlinearity is a necessary condition for chaos but not a sufficient one. Chaos occurs when a system depends in a sensitive way on its previous state. Even a tiny effect, such as a butterfly flying nearby, may be enough to vary the conditions such that the future is *entirely* different than what it might have been, *not* just a tiny bit different. The advent of computers has allowed chaos to be studied because we now have the capability of performing calculations of the time evolution of the properties of a system that includes these tiny variations in the initial conditions. Chaotic systems can only be solved numerically, and there are no simple, general ways to predict when a system will exhibit chaos.

Chaotic phenomena have been uncovered in practically all areas of science and engineering - in irregular heartbeats; the motion of planets in our solar system; water dripping from a tap; electrical circuits; weather patterns; epidemics; changing populations of insects, birds, and animals; and the motion of electrons in atoms. The list goes on and on. Henri Poincare is generally given credit for first recognizing the existence of chaos during his investigation of celestial mechanics at the end of the nineteenth century. He came to the realization that the motion of apparently simple systems, such as the planets in our solar system, can be extremely complicated. Although various investigators also eventually came to understand the existence of chaos, tremendous breakthroughs did not happen until the 1970s, when computers were readily available to calculate the longtime histories required to document the behavior.

The study of chaos has become widespread, and we will only be able to look at the rudimentary aspects of the phenomena. Specialized textbooks on the subject have become abundant for those desiring further study. For example, space does not permit us to discuss the fascinating area of fractals, the complicated patterns that arise from chaotic processes.

4.2 Nonlinear Oscillations

Consider a potential energy of the parabolic form

$$U(x) = \frac{1}{2} kx^2 \tag{4.2}$$

Then the corresponding force is

$$F(x) = -kx \tag{4.3}$$

This is just the case of simple harmonic motion discussed in Section 3.2. Now, suppose a particle moves in a potential well, which is some arbitrary function of distance (as in Figure 4-1).

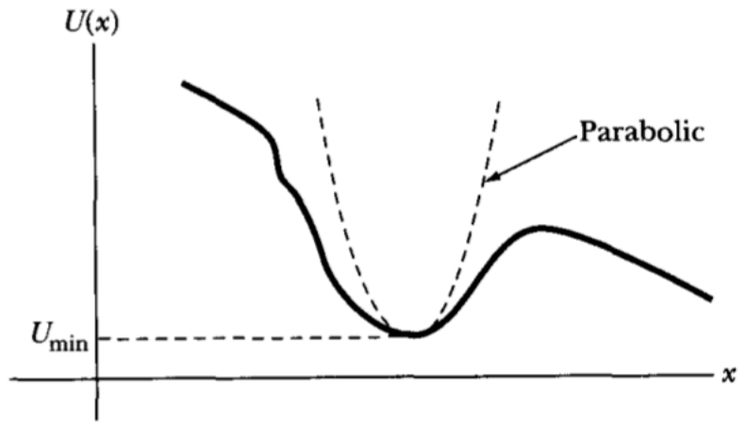


FIGURE 4-1 Arbitrary potential $U(x)$ indicating a parabolic region where simple harmonic motion is applicable.

Then, in the vicinity of the minimum of the well, we usually approximate the potential with a parabola. Therefore, if the energy of the particle is only slightly greater than U_{min} only small amplitudes are possible and the motion is approximately simple harmonic. If the energy is appreciably greater than U_{min} so that the amplitude of the motion cannot be considered small, then it may no longer be sufficiently accurate to make the approximation $U(x) \approx \frac{1}{2}kx^2$ and we must deal with a nonlinear force.

In many physical situations, the deviation of the force from linearity is *symmetric* about the equilibrium position (which we take to be at x = 0). In such cases, the magnitude of the force exerted on a particle is the same at - -x as at x; the direction of the force is opposite in the two cases. Therefore, in a symmetric situation, the first correction to a linear force must be a term proportional to x^;
hence.

$$F(x) \cong -kx + \varepsilon x^3 \tag{4.4}$$

where ε is usually a small quantity. The potential corresponding to such a force

$$U(x) = \frac{1}{2}kx^2 - \frac{1}{4}\varepsilon x^4 \tag{4.5}$$

Depending on the sign of the quantity ε, the force may either be greater or less than the linear approximation. If $\varepsilon > 0$, then the force is less than the linear term alone and the system is said to be soft; if $\varepsilon < 0$, then the force is greater and the system is hard. Figure 4-2 shows the form of the force and the potential for a soft and a hard system.

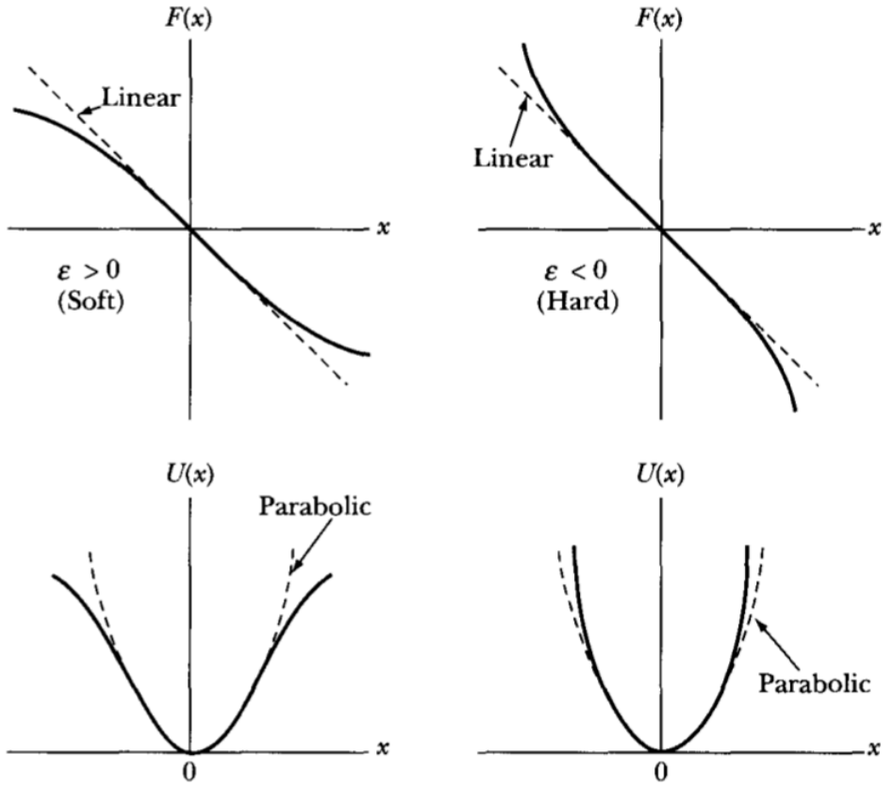
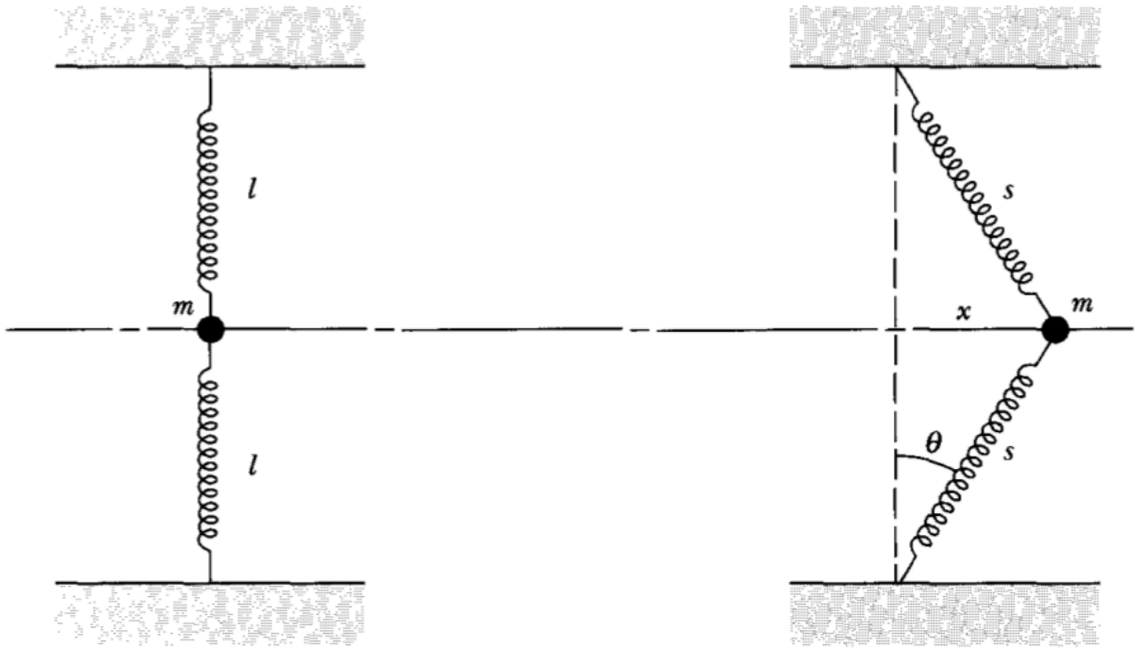


FIGURE 4-2 Force $F(x)$ and potential $U(x)$ for a soft and hard system when an x^3 term is added to the force.

EXAMPLE 4.1

Consider a particle of mass m suspended between two identical springs (Figure 4-3)



(a) Equilibrium position (b) Extended position

FIGURE 4-3 Example 4.1. A double spring system in (a) equilibrium and (b) extended positions.

Show that the system is nonlinear. Find the steady-state solution for a driving force F0 cos(ω t).

Solution. If both springs are in their unextended conditions (i.e., there is no tension, and therefore no potential energy, in either spring) when the particle is in its equilibrium position—and if we neglect gravitational forces—then when the particle is displaced from equilibrium (Figure 4-3b), each spring exerts a force -k(s - l) on the particle (k is the force constant of each spring). The net (horizontal) force on the particle is

$$F = -2k(s - l)\sin\theta \tag{4.6}$$

Now,

$$s = \sqrt{l^2 + x^2}$$

so

$$\sin\theta = \frac{x}{s} = \frac{x}{\sqrt{l^2 + x^2}}$$

Hence,

$$F = -\frac{2kx}{\sqrt{l^2 + x^2}} \cdot \left(\sqrt{l^2 + x^2} - l\right) = -2kx\left(1 - \frac{1}{\sqrt{1 + (x/l)^2}}\right) \tag{4.7}$$

If we consider x/l to be a small quantity and expand the radical, we find

$$F = -kl\left(\frac{x}{l}\right)^3\left[1 - \frac{3}{4}\left(\frac{x}{l}\right)^2 + \cdots\right]$$

If we neglect all terms except the leading term, we have, approximately,

$$F(x) \cong -(k/l^2)x^3 \tag{4.8}$$

Therefore, even if the amplitude of the motion is sufficiently restricted so that x/l is a small quantity, the force is still proportional to x^3. The system is therefore intrinsically nonlinear. However, if it had been necessary to stretch each spring a distance d to attach it to the mass when at the equilibrium position, then we would find for the force

$$F(x) \cong -2(kd/l)x - [k(l - d)/l^3]x^3 \tag{4.9}$$

and a linear term is introduced. For oscillations with small amplitude, the motion is approximately simple harmonic.

From Equation 4.9 we identify

$$\varepsilon' = -k(l - d)/l^3 < 0$$

Thus the system is hard.

If we have a driving force $F_0 \cos \omega t$, the equation of motion for the stretched spring (force of Equation 4.9) becomes

$$m\ddot{x} = -\frac{2kd}{l}x - \frac{k(l - d)}{l^3}x^3 + F_0 \cos \omega t \tag{4.10}$$

Let

$$\varepsilon = \frac{\varepsilon'}{m}, \quad a = \frac{2kd}{ml}, \quad \text{and} \quad G = \frac{F_0}{m} \tag{4.11}$$

then

$$\ddot{x} = -ax + \varepsilon x^3 + G \cos \omega t \tag{4.12}$$

Equation 4.12 is a difficult differential equation to solve. We can find the important characteristics of the solution by a method of successive approximations (perturbation technique). First, try a solution $x_1 = A \cos \omega t$, and insert x_1 into the right-hand side of Equation 4.12, which becomes

$$\ddot{x}_2 = -aA\cos\omega t + \varepsilon A^3\cos^3\omega t + G\cos\omega t \tag{4.13}$$

where the solution of Equation 4.13 is x= x_2. This equation can be solved for x_2 using the identity

$$\cos^3\omega t = \frac{3}{4}\cos\omega t + \frac{1}{4}\cos 3\omega t$$

Using this equation in Equation 4.13 gives

$$\ddot{x}_2 = -\left(aA - \frac{3}{4}\varepsilon A^3 - G\right)\cos\omega t + \frac{1}{4}\varepsilon A^3\cos 3\omega t \tag{4.14}$$

Integrating twice (with integration constants set equal to zero) gives

$$x_2 = \frac{1}{\omega^2}\left(aA - \frac{3}{4}\varepsilon A^3 - G\right)\cos\omega t - \frac{\varepsilon A^3}{36\omega^2}\cos 3\omega t \tag{4.15}$$

This is already a complicated solution. Under what conditions for ε, a, and x is x_2 a suitable solution? Numerical techniques with a computer can quickly yield a perturbative solution quite accurately. We have found that the amplitude depends on the driving frequency, but no resonance occurs at the natural frequency of the system.

Further discussion of solution methods for Equation 4.12 would take us too far afield of our present discussion. The result is that for some values of the driving frequency to, three different amplitudes may occur with "jumps" between the amplitudes. The amplitude may have a different value for a given to depending on whether to is increasing or decreasing (hysteresis). We present a simple case of this effect in Section 4.5.

In real physical situations, we are often concerned with symmetric forces and potentials. But some cases have asymmetric forms. For example,

$$F(x) = -kx + \lambda x^2 \tag{4.16}$$

The potential for which is

$$U(x) = \frac{1}{2}kx^2 - \frac{1}{3}\lambda x^3 \tag{4.17}$$

This case is illustrated in Figure 4-4 for A < 0; the system is hard for x > 0 and soft for x< 0.

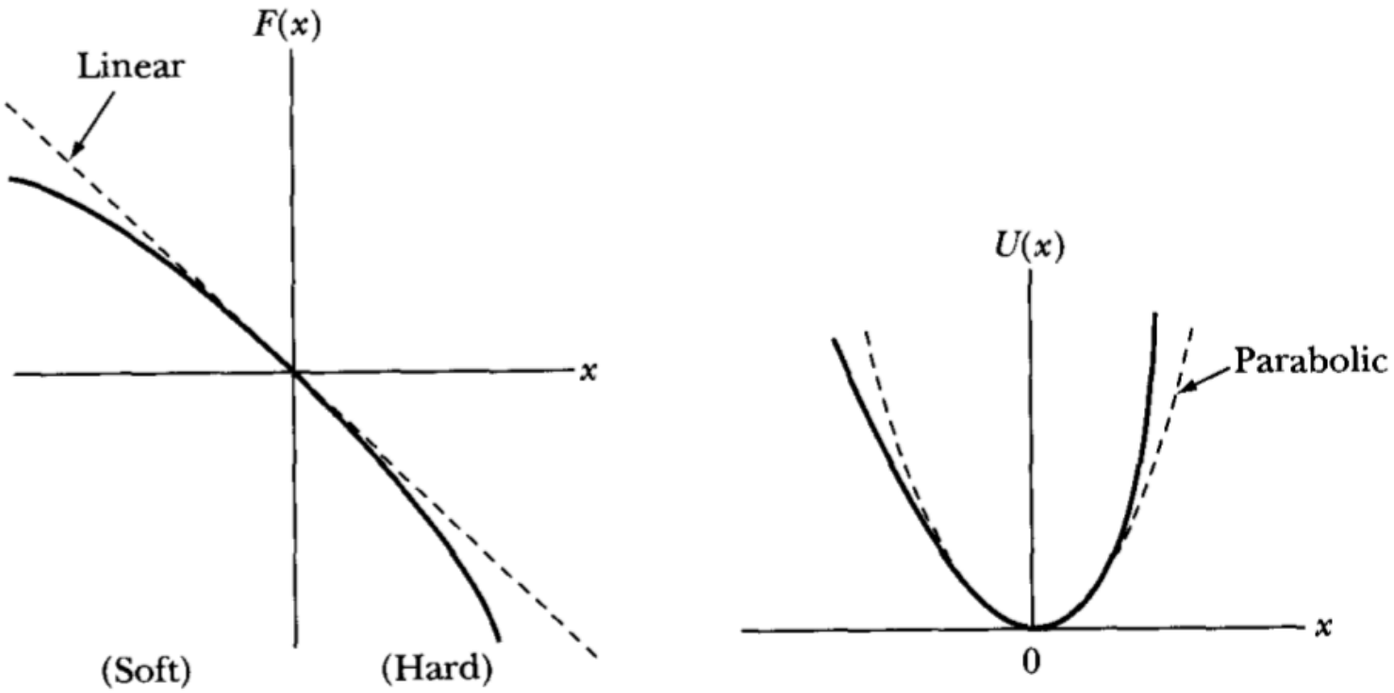


FIGURE 4-4 Example of asymmetric forces and potentials.

4.3 Phase Diagrams for Nonlinear Systems

The construction of a phase diagram for a nonlinear system may be accomplished by using Equation 2.97:

$$\dot{x}(x) \propto \sqrt{E - U(x)} \tag{4.18}$$

When U(x) is known, it is relatively easy to make a phase diagram for $\dot{x}$ (x). Computers, with their ever-improving graphics capability, make this a particularly easy task. However, in many cases it is difficult to obtain U(x), and we must resort to approximation procedures to eventually produce the phase diagram. On the other hand, it is relatively easy to obtain a qualitative picture of the phase diagram for the motion of a particle in an arbitrary potential. For example, consider the asymmetric potential shown in Figure 4-5a, which represents a system that is soft for x < 0 and hard for x > 0.

If no damping occurs, then because x is proportional to $\sqrt{E - U(x)}$, the phase diagram must be of the form shown in Figure 4-5b. Three of the oval phase paths are drawn, corresponding to the three values of the total energy indicated by the dotted lines in the potential diagram. For a total energy only slightly greater than that of the minimum of the potential, the oval phase paths approach ellipses. If the system is damped, then the oscillating particle will "spiral down the potential well" and eventually come to rest at the equilibrium position, x = 0. The equilibrium point at x = 0 in this case is called an **attractor**. An attractor is a set of point (or one point) in phase space toward which a system is "attracted" when damping is present.

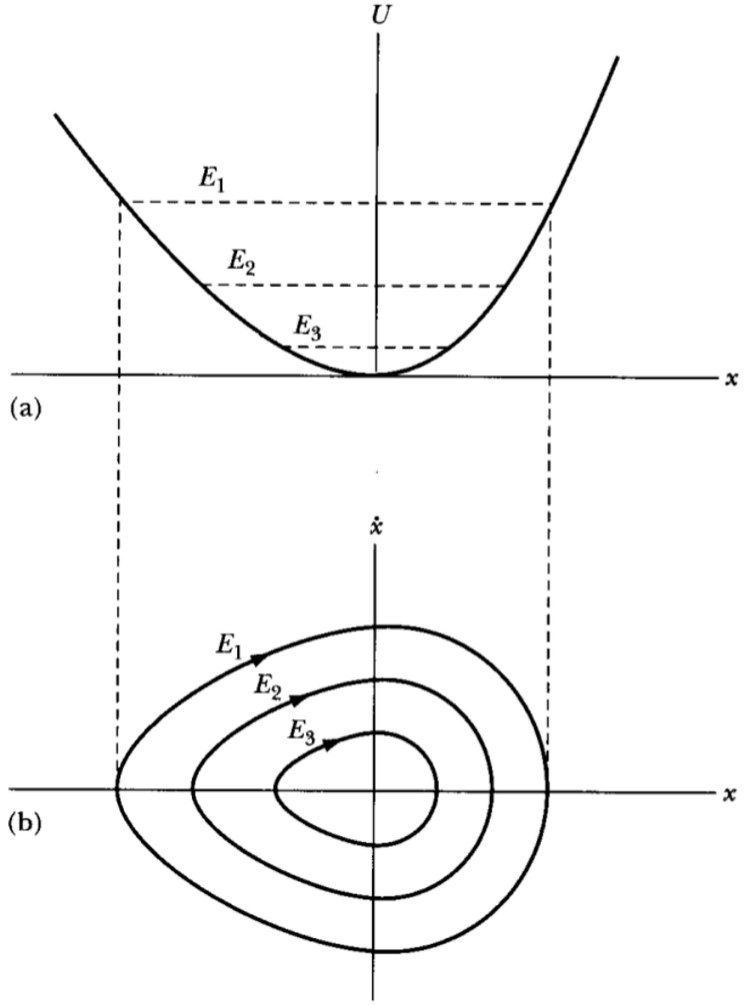


FIGURE 4-5 (a) Asymmetric potential and (b) phase diagram for bounded motion

For the case shown in Figure 4-5, if the total energy E of the particle is less than the height to which the potential rises on either side of x= 0, then the particle is "trapped" in the potential well (cf., the region $x_a < x < x_b$ in Figure 2-14). The point x = 0 is a position of stable equilibrium, because $(d^2U(x)/dx^2)_0 > 0$ (see Equation 2.103), and a small disturbance results in locally bounded motion.

In the vicinity of the maximum of a potential, a qualitatively different type of motion occurs (Figure 4-6).

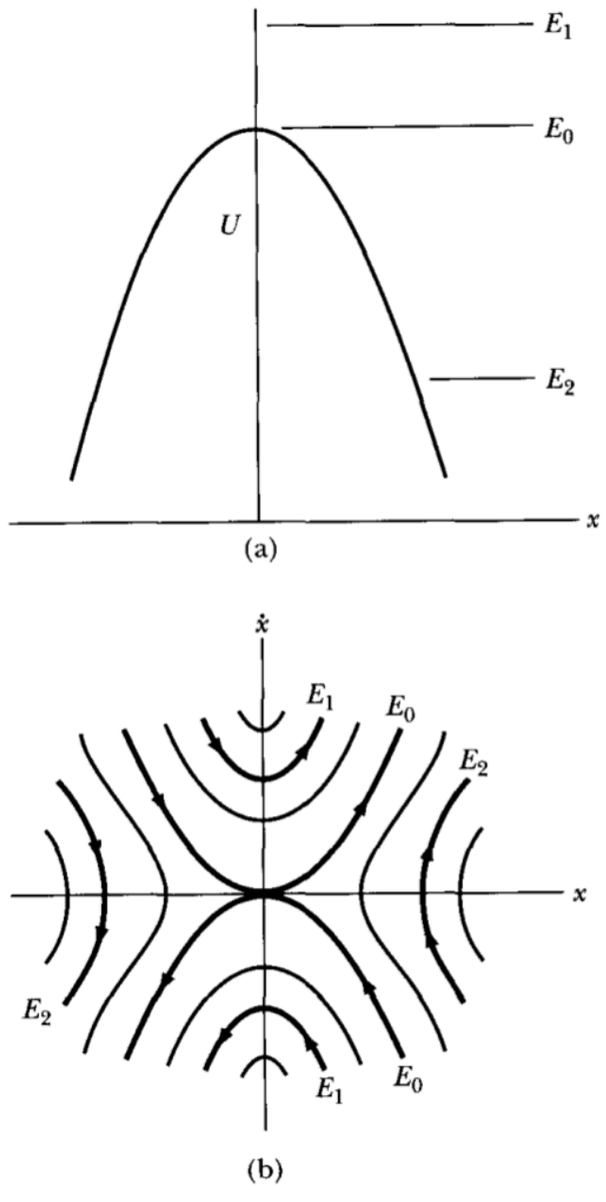


FIGURE 4-6 (a) Inverted asymmetric potential and (b) phase diagram for unbounded motion.

Here the point x = 0 is one of *unstable* equilibrium, because if a particle is at rest at this point, then a slight disturbance will result in locally unbounded motion* Similarly, $(d^2U(x)/dx^2)_0 < 0$ gives unstable equilibrium.

If the potential in Figure 4-6a were parabolic - if $U(x) = \frac{1}{2}kx^2$ - then the phase paths corresponding to the energy Eq would be straight lines and those corresponding to the energies E_1 and E_2 would be hyperbolas. This is therefore the limit to which the phase paths of Figure 4-6 would approach if the nonlinear term in the expression for the force were made to decrease in magnitude.

By referring to the phase paths for the potentials shown in Figures 4-5 and 4-6, we can rapidly construct a phase diagram for any arbitrary potential (such as that in Figure 2-14).

An important type of nonlinear equation was extensively studied by van der Pol in his investigation of nonlinear oscillations in vacuum tube circuits of early radios. This equation has the form

$$\ddot{x} + \mu(x^2 - a^2)\dot{x} + \omega_0^2 x = 0 \tag{4.19}$$

where μ is a small, positive parameter. A system described by van der Pol's equation has the following interesting property. If the amplitude |x| exceeds the critical value |a|, then the coefficient of $\dot{x}$ is positive and the system is damped. But if |x| < |a|, then negative damping occurs; that is, the amplitude of the motion increases. It follows that there must be some amplitude for which the motion neither increases nor decreases with time. Such a curve in the phase plane is called the **limit cycle** (Figure 4-7) and is the attractor for this system.

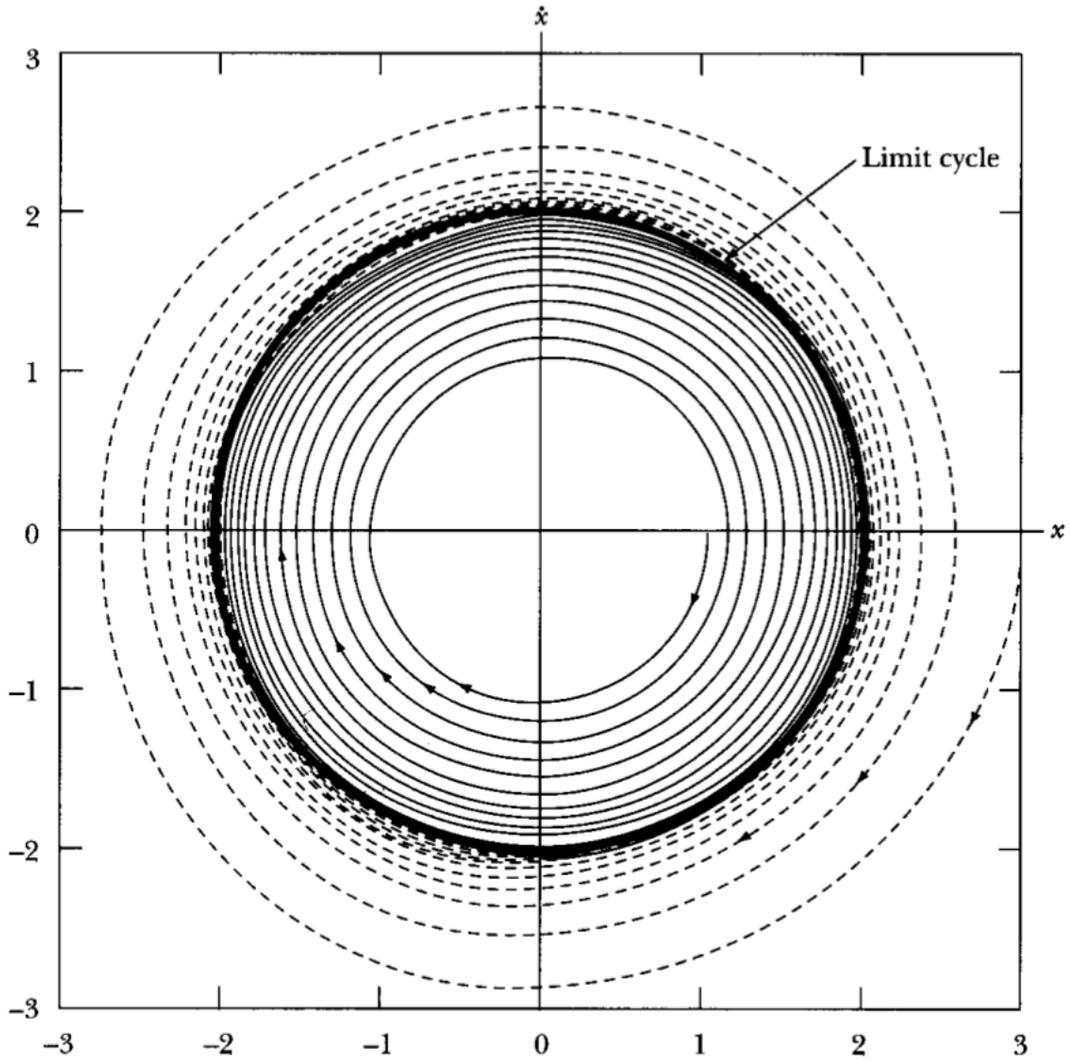


FIGURE 4-7 Phase diagram for the solution of the van der Pol Equation 4.20. The damping term is $\mu = 0.05$, and the solution very slowly approaches the limit cycle at 2. Positive and negative damping occur, respectively, for $|x|$ values outside and inside the limit cycle at 2. The solid and dashed lines have initial $(x, \dot{x})$ values of (1.0, 0) and (3.0, 0), respectively.

Phase paths outside the limit cycle spiral *inward*, and those inside the limit cycle spiral outward. In as much as the limit cycle defines locally *bounded* motion, we may refer to the situation it represents as stable.

A system described by van der Pol's equation is *self-limiting*; that is, once set into motion under conditions that lead to an increasing amplitude, the amplitude is automatically prevented from growing without bound. The system has this property whether the initial amplitude is greater or smaller than the critical (limiting) amplitude x_0.

Now let us turn to the numerical calculation of van der Pol's Equation 4.19. In order to make the calculation simpler and to be able to examine the system's motion, we let a = 1 and ω_0 = 1 with appropriate units. Equation 4.19 becomes

$$\ddot{x} + \mu(x^2 - 1)\dot{x} + x = 0 \tag{4.20}$$

In our case, we used a computer to solve this differential equation. We use a value of μ= 0.05, which will give a small damping term. It will take some time for the solution to reach the limit cycle. We show the calculation for two initial values of x(x_0 =1.0 and 3.0) in Figure 4-7; in both cases, we let the initial value of $\dot{x}$ = 0. Note that in this case the limit cycle is a circle of radius 2.

In both cases, when the initial values are both inside and outside the limit cycle, the solution spirals toward the limit cycle. If we set x_0 = 2 (with $\dot{x}_0 = 0$), the motion remains at the limit cycle. The solution of the circle in this case is a result of our special values for a and ω_0 above. If we use a large damping term, μ= 0.5, the solution reaches the limit cycle much more quickly, and the limit cycle is distorted as shown in Figure 4-8.

For a small value of μ(0.05) the x and $\dot{x}$ terms are sinusoidal with time, but for higher values of μ(0.05) the sinusoidal shapes become skewed The van der Pol oscillator is a nice system for studying nonlinear behavior.

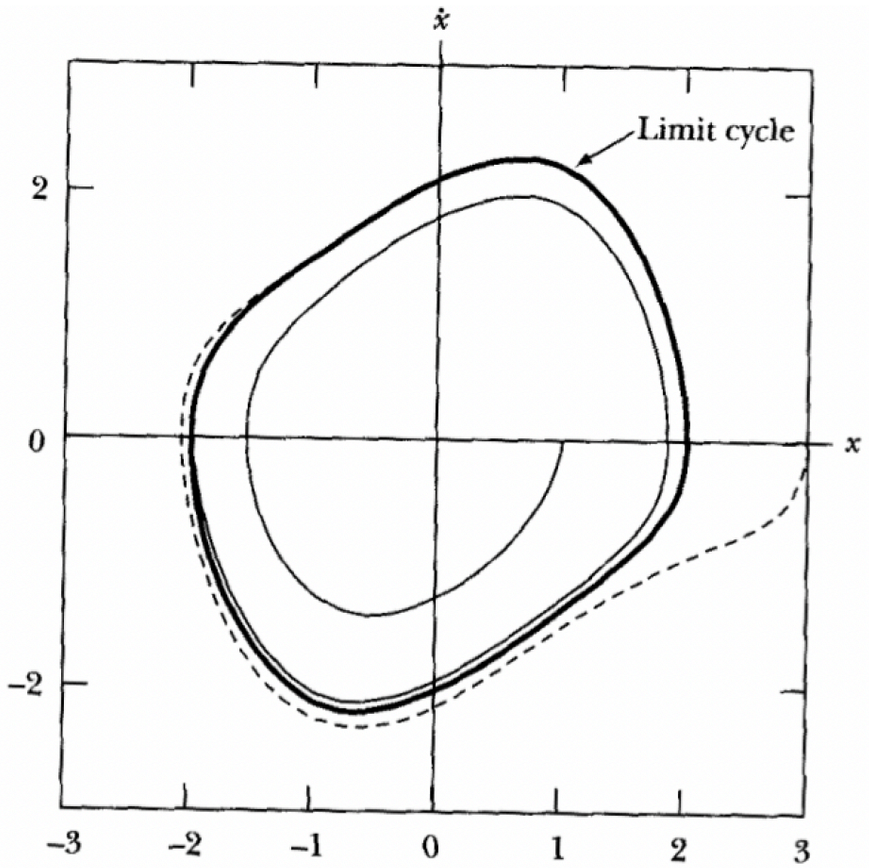


FIGURE 4-8 Similar calculation to Figure 4-7 for the solution of the van der Pol Equation 4.20. In this case the damping parameter μ = 0.5. Note that the solution reaches the limit cycle (now skewed) much more quickly.

4.4 Plane Pendulum

The solutions of certain types of nonlinear oscillation problems can be expressed in closed form by elliptic integrals(Appendix B). An example of this type is the plane pendulum. Consider a particle of mass m constrained by a weightless, extensionless rod to move in a vertical circle of radius l (Figure 4-9).

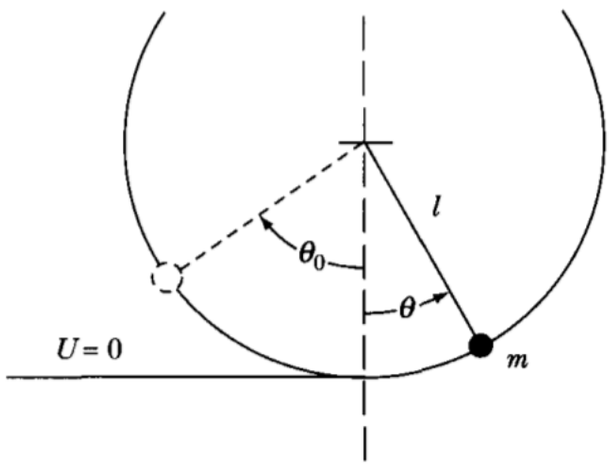


FIGURE 4-9 The plane pendulum where the mass m is not required to oscillate in small angles. The angle $\theta > 0$ is in the counterclockwise direction so that $\theta_0 < 0$.

The gravitational force acts downward, but the component of this force influencing the motion is *perpendicular* to the support rod. This force component, shown in Figure 4-10, is simply $F(\theta) = -mg \sin \theta$. The plane pendulum is a nonlinear system with a symmetric restoring force. It is only for small angular deviations that a linear approximation may be used.

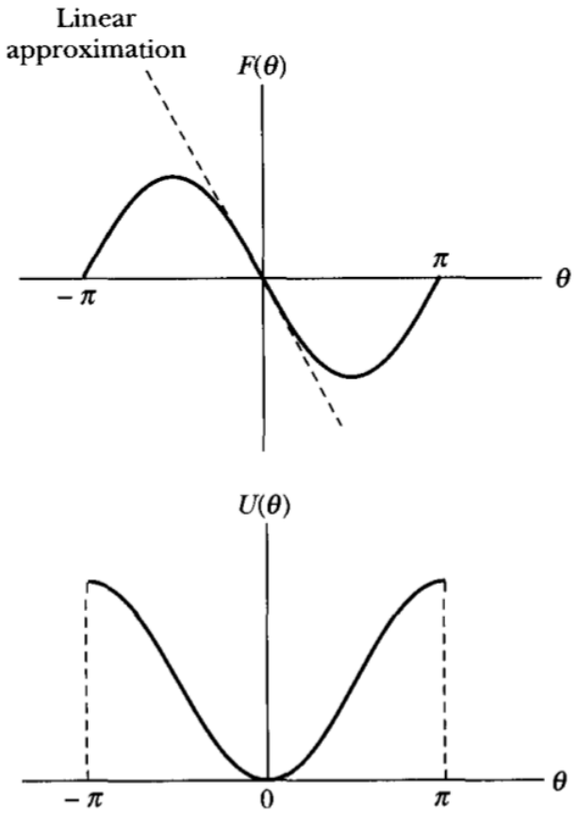


FIGURE 4-10 The component of the force, $F(\theta)$, and its associated potential that acts on the plane pendulum. Notice that the force is nonlinear.

We obtain the equation of motion for the plane pendulum by equating the torque about the support axis to the product of the angular acceleration and the rotational inertia about the same axis:

$$I\ddot{\theta} = lF$$

or, because because I = m l^2 and F= - mg sin θ ,

$$\boxed{\ddot{\theta} + \omega_0^2 \sin\theta = 0} \tag{4.21}$$

where

$$\omega_0^2 \equiv \frac{g}{l} \tag{4.22}$$

If the amplitude of the motion is small, we may approximate $\sin\theta \simeq \theta$, and the equation of motion becomes identical with that for the simple harmonic oscillator:

$$\ddot{\theta} + \omega_0^2 \theta = 0$$

In this approximation, the period is given by the familiar expression

$$\tau \cong 2\pi\sqrt{\frac{l}{g}}$$

If we wish to obtain the general result for the period in the event that the amplitude is finite, we may begin with Equation 4.21. But because the system is conservative, we can use the fact that

$$T + U = E = \text{constant}$$

to obtain a solution by considering the energy of the system rather than by solving the equation of motion.

If we take the zero of potential energy to be the lowest point on the circular path described by the pendulum bob (i.e., $\theta = 0$; see Figure 4-10), the kinetic and potential energies can be expressed as

$$\left.\begin{aligned} T &= \frac{1}{2} I\omega^2 = \frac{1}{2} m l^2 \dot{\theta}^2 \\ U &= mgl(1 - \cos\theta) \end{aligned}\right\} \tag{4.23}$$

If we let $\theta = \theta_0$ the highest point of the motion, then

$$T(\theta = \theta_0) = 0$$
$$U(\theta = \theta_0) = E = mgl(1 - \cos\theta_0)$$

Using the trigonometric identity

$$\cos\theta = 1 - 2\sin^2(\theta/2)$$

we have

$$E = 2mgl \sin^2(\theta_0/2) \tag{4.24}$$

and

$$U = 2mgl \sin^2(\theta/2) \tag{4.25}$$

Expressing the kinetic energy as the difference between the total energy and the potential energy, we have T= E — U,

$$\frac{1}{2} m l^2 \dot{\theta}^2 = 2mgl\,[\sin^2(\theta_0/2) - \sin^2(\theta/2)]$$

or

$$\dot{\theta} = 2\sqrt{\frac{g}{l}}\,[\sin^2(\theta_0/2) - \sin^2(\theta/2)]^{1/2} \tag{4.26}$$

from which

$$dt = \frac{1}{2}\sqrt{\frac{l}{g}}\,[\sin^2(\theta_0/2) - \sin^2(\theta/2)]^{-1/2}\,d\theta$$

This equation may be integrated to obtain the period τ . Because the motion is symmetric, the integral over θ from $\theta = 0$ to $\theta = \theta_0$ yields $\tau/4$; hence

$$\tau = 2\sqrt{\frac{l}{g}}\int_0^{\theta_0}[\sin^2(\theta_0/2) - \sin^2(\theta/2)]^{-1/2}\,d\theta \tag{4.27}$$

That this is actually an elliptic integral of the first kind may be seen more clearly by making the substitutions

$$z = \frac{\sin(\theta/2)}{\sin(\theta_0/2)}, \qquad k = \sin(\theta_0/2)$$

Then

$$dz = \frac{\cos(\theta/2)}{2\,\sin(\theta_0/2)}\,d\theta = \frac{\sqrt{1 - k^2z^2}}{2k}\,d\theta$$

from which

$$\tau = 4\sqrt{\frac{l}{g}}\int_0^1 [(1 - z^2)(1 - k^2z^2)]^{-1/2}dz \tag{4.28}$$

Numerical values for integrals of this type can be found in various tables.

For oscillatory motion to result, $|\theta_0| < \pi$. or, equivalently, $\sin\theta_0/2 = k$, where -1 < k < +1. For this case, we can evaluate the integral in Equation 4.28 by expanding $(1 - k^2z^2)^{-1/2}$ in a power series:

$$(1-k^2z^2)^{-1/2} = 1 + \frac{k^2z^2}{2} + \frac{3k^4z^4}{8} + \cdots$$

Then, the expression for the period becomes

$$\tau = 4\sqrt{\frac{l}{g}}\int_0^1 \frac{dz}{(1-z^2)^{1/2}}\left[1+\frac{k^2z^2}{2}+\frac{3k^4z^4}{8}+\cdots\right]$$

$$= 4\sqrt{\frac{l}{g}}\left[\frac{\pi}{2}+\frac{k^2}{2}\cdot\frac{1}{2}\cdot\frac{\pi}{2}+\frac{3k^4}{8}\cdot\frac{3}{8}\cdot\frac{\pi}{2}+\cdots\right]$$

$$= 2\pi\sqrt{\frac{l}{g}}\left[1+\frac{k^2}{4}+\frac{9k^4}{64}+\cdots\right]$$

If $|k|$ is large (i.e., near 1), then we need many terms to produce a reasonably accurate result. But for small k, the expansion converges rapidly. And because $k = \sin(\theta_0/2)$, then $k \simeq (\theta_0/2) - (\theta_0^3/48)$; the result, correct to the fourth order, is

$$\tau \cong 2\pi\sqrt{\frac{l}{g}}\left[1+\frac{1}{16}\theta_0^2+\frac{11}{3072}\theta_0^4\right] \tag{4.29}$$

Therefore, although the plane pendulum is not isochronous, it is very nearly so for small amplitudes of oscillation.

We may construct the phase diagram for the plane pendulum in Figure 4-11 because Equation 4.26 provides the necessary relationship $\dot{\theta} = \dot{\theta}(\theta)$. The parameter θ_0 specifies the total energy through Equation 4.24. If θ and θ_0 are small angles, then Equation 4.26 can be written as

$$\left(\sqrt{\frac{l}{g}}\dot{\theta}\right)^2 + \theta^2 \cong \theta_0^2 \tag{4.30}$$

If the coordinates of the phase plane are θ and $\dot{\theta}/\sqrt{g/l}$, then the phase paths near $\theta = 0$ are approximately circles. This result is expected, because for small θ_0 , the motion is approximately simple harmonic.

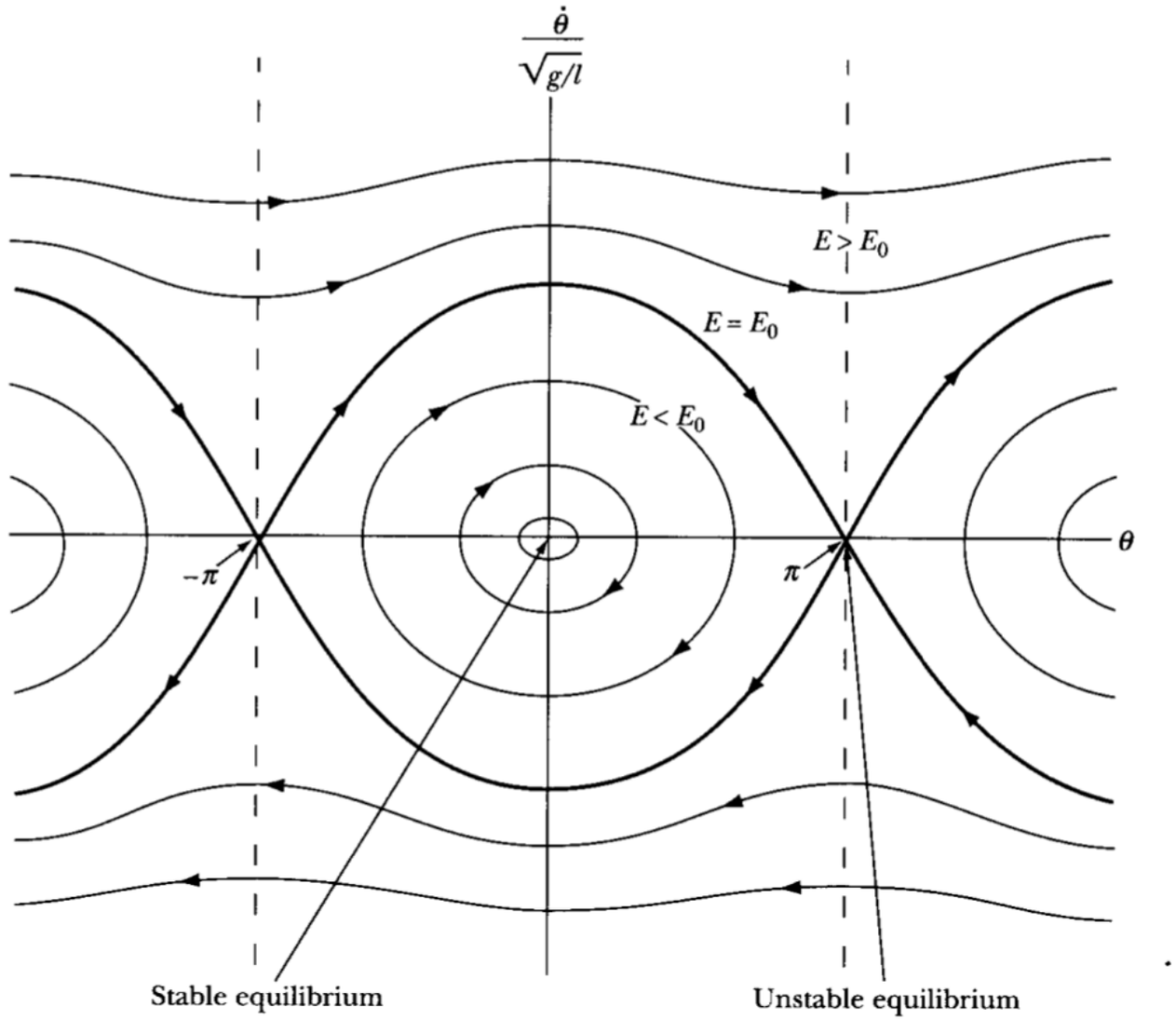


FIGURE 4-11 The phase diagram for the plane pendulum. Note the stable and unstable equilibrium points and the regions of bounded and unbounded motion.

For $-\pi < \theta < \pi$ and $E < 2mgl = E_0$, the situation is equivalent to a particle bound in the potential well $U(\theta) = mgl(1 - \cos\theta)$ (see Figure 4-10). The phase paths are therefore closed curves for this region and are given by Equation 4.26. Because the potential is periodic in θ, exactly the same phase paths exist for the regions $\pi < \theta < 3\pi$, $-3\pi < \theta - \pi$, and so forth. The points $\theta = \cdots, -2\pi, 0, 2\pi, \cdots$ along the θ-axis are positions of *stable* equilibrium and are the attractors when the undriven pendulum is damped.

For values of the total energy exceeding E_0, the motion is no longer oscillatory - although it is still periodic. This situation corresponds to the pendulum executing complete revolutions about its support axis. Normally the phase space diagram is plotted for only one complete cycle or a "unit cell," in this case over the interval $-\pi < \theta < \pi$. We denote this region in Figure 4-11 between the dashed lines at angles $-\pi$ and π. One can follow a phase path by noting that motion that exits on the left of the cell re-enters on the right and vice versa.

If the total energy equals E_0, then Equation 4.24 shows that $\theta_0 = \pm\pi$. In this case, Equation 4.26 reduces to

$$\dot{\theta} = \pm 2\sqrt{\frac{g}{l}}\cos(\theta/2) \tag{4.31}$$

so the phase paths for $E = E_0$ are just cosine functions (see the heavy curves in The phase paths for Figure 4-11). There are two branches, depending on the direction of motion.

The phase paths for $E = E_0$ do not actually represent possible continuous motions of the pendulum. If the pendulum were at rest at, say, $\theta = \pi$ (which is a point on the $E = E_0$ phase paths), then any small disturbance would cause the motion to follow *closely but not exactly on* one of the phase paths that diverges from $\theta = \pi$, because the total energy would he $E = E_0 + \delta$, where δ is a small but nonzero quantity. If the motion were along one of the $E = E_0$ phase paths, the pendulum would reach one of the points $\theta = n\pi$ with exactly zero velocity, but only after an infinite time! (This may be verified by evaluating Equation 4.27 for $\theta_0 = \pi$; the result is $\tau \to \infty$.)

A phase path separating locally bounded motion from locally unbounded motion (such as the path for $E = E_0$ in Figure 4-11) is called a **separatrix**. A separatrix always passes through a point of unstable equilibrium. The motion in the vicinity of such a separatrix is extremely sensitive to initial conditions because points on either side of the separatrix have very different trajectories.

4.5 Jumps, Hysteresis, and Phase Lags

In Example 4.1 we considered a particle of mass m suspended between two springs. We showed that the system was nonlinear and mentioned the phenomena of jumps in amplitude and hysteresis effects. Now, we want to examine such phenomena more carefully.

Consider a harmonic oscillator subjected to an external force $F(t) = F_g \cos \omega t$ and a resistive viscous force $-r\dot{x}$, where r is a constant. The equation of motion for a particle of mass m connected to a spring with force constant k is

$$m\ddot{x} = -r\dot{x} - kx + F_0 \cos \omega t \tag{4.32}$$

A solution to Equation 4.32 is

$$x(t) = A(\omega) \cos [\omega t - \phi(\omega)] \tag{4.33}$$

where

$$A(\omega) = \frac{F_0}{[(k - m\omega^2)^2 + (r\omega)^2]^{1/2}} \tag{4.34}$$

and

$$\tan[\phi(\omega)] = \frac{r\omega}{(k - m\omega^2)} \tag{4.35}$$

You can verify that Equation 4.33 is a particular solution by substitution into Equation 4.32.

If the spring constant k depends on x as k(x), then we have a nonlinear oscillator. An often used dependence is

$$k(x) = (1 + \beta x^2) k_0 \tag{4.36}$$

and the resulting equation of motion in Equation 4.32 is known as the Duffing equation. It has been widely studied through perturbation techniques with solutions similar to Equation 4.33 but with complicated results for $A(\omega)$ and $\phi(\omega)$ as shown in Figure 4-12.

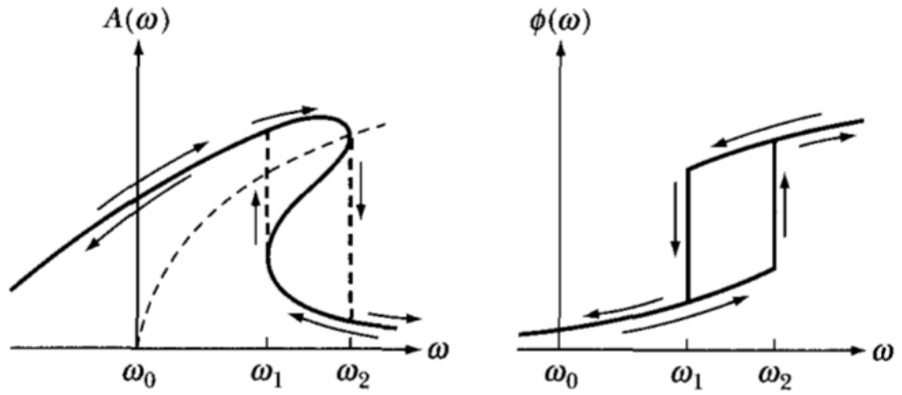


FIGURE 4-12 The amplitude $A(\omega)$ and phase angle $\phi(\omega)$ as a function of the angular frequency ω. Notice the "jumps" at ω_1 and ω_2 depending on the direction of change of ω.

As ω increases, A(ω) increases to its peak until it reaches $\omega = \omega_2$, where the amplitude suddenly decreases by a large factor. As ω decreases from large values, the amplitude slowly increases until $\omega = \omega_1$, where the amplitude suddenly approximately doubles. These are the "jumps" referred to earlier. The amplitude between ω_1 and ω_2 depends on whether ω is increasing or decreasing (hysteresis effect). Similarly strange phenomena occur for the phase $\phi(\omega)$ in Figure 4-12. The physical explanation of Figure 4-12 is not very transparent, so we consider a simpler dependence of k as shown in Figure 4-13.

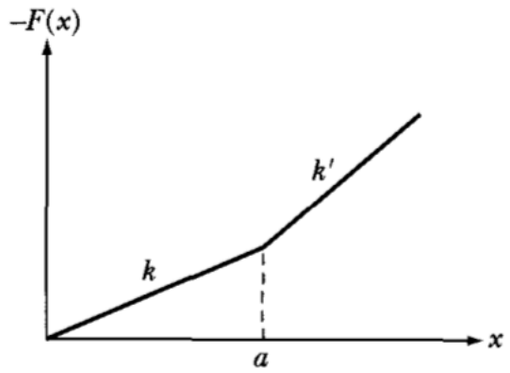


FIGURE 4-13 A simpler dependence of $F(x)$ on the spring constant k than that in Equation 4.36.

$$\begin{aligned} F(x) &= -kx && x \le a \\ &\simeq -k'x - c && x \ge a \end{aligned} \tag{4.37}$$

The Duffing equation represents a situation with many values of a, because k(x) continuously varies in Equation 4.36. Our example of an anharmonic oscillator allows simpler mathematics.

Figure 4-14 shows the harmonic response curves A(ω) for k and k' (with k < k').

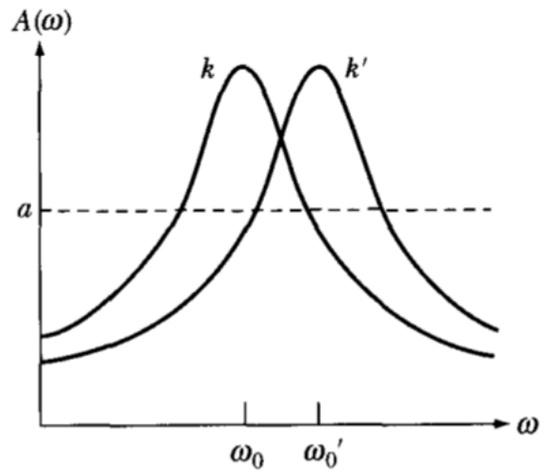


FIGURE 4-14 The values of $A(\omega)$ for the two values of k shown in Figure 4-12.

For very large values of a (a —> ∞), we have a linear oscillator with force constant k (because x < a, see Figure 4-13) and a resonance frequency $\omega_0 = (k/m)^{1/2}$. For very small values of a(a —> 0), the force constant is k' and $\omega_0' = (k'/m)^{1/2}$

We want to consider intermediate values of a, where both k and k' are effective. We consider the situation in which a is much smaller than the maximum amplitude of A(ω). If we start at small values of ω , our system has small vibrations that follow the amplitude curve for k. The amplitude moves up the tail of the A(ω) curve for k as shown in Figure 4-15.

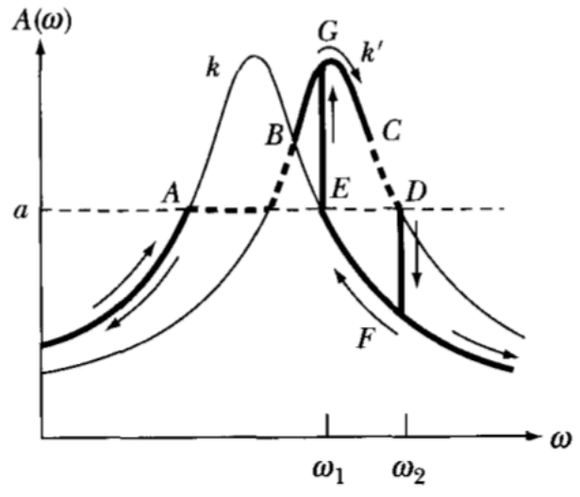


FIGURE 4-15 The bold lines and arrows help follow the path as ω increases and decreases.

However, when the vibration amplitude A(ω) is larger than the critical amplitude a, the force constant k' is effective. For these larger amplitudes, the system follows A'(ω) for force constant k'. This is represented by the solid bold line from B to C in Figure 4-15.

Between A and B, as the frequency increases, the system follows the simplified amplitude rise shown by the dashed line in Figure 4-15. Continuing to increase the driving frequency ω at C, we again reach the critical amplitude a at point D. If ω is only slightly increased, the system must follow A(ω) for k, and the amplitude suddenly jumps down from A'(ω) at point D to A(ω) at point F at $\omega = \omega_2$ As ω continues increasing above ω_2 , the system follows the A(co) curve.

Now let us see what happens if we decrease ω from large values. The system follows A(ω) until $\omega = \omega_1$, where A(ω) = a. If ω is barely decreased, the amplitude increases above a, and the system must follow A'(ω). Therefore the amplitude jumps from E to G. As to continues decreasing, it follows a similar path as before.

A hysteresis effect occurs because the system behaves differently depending on whether ω is increasing or decreasing. Two amplitude jumps occur, one for ω increasing and one for ω decreasing. The system's paths are ABGCDF (to increasing) and FEGBA {to decreasing).

Similar phenomena occur for the phase lag $\phi(\omega)$. Figure 4-16a shows the phase curves $\phi(\omega)$) and $\phi(\omega')$) for the linear harmonic oscillators. Using the same arguments as applied to A(ω), we depict the system's paths in Figure 4-16b by the bold lines and the arrows.

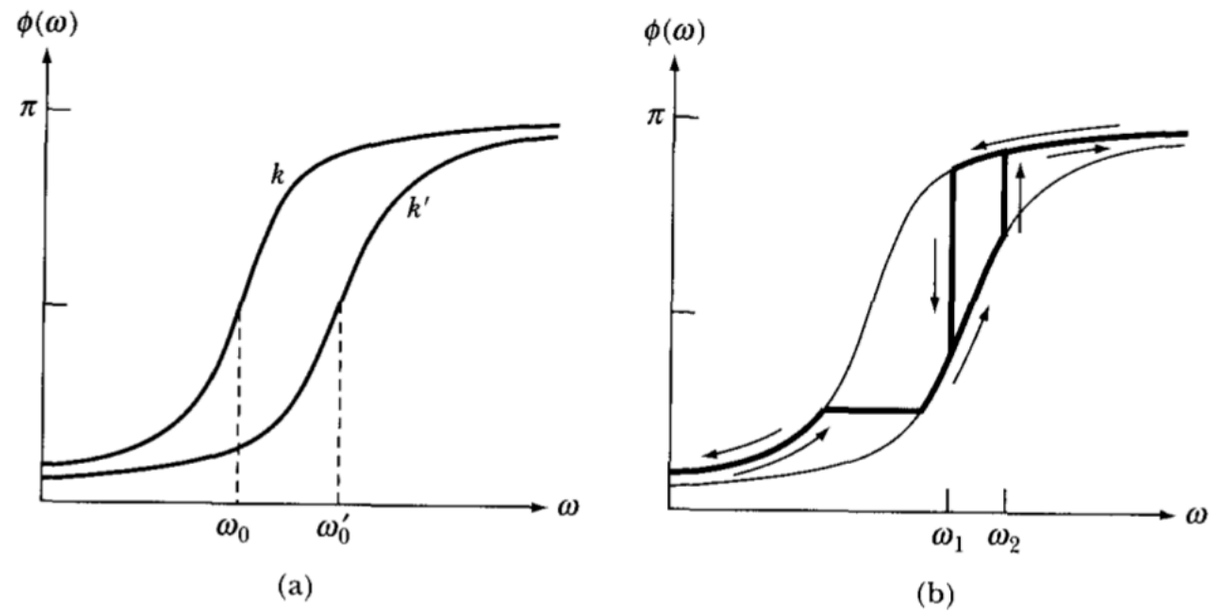
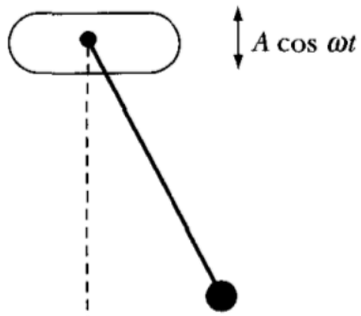


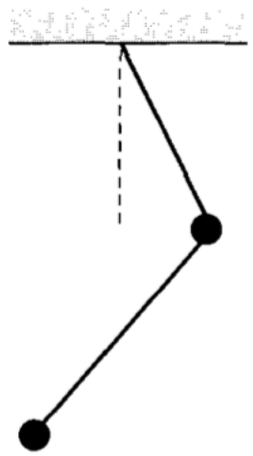
FIGURE 4-16 The phase angle $\phi(\omega)$ for k and k' is shown in (a), and the system's path is shown in (b).

4.6 Chaos in a Pendulum

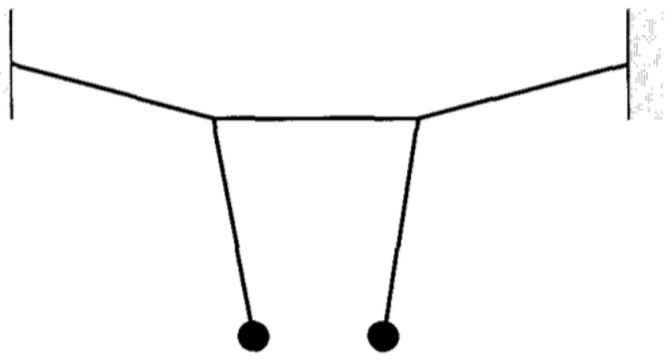
We will use the damped and driven pendulum to introduce several chaos concepts. The simple motion of a pendulum is well understood after hundreds of years of study, but its chaotic motion has been extensively studied only in the past few years.



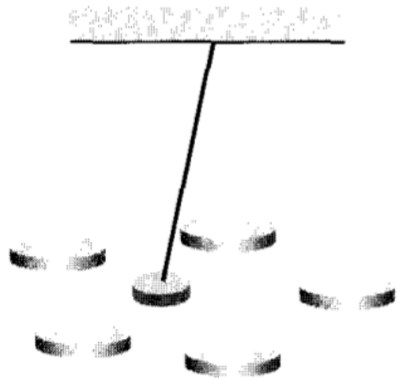
(a) Forced pivot



(b) Double pendulum



(c) Coupled pendulums



(d) Magnetic pendulum

FIGURE 4-17 Examples of pendulums that have chaotic motion.

4-17a, the double pendulum (Figure 4-17b), coupled pendulums (Figure 4-17c), and a pendulum oscillating between magnets (Figure 4-17d). The damped and driven pendulum that we will consider is driven around its pivot point, and the geometry is displayed in Figure 4-18.

Among the motions of pendula that have been found to be chaotic are a pendulum with a forced oscillating support as shown in Figure 4-17a, the double pendulum (Figure 4-17b), coupled pendulums (Figure 4-17c), and a pendulum oscillating between magnets (Figure 4-17d).

The damped and driven pendulum that we will consider is driven around its pivot point, and the geometry is displayed in Figure 4-18.

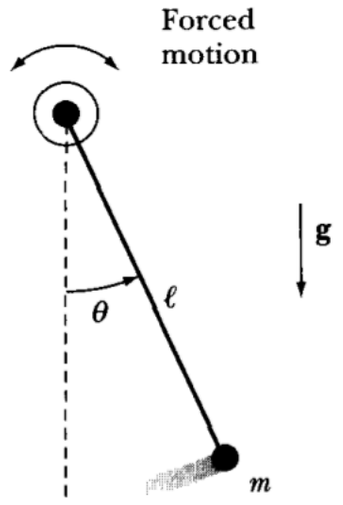


FIGURE 4-18 A damped pendulum is driven about its pivot point.

The torque around the pivot point can be written as

$$N = I\frac{d^2\theta}{dt^2} = I\ddot{\theta} = -b\dot{\theta} - mg\ell \sin\theta + N_d \cos\omega_d t \tag{4.38}$$

where I is the moment of inertia, b is the damping coefficient, and N_d is the driving torque of angular frequency ω_d . If we divide by $I = m\ \ell^2$, we have

$$\ddot{\theta} = -\frac{b}{m\ell^2}\dot{\theta} - \frac{g}{\ell}\sin\theta + \frac{N_d}{m\ell^2}\cos\omega_d t \tag{4.39}$$

We will eventually want to deal with this equation with a computer, and it will be much easier in that case to use dimensionless parameters. Let us divide Equation 4.39 by $\omega_0^2 = g/\ell$ and define the dimensionless time $t' = t/t_0$ with $t_0 = 1/\omega_0$ and the dimensionless driving frequency $\omega = \omega_d/\omega_0$. The new dimensionless variables and parameters are

$$x = \theta \qquad \text{oscillating variable} \tag{4.40a}$$

$$c = \frac{b}{m\ell^2\omega_0} \qquad \text{damping coefficient} \tag{4.40b}$$

$$F = \frac{N_d}{m\ell^2\omega_0^2} = \frac{N_d}{mg\ell} \qquad \text{driving force strength} \tag{4.40c}$$

$$t' = \frac{t}{t_0} = \sqrt{\frac{g}{\ell}}\,t \qquad \text{dimensionless time} \tag{4.40d}$$

$$\omega = \frac{\omega_d}{\omega_0} = \sqrt{\frac{\ell}{g}}\,\omega_d \qquad \text{driving angular frequency} \tag{4.40e}$$

Note that

$$\dot{x} = \frac{dx}{dt'} = \frac{d\theta}{dt}\frac{dt}{dt'} = \frac{d\theta}{dt}\frac{1}{\omega_0}$$

$$\ddot{x} = \frac{d^2x}{dt'^2} = \frac{d^2\theta}{dt^2}\left(\frac{dt}{dt'}\right)^2 = \frac{d^2\theta}{dt^2}\frac{1}{{\omega_0}^2} = \frac{\ddot{\theta}}{{\omega_0}^2}$$

Using these variables and parameters. Equation 4.39 becomes

$$\ddot{x} = -c\dot{x} - \sin x + F\cos\omega t' \tag{4.41}$$

Equation 4.41 is a nonlinear equation of the form first presented in Equation 4.1. We will use numerical methods to solve this equation for x, given the parameters c, F, and ω. The techniques mentioned in Chapter 3 are used to solve this equation, depending on the accuracy desired and computer speed available, and commercial software programs are available.

Equation 4.41, a second-order differential equation, can be reduced to two first-order equations by making the substitution

$$y = \frac{dx}{dt'} \tag{4.42}$$

Equation 4.41 becomes a first-order differential equation

$$\frac{dy}{dt'} = -cy - \sin x + F\cos z \tag{4.43}$$

where we have also made the substitution $z = \omega t'$. Equations 4.42 and 4.43 are the first-order differential equations.

We present the results of numerical methods solutions in Figure 4-19.

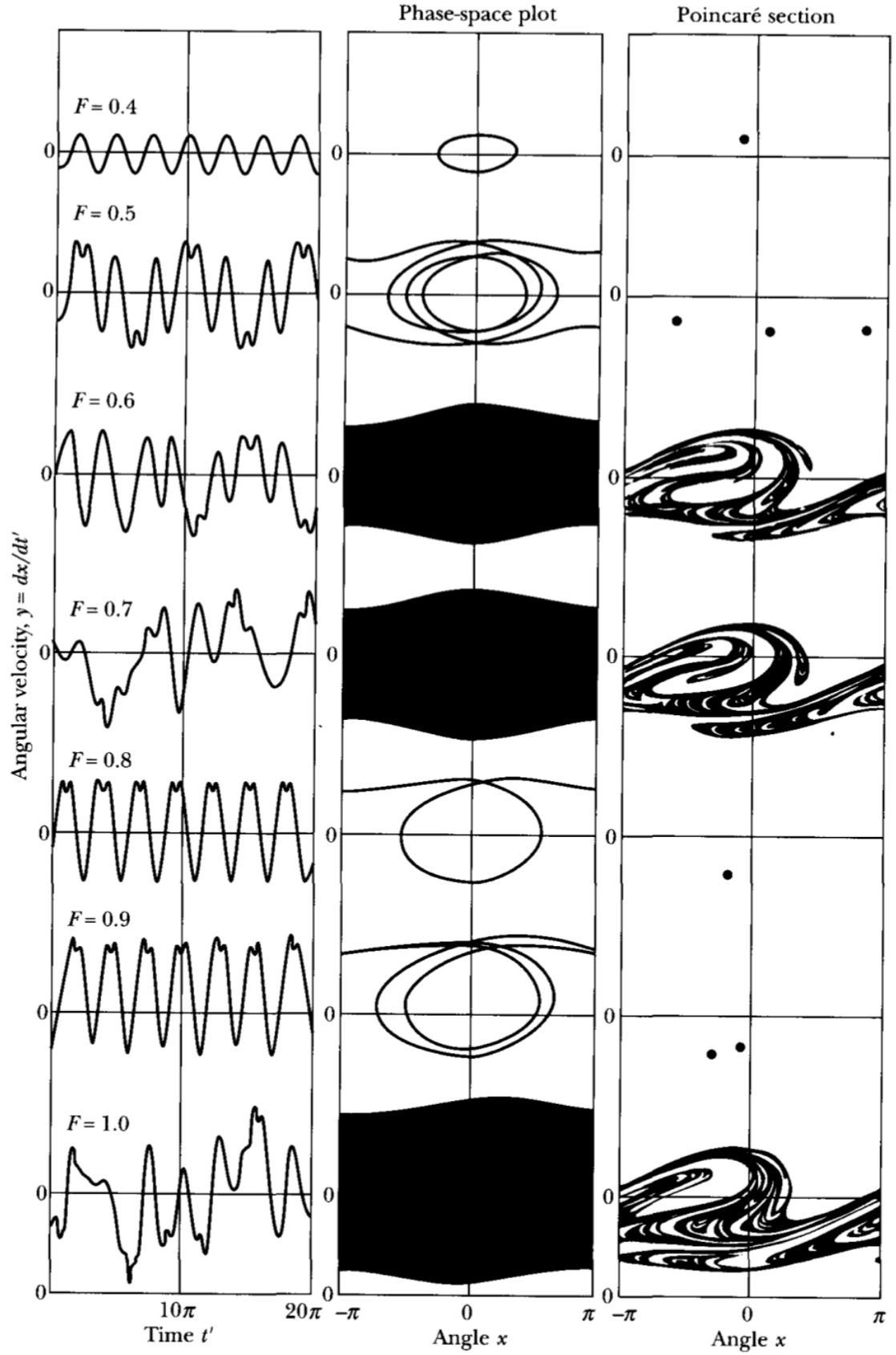


FIGURE 4-19 The damped and driven pendulum for various values of the driving force strength. The angular velocity versus time is shown on the left, and phase diagrams are in the center. Poincaré sections are shown on the right. Note that motion is chaotic for the driving force F values of 0.6, 0.7, and 1.0.

We leave the parameters c and ω set at 0.05 and 0.7, respectively, and vary only the driving strength Fin steps of 0.1 from 0.4 to 1.0. The results are that the motion is periodic for F values of 0.4, 0.5, 0.8, and 0.9 but is chaotic for 0.6, 0.7, and 1.0. These results indicate the beautiful and surprising results obtained from nonlinear dynamics. The left side of Figure 4-19 displays y = dx/dt' (angular velocity) versus time long after the initial motion (i.e., transient effects have died out). The value of F= 0.4 shows simple harmonic motion, but the results for 0.5, 0.8, and 0.9, although periodic, are hardly simple.

We can learn more by examining the phase space plots, shown in the middle column of Figure 4-19 (note that we present only a unit cell of the phase diagram from $-\pi$ to π . As expected, the result for F = 0.4 shows the results seen previously in Chapter 3 (Figure 3-5). The phase plot for F= 0.5 shows one long cycle that includes two complete revolutions and two oscillations. The entire allowed area in the phase plane is accessed chaotically for F= 0.6 and 0.7, but for F = 0.8, the motion becomes periodic again with one complete revolution and an oscillation. The result for F = 0.9 is interesting, because there appears to be two different revolutions in one cycle, each similar to the one for F= 0.8. This result is called period doubling (i.e., the period for F = 0.9 is twice the period for F= 0.8). After close inspection, this effect can also be observed from the dx/dt' versus time plot, shown on the left column of Figure 4-19.

Poincare Section

Henry Poincare invented a technique to simplify the representations of phase space diagrams, which can become quite complicated. It is equivalent to taking a stroboscopic view of the phase space diagram. A three-dimensional phase diagram plots y($= \dot{x} = \dot{\theta}$) versus x ($= \theta$) versus z($= \omega t'$). The left column of Figure 4-19 is a projection of this plot onto a y-z plane, showing points that correspond to various values of phase angle x. The middle column of Figure 4-19 is a projection onto a y-z plane, showing points belonging to various values of z. In Figure 4-20 we show the three-dimensional phase space diagram intersected by a set of y-x planes, perpendicular to the z-axis, at equal z intervals.

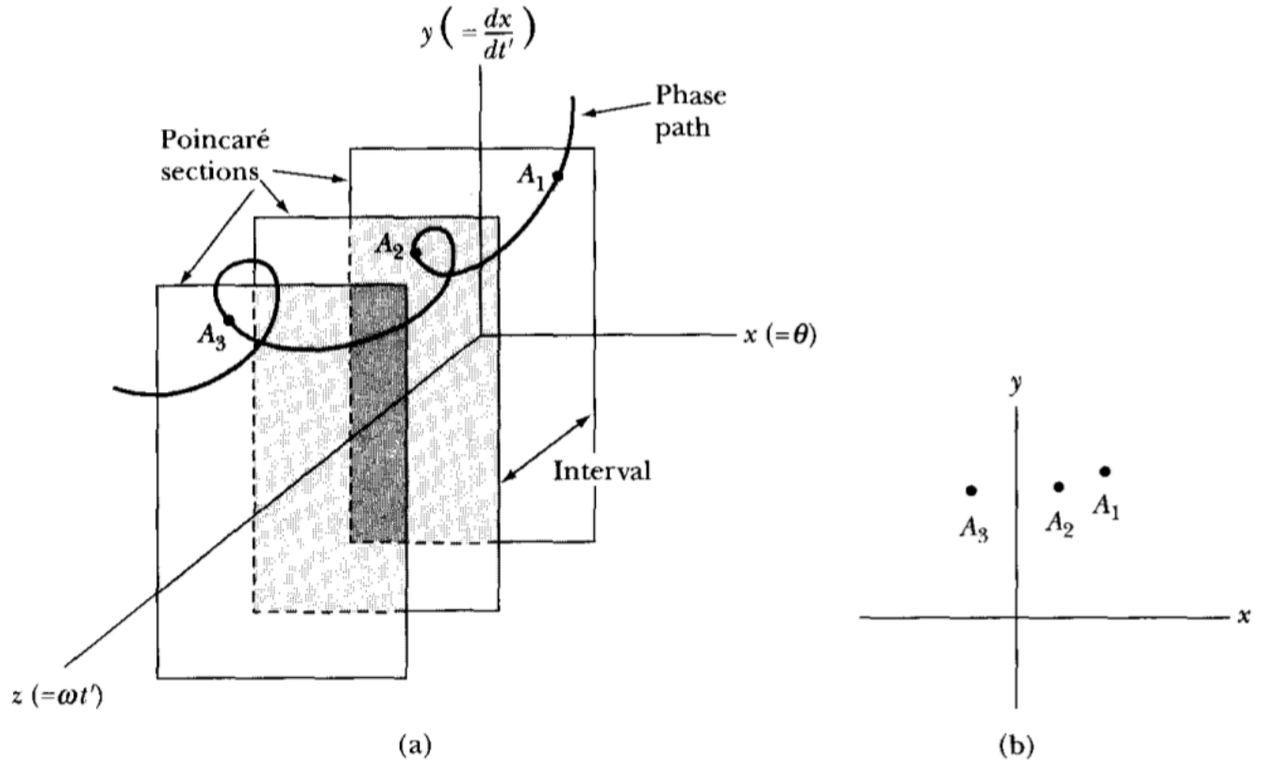


FIGURE 4-20 (a) Poincaré plot, a three-dimensional phase diagram, showing three Poincaré sections and the phase path. The sections are projections along the y-x plane. (b) The points A_i are the phase path intersections with the section plots. They are plotted here on the y-x plane to help visualize the motion in phase space.

A *Poincare section plot* is the sequence of points formed by the intersections of the phase path with these parallel planes in phase space, projected onto one of the planes. The phase path pierces the planes as a function of angular speed (y = $\dot{\theta}$), time (z = ω t'), and phase angle (x = θ). The points on the intersections are labeled as A_1, A_2, A_3, etc. This set of points A, forms a pattern when projected onto one of the planes (Figure 4-20b) that sometimes will be a recognizable curve, but sometimes will appear irregular.

For simple harmonic motion, such as F = 0.4 in Figure 4-19, all the points projected are the same (or in a smooth curve, depending on the z spacing of the y-x planes). Poincare realized that the simple curves represent motion with possibly analytic solutions, but the many complicated, apparently irregular, curves represent chaos. The Poincare section curve effectively reduces an N-dimensional diagram to (N - 1)-dimensions for graphical purposes and often helps visualize the motion in phase space.

For the case of the damped and driven pendulum, the regularity of the dynamical motion is due to the forcing period, and a complete description of the dynamical motion depends on three parameters. We can take those parameters to be x (angle θ), y = dx/dt' (angular frequency), and z = ω t' (phase of the driving force). A complete description of the motion in phase space would require three-dimensional phase diagrams rather than displaying just two parameters as in Figure 4-19. All the values of z are included in the middle column of Figure 4-19, so we choose to take the stroboscopic sections of the motion for just the values of $z = 2n\pi$ (n = 0, 1, 2, . . .), which is at a frequency equal to that of the driving force.

We show the Poincare section for the pendulum in the right column of Figure 4-19 for the same systems displayed in the left and middle columns. For the simple motion of F = 0.4, the system always comes back to the same position of (x, y) after z goes through 2π. Therefore, we expect the Poincare section to show only one point, and that is what we find in the top figure of the right column of Figure 4-19. The motion for F = 0.8 also shows only one point, but F = 0.5 and 0.9 show three and two points, respectively, because of the more complex motion. The number of points n on the Poincare section here shows that the new period $T = T_0 n/m$, where $T_0 = 2\pi/\omega$ is the period of the driven force and m is an integer (m = 2 for the F = 0.5 plot and m = 1 for the F = 0.9 plot). The chaotic motions for F = 0.6, 0.7, and 1.0 display the complicated variation of points expected for chaotic motion with a period $T \to \infty$. The Poincare sections are also rich in structure for chaotic motion.

On three occasions thus far (Figures 4-5, 4-7, and 4-11), we have pointed out attractors, a set of points (or a point) on which the motion converges for dissipative systems. The regions traversed in phase space are strictly bounded when there is an attractor In chaotic motion, nearby trajectories in phase space are continually diverging from one another but must eventually return to the attractor Because the attractors in these chaotic motions, called strange or chaotic attractors, are necessarily bounded in phase space, the attractors must fold back into the nearby regions of phase space. Strange attractors create intricate patterns, because the folding and stretching of the trajectories must occur such that no trajectory in phase space intersects, which is ruled out by the deterministic dyamical motion. T

The Poincare sections of Figure 4-19 reveal the folded, layered structure of the attractors. Chaotic attractors are fractals, but space does not permit further discussion of this extremely interesting phenomenon.

4.7 Mapping

If we use n to denote the time sequence of a system and x to denote a physical observable of the system, we can describe the progression of a nonlinear system at a particular moment by investigating how the *(n + 1)*th state (or iterate) depends on the *n*th state. An example of such a simple, nonlinear behavior is $x_{n+1} = (2x_n+3)^2$. This relationship, $x_{n+1} = f(x_n)$, is called **mapping** and is often used to describe the progression of the system. The Poincare section plots previously discussed are examples of two-dimensional maps. A physical example appropriate for mapping might be the temperature of the space shuttle orbiter tiles while the shuttle descends through the atmosphere. After the orbiter has been on the ground for some time, the temperature T_{n+1} is the same as T_n but this was not true while the shuttle plummeted through the atmosphere from its earth orbit. Modeling the tile temperatures correctly with a mathematical model is difficult, and linear assumptions are often first assumed in such calculations with nonlinear terms added to make more realistic calculations.

We can write a *difference equation* using $f(a, x_n)$ where x_n is restricted to a real number in the interval (0, 1) between 0 and 1, and a is a model-dependent parameter

$$x_{n+1} = f(\alpha, x_n) \tag{4.44}$$

The function $f(a, x_n)$ generates the value of x_{n+1} from x_n and the collection of points generated is said to be a **map** of the function itself. The equations, which are often nonlinear, are amenable to numerical solution by iteration, Starting with x_1. We will restrict ourselves here to one-dimensional maps, but two-dimensional (and higher order) equations are possible.

Mapping can best be understood by looking at an example. Let us consider the "logistic" equation, a simple one-dimensional equation given by

$$f(\alpha, x) = \alpha x(1 - x) \tag{4.45}$$

so that the iterative equation becomes

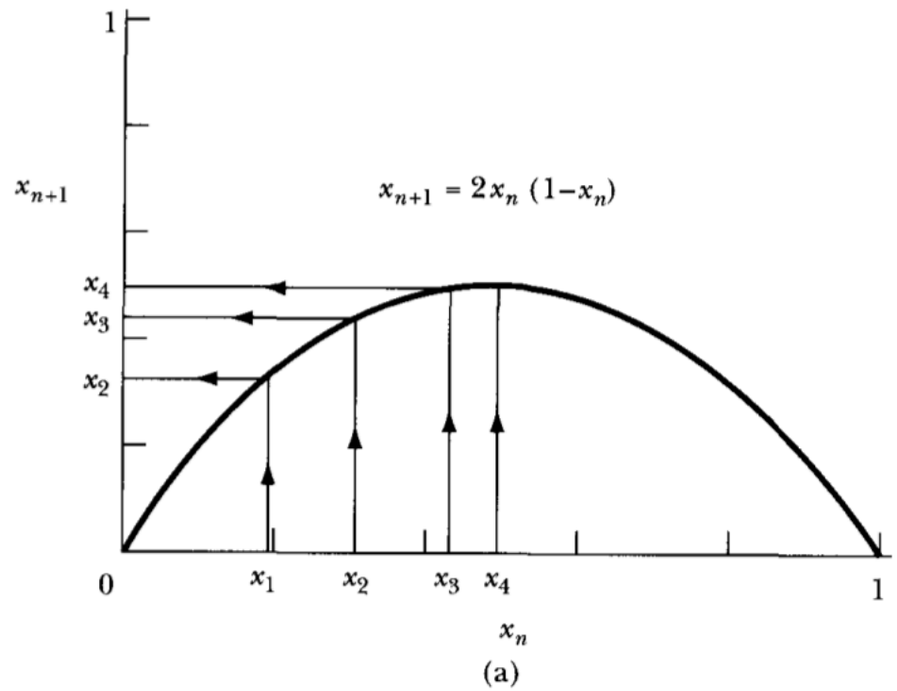
$$x_{n+1} = \alpha x_n(1 - x_n) \tag{4.46}$$

In this discussion we will use the logistic equation in a biological application example of studying the population growth of fish in a pond, where the pond is well isolated from external effects such as weather. The iterations, or n values, represent the annual fish population, where x_1 is the number of fish in the pond at the beginning of the first year of the study. If x_1 is small, the fish population may grow rapidly in the early years because of available resources, but overpopulation may eventually deplete the number of fish. The population x_n is scaled so that its value fits in the interval (0, 1) between 0 and 1. The factor α is a model-dependent parameter representing average effects of environmental factors (e.g., fishermen, floods, drought, predators) that may affect the fish. The factor a may be varied as desired in the study, but experience shows that a should be limited in this example to the interval (0, 4) to prevent the fish population from becoming negative or infinite.

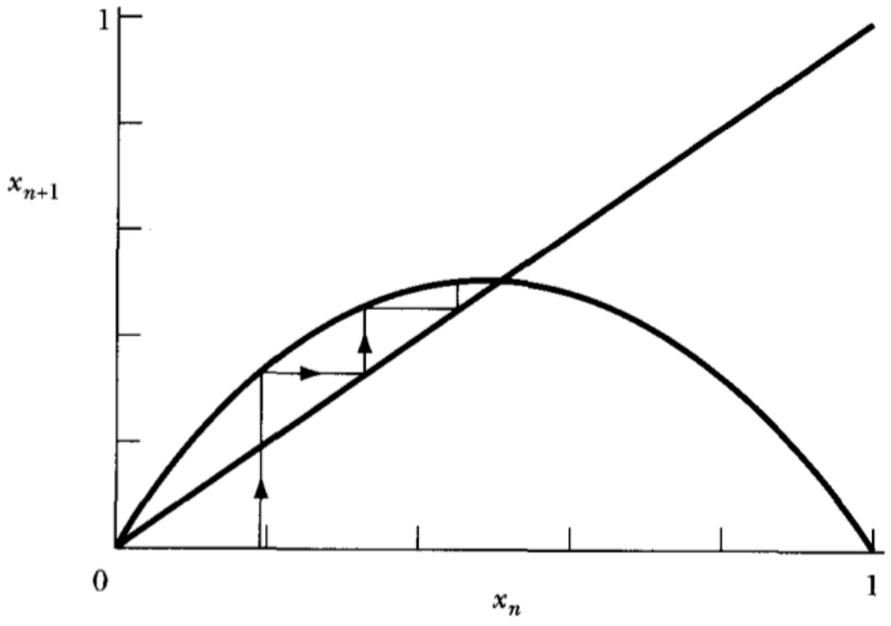
The results of the logistic equation are most easily observed by graphical means in a map called the *logistic map*. The iteration x_{n+1} is plotted versus x_n in Figure 4-21a for a value of $\alpha = 2.0$. Starting with an initial value x_1 on the horizontal (xn) axis, we move up until we intersect with the curve $x_{n+1} = 2x_n(1 - x_n)$, and then we move to the left where we find x_2 on the vertical axis (x_{n+1}). We then start with this value of x_2 on the horizontal axis and repeat the process to find x_3 on the vertical axis. If we do this for a few iterations, we converge on the value x = 0.5, and the fish population stabilizes at half its maximum. We arrive at this result independent of our initial value of x_1 as long as it is not 0 or 1.

An easier way to follow the process is to add the 45° line, $x_{n+1} = x_n$ to the same graph. Then after initially intersecting the curve from x_1, one moves horizontally to intersect with the 45° line to find x_2 and then moves up vertically to find the next iterative value of x_3.

This process can go on and reach the same result as in Figure 4-2la. We show the process in Figure 4-2lb to indicate that this method is easier to use than the one without the 45° line.



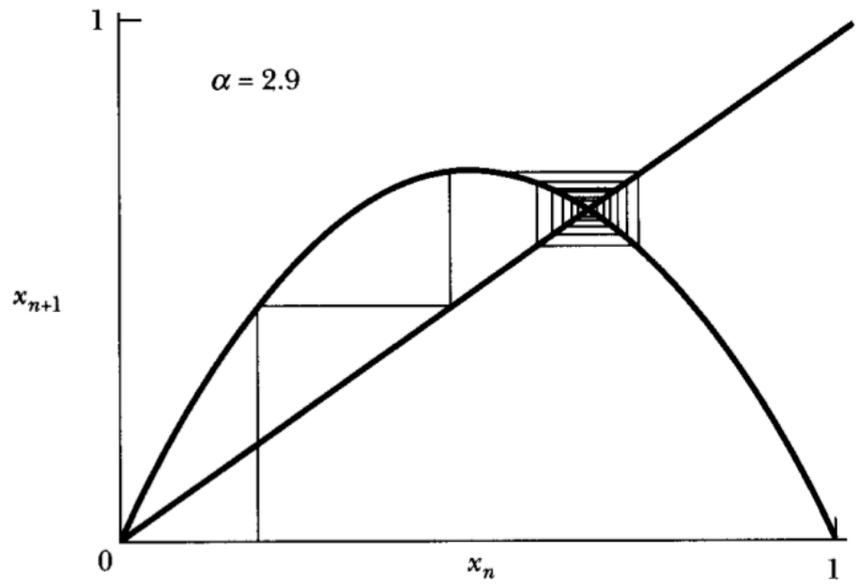
(a)



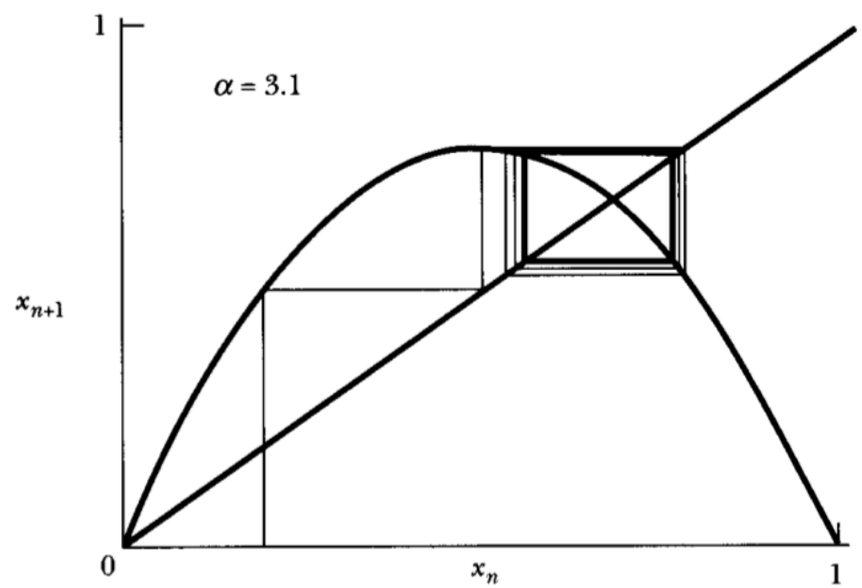
(b)

FIGURE 4-21 Techniques for producing a map of the logistics equation.

In practice, we want to study the behavior of the system when the model parameter α is varied. In the present case, for values of α less than 3.0, stable populations will result (Figure 4-22a). The solutions follow a square spiral path to the central, final value. For values of α just above 3.0, more than one solution for the fish population occurs (Figure 4-22b). The solutions follow a path similar to the square spiral, which converges to the two points at which the square intersects the "iteration line", rather than to a single point. Such a change in the number of solutions to an equation, when a parameter such as α is varied, is called a **bifurcation**.



(a)



(b)

FIGURE 4-22 Logistic equation map for α values of 2.9 and 3.1, indicating stable populations in (a) and multiple possible solutions for $\alpha > 3.0$ in (b).

We obtain a more general view of the global picture by plotting a *bifurcation diagram*, which consists of x_n, determined after many iterations to avoid initial effects, plotted as a function of the model α parameter a. Many new interesting effects emerge indicating regions and *windows* of stability as well as those of chaotic dynamics. We show the bifurcation diagram in Figure 4-23 for the logistic equation over the range of α values from 2.8 to 4.0.

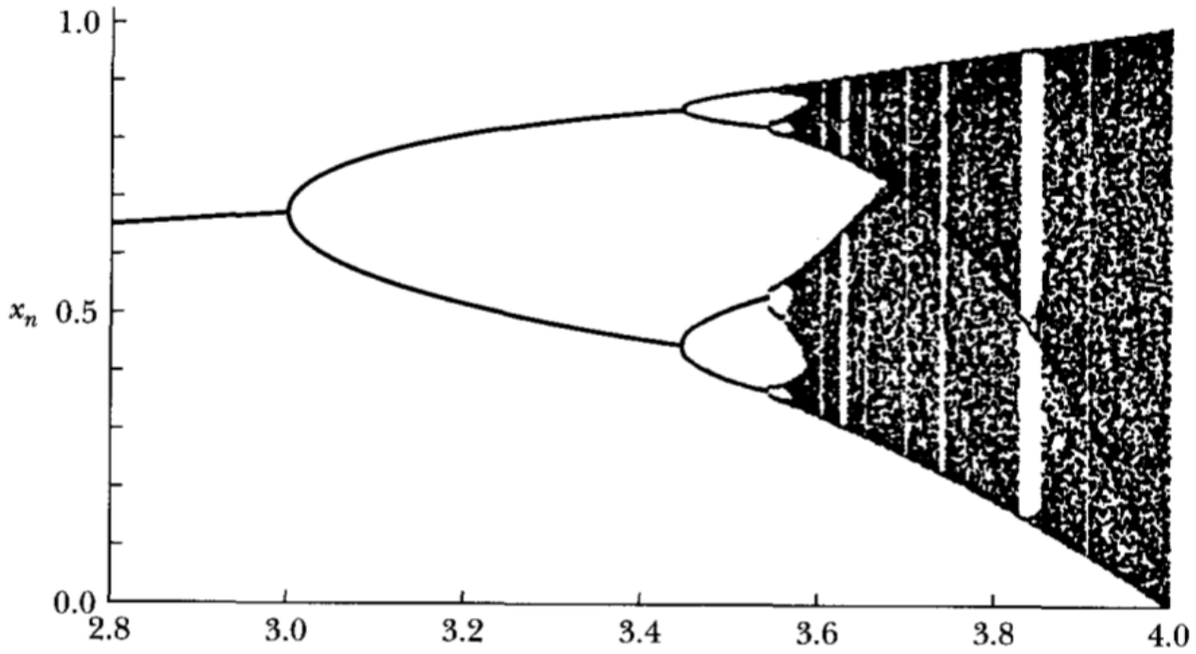


FIGURE 4-23 Bifurcation diagram for the logistic equation map.

For the value of $\alpha = 2.9$ shown in Figure 4-22a, we observe that after a few iterations, a stable configuration for x = 0.655 results. An *N cycle* is an orbit that returns to its original position after *N* iterations, that is, $x_{N+i} = x_i$. The period for $\alpha = 2.9$ is then a one cycle. For $\alpha = 3.1$ (Figure 4-22b), the value of x oscillates between 0.558 and 0.765 (two cycle) after a few iterations evolve. The bifurcation occurring at 3.0 is called a *pitchfork bifurcation* because of the obvious shape of the diagram caused by the splitting. At $\alpha = 3.1$, the period doubling effect has $x_{n+2} = x_n$. At $\alpha = 3.45$, the two-cycle bifurcation evolves into a four cycle, and the bifurcation and period doubling continues up to an infinite number of cycles near $\alpha = 3.57$.

Chaos occurs for many of the a values between 3.57 and 4.0, but there are still windows of periodic motion, with an especially wide window around 3.84. A really interesting behavior occurs for $\alpha = 3.82831$. An apparent periodic cycle of 3 years seems to occur for several periods, but then it suddenly violently changes for a few years, and then returns again to the 3-year cycle. This intermittent behavior could certainly prove devastating to a biological study operating over several years that suddenly turns chaotic without apparent reason.

EXAMPLE 4.2

Let $\Delta\alpha_n = \alpha_n - \alpha_{n-1}$ be the width between successive period doubling bifurcations of the logistic map that we have been discussing. For example, from Figure 4-23, we let $\alpha_1 = 3.0$ where the first bifurcation occurs and $\alpha_2 = 3.449490$ where the next one occurs. Let δ_n be defined as the ratio

$$\delta_n = \frac{\Delta\alpha_n}{\Delta\alpha_{n+1}} \tag{4.47}$$

and let $\delta_n \to \delta$ as $n \to \infty$. Find δ_n for the first few bifurcations and the limit δ.

Solution. We make a table

n	α_n	$\Delta\alpha$	δ_n
1	3.0		
2	3.449490	0.449490	4.7515
3	3.544090	0.094600	4.6562
4	3.564407	0.020317	4.6684
5	3.568759	0.004352	
∞	3.5699456		4.6692

As α_n approaches the limit 3.5699456, the number of period doublings approaches infinity, and the ratio δ_n called *Feigenbaum's number*, approaches 4.669202. This result was first found by Mitchell Feigenbaum in the 1970s, and he found that the limit δ was a universal property of the period doubling route to chaos when the function $f(\alpha, x)$ has a quadratic maximum. It is a remarkable fact that this universality is not confined to one-dimensional mappings; it is also true for two-dimensional maps and has been confirmed for several cases. Feigenbaum claims to have found this result using a programmable hand calculator. The calculation obviously has to be carried to many significant figures to establish its accuracy, and such a calculation was not possible before such calculators (or computers) were available.

4.8 Chaos Identification

In our driven and damped pendulum, we found that chaotic motion occurs for some values of the parameters, but not for others. What are the characteristics of chaos and how can we identify them? Chaos does not represent periodic motion, and its limiting motion will not be periodic. Chaos can generally be described as having a sensitive dependence on initial conditions. We can demonstrate this effect by the following example.

EXAMPLE 4.3

Consider the nonlinear relation $x_{n+1} = f(\alpha, x_n) = \alpha x_n(1 - x_n^2)$. Let α = 2.5 and make two numerical calculations with initial x_1 values of 0.700000000 and 0.700000001. Plot the results and find the iteration w where the solutions have clearly diverged.

Solution. The iterative equation that we are considering is

$$x_{n+1} = \alpha x_n(1 - x_n{}^2) \tag{4.48}$$

We perform a short numeric calculation and plot the results of iterations for the two initial values on the same graph. The result is shown in Figure 4-24 where there is no observed difference for x_{n+1} until n reaches at least 30.

By n = 39, the difference in the two results is marked, despite the original values differing by only 1 part in 10^8.

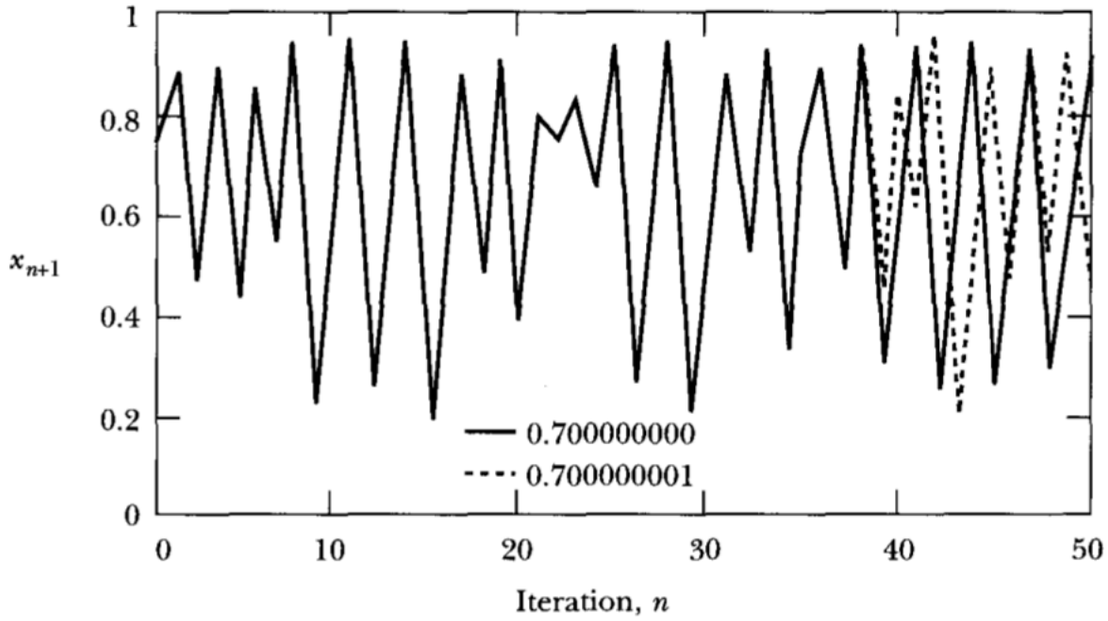


FIGURE 4-24 Example 4.3. The $n + 1$ iterative state is plotted versus the number of iterations and shows two eventual results for slightly different initial conditions of x_1.

If the computations are made without error, and the difference between iterated values doubled on the average for each iteration, then there will be an exponential increase such as

$$2^n = e^{n \ln 2}$$

where n is the number of iterations undergone. For the iterates to be separated by the order of unity (the size of the attractor), we will have

$$2^n 10^{-8} \sim 1$$

which gives n = 27. That is, after 27 iterations, the difference between the two iterates reaches the full range of x_n. To have the results differ by unity for n = 40 iterations, we would have to know the initial values with a precision of 1 part in 10^{12}!

The previous example indicates the sensitive dependence on initial conditions that is characteristic of chaos.

The two results can still be determined in this case, but it is rare to know the initial values to a precision of 10^{-8}. If we add another factor of 10 to the precision of x_1, we gain only four interative steps of agreement in the calculation. We must accept the reality that increasing the precision of the initial conditions only gains us a little in the accuracy of the ultimate measurement. This exponential growth of an initial error will ultimately prevent us from predicting the outcome of a measurement.

The effect of sensitive dependence on initial conditions has been called the **"butterfly" effect**. A butterfly moving slowly through the air may cause an extremely small effect on the airflow that will prevent us from predicting the weather patterns next week. Background noise or thermal effects will usually add uncertainties larger than the ones we have discussed here, and we cannot distinguish these effects from measurement errors. Precise predictive power of many steps is just not possible.

Lyapunov Exponents

One method to quantify the sensitive dependence on initial conditions for chaotic behavior uses the Lyapunov characteristic exponent. It is named after the Russian mathematician A. M. Lyapunov. There are as many Lyapunov exponents for a particular system as there are variables. We will limit ourselves at first to considering only one variable and therefore one exponent. Consider a system with two initial states differing by a small amount; we call the initial states x_0 and $x_0 + \varepsilon$. We want to investigate the eventual values of x_n after n iterations from the two initial values. The Lyapunov exponent λ represents the coefficient of the average exponential growth per unit time between the two states. After n iterations, the difference d_n between the two x_n values is approximately

$$d_n = \varepsilon e^{n\lambda} \tag{4.49}$$

From this equation, we can see that if λ is negative, the two orbits will eventually converge, but if positive, the nearby trajectories diverge and chaos results.

Let us look at a one-dimensional map described by $x_{n+1} = f(x_n)$. The initial difference between the states is $d_0 = \varepsilon$, and after one iteration, the difference d_1 is

$$d_1 = f(x_0 + \varepsilon) - f(x_0) \simeq \varepsilon \left.\frac{df}{dx}\right|_{x_0}$$

where the last result on the right side occurs because ε is very small. After n iterations, the difference d_n between the two initially nearby states is given by

$$d_n = f^n(x + \varepsilon) - f^n(x_0) = \varepsilon e^{n\lambda} \tag{4.50}$$

where we have indicated the *n*th iterate of the map f(x) by the superscript n. If we divide by ε and take the logarithm of both sides, we have

$$\ln\left(\frac{f^n(x + \varepsilon) - f^n(x_0)}{\varepsilon}\right) = \ln(e^{n\lambda}) = n\lambda$$

and because ε is very small, we have for λ,

$$\lambda = \frac{1}{n}\ln\left(\frac{f^n(x + \varepsilon) - f^n(x_0)}{\varepsilon}\right) = \frac{1}{n}\ln\left|\left.\frac{df^n(x)}{dx}\right|_{x_0}\right| \tag{4.51}$$

The value of $f^n(x_0)$ is obtained by iterating the function $f(x_0)$ n times.

$$f^n(x_0) = f(f(\cdots(f(x_0))\cdots))$$

We use the derivative chain rule of the *n*th iterate to obtain

$$\left.\frac{df^n(x)}{dx}\right|_{x_0} = \left.\frac{df}{dx}\right|_{x_{n-1}} \left.\frac{df}{dx}\right|_{x_{n-2}} \cdots \left.\frac{df}{dx}\right|_{x_0}$$

We take the limit as $n \to \infty$ and finally obtain

$$\lambda = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \ln\left|\frac{df(x_i)}{dx}\right| \tag{4.52}$$

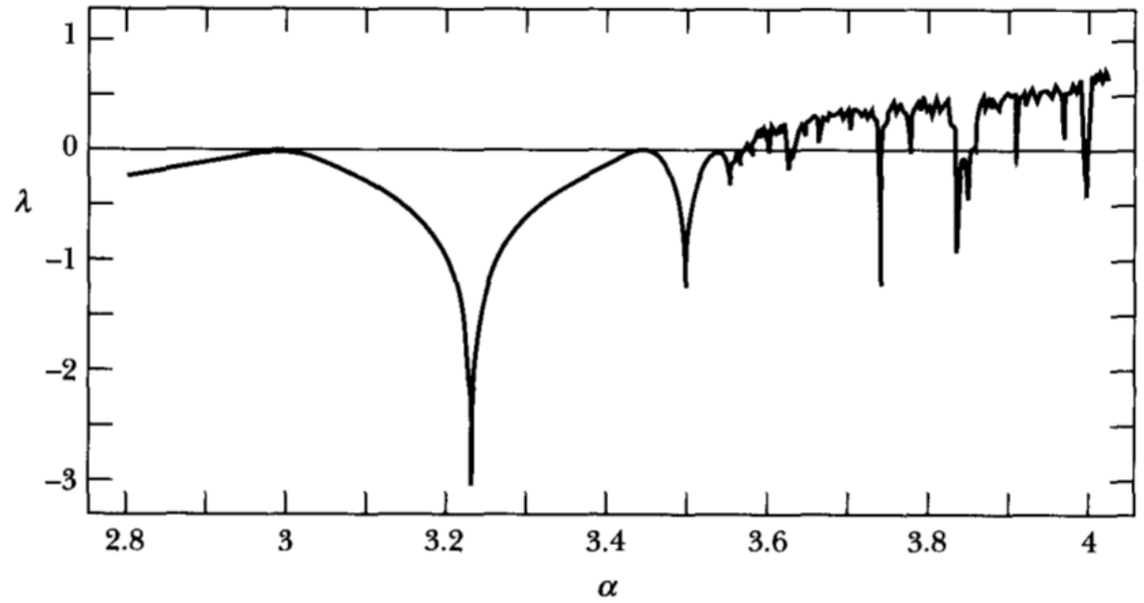


FIGURE 4-25 Lyapunov exponent as a function of α for the logistic equation map. A value of $\lambda > 0$ indicates chaos.

We plot the Lyapunov exponent as a function of a in Figure 4-25 for the logistic map. We note the agreement of the sign of λ with the discussion of chaotic behavior in Section 4.6. The value of λ is zero when bifurcation occurs, because $|df/dx| = 1$, and the solution becomes unstable. A super-stable point occurs where $df(x)/dx = 0$, and this implies that $\lambda = -\infty$. From Figure 4-25 as λ goes above 0, we see there are windows where λ returns to $\lambda < 0$ and periodic orbits occur amid the chaotic behavior. The relatively wide window just above 3.8 is apparent.

Remember that for n dimensional maps, there will be n Lyapunov exponents. Only one of them need be positive for chaos to occur. For dissipative systems, the phase space volume will decrease as time passes. This means the sum of the Lyapunov exponents will be negative.

The calculation of Lyapunov exponents for the damped and driven pendulum is difficult, because one has to deal with the solutions of differential equations rather than maps such as those of the logistic equation. Nevertheless, these calculations have been done, and we show in Figure 4-26 the Lyapunov exponents, three of them because of the three dimensions.

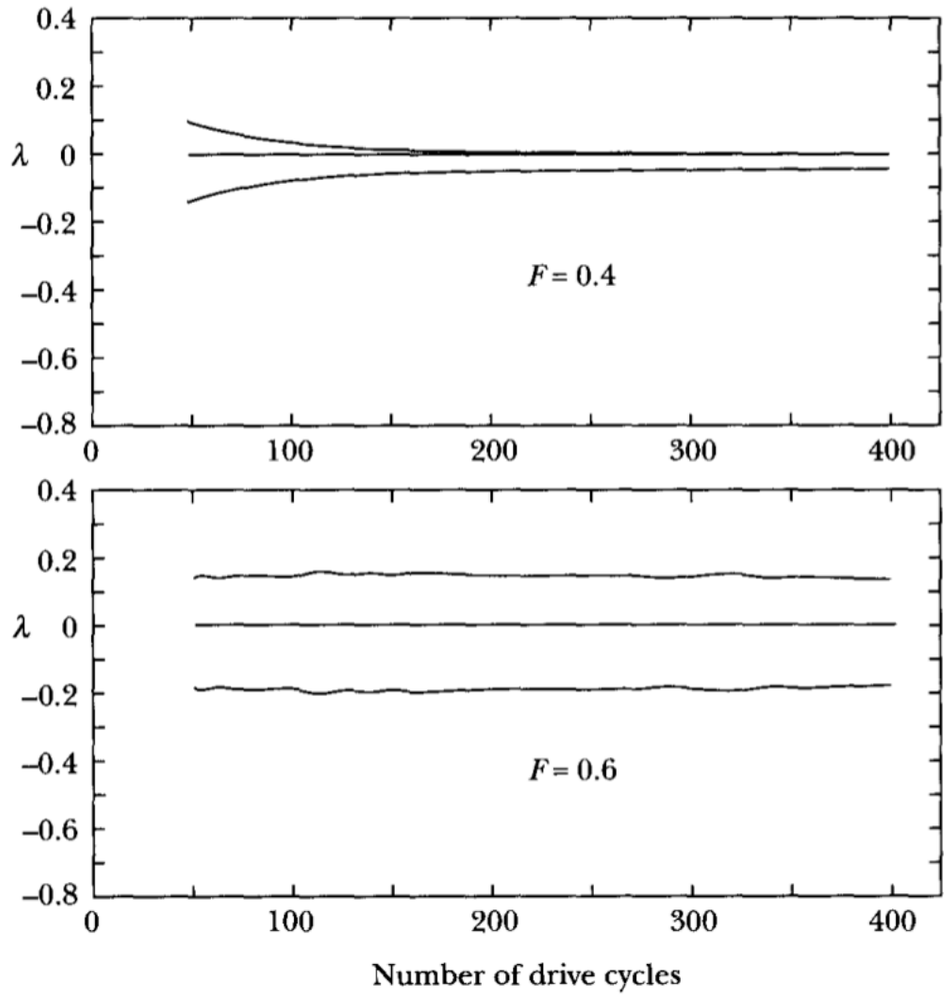


FIGURE 4-26 The three Lyapunov exponents for the damped and driven pendulum. The values of λ are those approached as $t \to \infty$ (large number of cycles).

The parameters are the same as those discussed in Section 4.6: c = 0.05, ω = 0.7, and F = 0.4 (periodic) and F = 0.6 (chaotic). For both cases, we must make at least several hundred iterations to make sure transient effects have died out. Note that one of the Lyapunov exponents is zero, because it does not contribute to the expansion or contraction of the phase space volume. For the case of F = 0.4, none of the Lyapunov exponents is greater than zero after 350 iterations, but for the F = 0.6 driven case, one of the exponents is still well above zero. The motion is chaotic for F = 0.6, as we found earlier in Figure 4-19. However, because the motion described in Figure 4-26 is damped, the sum of the three Lyapunov exponents is negative for both cases, as it should be.

CHAPTER 5 - Gravitation

5.1 Introduction

By 1666, Newton had formulated and numerically checked the gravitation law he eventually published in his book Principia in 1687. Newton waited almost 20 years to publish his results because he could not justify his method of numerical calculation in which he considered Earth and the Moon as point masses. With mathematics formulated on calculus (which Newton later invented), we have a much easier time proving the problem Newton found so difficult in the seventeenth century.

Newton's law of universal gravitation states that *each mass particle attracts every other particle in the universe with a force that varies directly as the product of the two masses and inversely as the square of the distance between them*. In mathematical form, we write the law as

$$\mathbf{F} = -G\frac{mM}{r^2}\mathbf{e}_r \tag{5.1}$$

where at a distance r from a particle of mass M a second particle of mass m experiences an attractive force (see Figure 5-1).

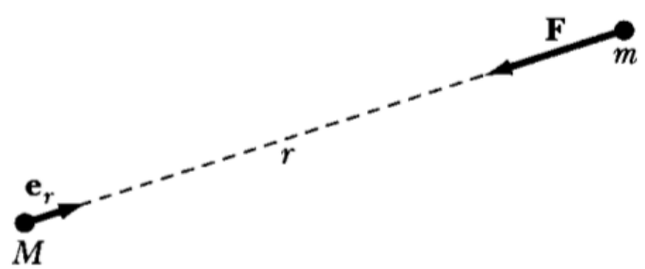


FIGURE 5-1 Particle *m* feels an attractive gravitational force toward *M*.

The unit vector $\mathbf{e}_r$ points from M to m, and the minus sign ensures that the force is attractive - that is, that m is attracted toward M.

A laboratory verification of the law and a determination of the value of G was made in 1798 by the English physicist Henry Cavendish. Cavendish's experiment, described in many elementary physics texts, used a torsion balance with two small spheres fixed at the ends of a light rod. The two spheres were attracted to two other large spheres that could be placed on either side of the smaller spheres. The official value for G is $6.673 \pm 0.010 \times 10^{-11}$ N m^2/kg^2. Interestingly, although G is perhaps the oldest known of the fundamental constants, we know it with less precision than we know most of the modern fundamental constants such as e, c, and $\hbar$. Considerable research is ongoing today to improve the precision of G.

In the form of Equation 5.1, the law strictly applies only to *point particles*. If one or both of the particles is replaced by a body with a certain extension, we must make an additional hypothesis before we can calculate the force. We must assume that the gravitational force field is a *linear field*. In other words, we assume that it is possible to calculate the net gravitational force on a particle due to many other particles by simply taking the vector sum of all the individual forces. For a body consisting of a continuous distribution of matter, the sum becomes an integral (Figure 5-2):

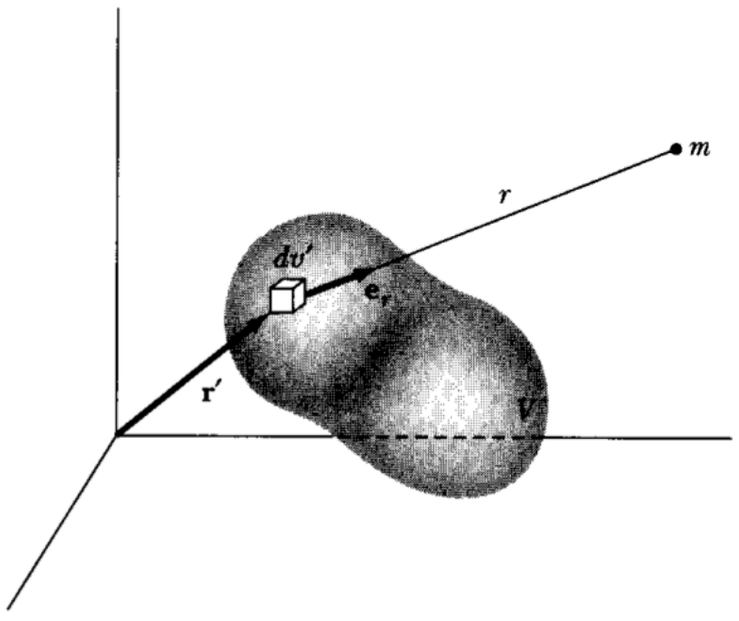


FIGURE 5-2 To find the gravitational force between a point mass m and a continuous distribution of matter, we integrate the mass density over the volume.

$$\mathbf{F} = -Gm\int_V \frac{\rho(\mathbf{r}')\mathbf{e}_r}{r^2}dv' \tag{5.2}$$

where $\rho(\mathbf{r}')$ is the mass density and dv' is the element of volume at the position defined by the vector **r**' from the (arbitrary) origin to the point within the mass distribution.

If both the body of mass M and the body of mass m have finite extension, a second integration over the volume of m will be necessary to compute the total gravitational force.

The **gravitational field vector g** is the vector representing the force per unit mass exerted on a particle in the field of a body of mass M. Thus

$$\mathbf{g} = \frac{\mathbf{F}}{m} = -G\frac{M}{r^2}\mathbf{e}_r \tag{5.3}$$

or

$$\boxed{\mathbf{g} = -G\int_V \frac{\rho(\mathbf{r}')\mathbf{e}_r}{r^2}dv'} \tag{5.4}$$

Note that the direction of $\mathbf{e}_r$ varies with r' (in Figure 5-2).

The quantity **g** has the dimensions of force per unit mass, also equal to acceleration. In fact, near the surface of the earth, the magnitude of **g** is just the quantity that we call the gravitational acceleration constant. Measurement with a simple pendulum (or some more sophisticated variation) is sufficient to show that **|g|** is approximately 9.80 m/s^2 (or 9.80 N/kg) at the surface of the earth.

5.2 Gravitational Potential

The gravitational field vector **g** varies as $1/r^2$ and therefore satisfies the requirement ($\nabla \times \mathbf{g} = 0$) that permits **g** to be represented as the gradient of a scalar function. Hence, we can write

$$\boxed{\mathbf{g} \equiv -\nabla \Phi} \tag{5.5}$$

where Φ is called the gravitational potential and has dimensions of (*force per unit mass*) x (*distance*), or *energy per unit mass*.

Because **g** has only a radial variation, the potential Φ can have at most a variation with r. Therefore, using Equation 5.3 for **g**, we have

$$\nabla \Phi = \frac{d\Phi}{dr}\mathbf{e}_r = G\frac{M}{r^2}\mathbf{e}_r$$

Integrating, we obtain

$$\boxed{\Phi = -G\frac{M}{r}} \tag{5.6}$$

The possible constant of integration has been suppressed, because the potential is undetermined to within an additive constant; that is, only differences in potential are meaningful, not particular values. We usually remove the ambiguity in the value of the potential by arbitrarily requiring that $\Phi \to 0$ as $r \to \infty$; then Equation 5.6 correctly gives the potential for this condition.

The potential due to a continuous distribution of matter is

$$\Phi = -G\int_V \frac{\rho(\mathbf{r}')}{r}\, dv' \tag{5.7}$$

Similarly, if the mass is distributed onlv over a thin shell (i.e., a *surface* distribution), then

$$\Phi = -G\int_S \frac{\rho_s}{r}\, da' \tag{5.8}$$

(5.8)

where ρ is the surface density of mass (or *areal mass density*).

Finally, if there is a line source with linear mass density ρ_l , then

$$\Phi = -G\int_\Gamma \frac{\rho_l}{r}\, ds'$$

The physical significance of the gravitational potential function becomes clear if we consider the work per unit mass dW′ that must be done by an outside agent on a body in a gravitational field to displace the body a distance d**r**. In this case, work is equal to the scalar product of the force and the displacement. Thus, for the work done on the body per unit mass, we have

$$\begin{aligned} dW' &= -\mathbf{g}\cdot d\mathbf{r} = (\nabla\Phi)\cdot d\mathbf{r} \\ &= \sum_i \frac{\partial\Phi}{\partial x_i}\, dx_i = d\Phi \end{aligned} \tag{5.10}$$

because Φ is a function only of the coordinates of the point at which it is measured:

$$\Phi = \Phi(x_1, x_2, x_3) = \Phi(x_i)$$

Therefore the amount of work per unit mass that must be done on a body to move it from one position to another in a gravitational field is equal to the difference in potential at the two points.

If the final position is farther from the source of mass M than the initial position, work has been done *on* the unit mass. The positions of the two points are arbitrary, and we may take one of them to be at infinity.

If we define the potential to be zero at infinity, we may interpret Φ at any point to be the work per unit mass required to bring the body from infinity to that point. The *potential energy* is equal to the mass of the body multiplied by the potential Φ. If U is the potential energy, then

$$U = m\Phi \tag{5.11}$$

and the force on a body is given by the negative of the gradient of the potential energy of that body,

$$\mathbf{F} = -\nabla U \tag{5.12}$$

which is just the expression we have previously used (Equation 2.88).

We note that both the potential and the potential energy *increase* when work is done *on* the body. (The potential, according to our definition, is always negative and only approaches its maximum value, that is, zero, as r tends to infinity.)

A certain potential energy exists whenever a body is placed in the gravitational field of a source mass. This potential energy resides in the *field* but it is customary under these circumstances to speak of the potential energy "of the body". We shall continue this practice here. We may also consider the source mass itself to have an intrinsic potential energy. This potential energy is equal to the gravitational energy released when the body was formed or, conversely, is equal to the energy that must be supplied (i.e., the work that must be done) to disperse the mass over the sphere at infinity. For example, when interstellar gas condenses to form a star, the gravitational energy released goes largely into the initial heating of the star As the temperature increases, energy is radiated away as electromagnetic radiation. In all the problems we treat, the structure of the bodies is considered to remain unchanged during the process we are studying. Thus, there is no change in the intrinsic potential energy, and it may be neglected for the purposes of whatever calculation we are making.

EXAMPLE 5.1

What is the gravitational potential both inside and outside a spherical shell of inner radius b and outer radius a?

Solution. One of the important problems of gravitational theory concerns the calculation of the gravitational force due to a homogeneous sphere. This problem is a special case of the more general calculation for a homogeneous spherical shell. A solution to the problem of the shell can be obtained by directly computing the force on an arbitrary object of unit mass brought into the field, but it is easier to use the potential method.

We consider the shell shown in Figure 5-3 and calculate the potential at point P a distance R from the center of the shell.

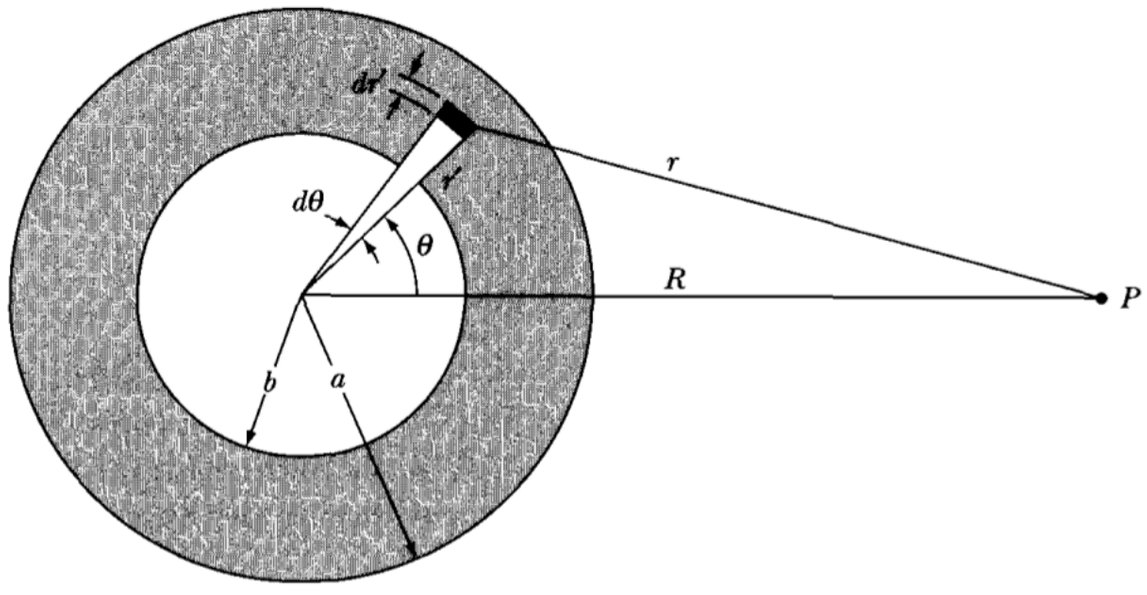


FIGURE 5-3 The geometry for finding the gravitational potential at point *P* due to a spherical shell of mass.

Because the problem has symmetry about the line connecting the center of the sphere and the field point P, the azimuthal angle ϕ is not shown in Figure 5-3 and we can immediately integrate over $d\phi$ in the expression for the potential. Thus,

$$\Phi = -G\int_V \frac{\rho(r')}{r}dv'$$
$$= -2\pi\rho G\int_b^a r'^2 dr' \int_0^\pi \frac{\sin\theta}{r}d\theta \tag{5.13}$$

where we have assumed a homogeneous mass distribution for the shell, $\rho(r') = \rho$. According to the law of cosines,

$$r^2 = r'^2 + R^2 - 2r'R\cos\theta \tag{5.14}$$

Because R is a constant, for a given r' we may differentiate this equation and obtain

$$2r\,dr = 2r'R\sin\theta\,d\theta$$

or

$$\frac{\sin\theta}{r}\,d\theta = \frac{dr}{r'R} \tag{5.15}$$

Substituting this expression into Equation 5.13, we have

$$\Phi = -\frac{2\pi\rho G}{R}\int_b^a r'\,dr' \int_{r_{min}}^{r_{max}} dr \tag{5.16}$$

The limits on the integral over dr depend on the location of point P. If P is *outside* the shell, then

$$\Phi(R > a) = -\frac{2\pi\rho G}{R}\int_b^a r'\,dr' \int_{R-r'}^{R+r'} dr$$
$$= -\frac{4\pi\rho G}{R}\int_b^a r'^2\,dr' \tag{5.17}$$
$$= -\frac{4}{3}\frac{\pi\rho G}{R}(a^3 - b^3)$$

But the mass M of the shell is

$$M = \frac{4}{3}\pi\rho(a^3 - b^3) \tag{5.18}$$

so the potential is

$$\boxed{\Phi(R > a) = -\frac{GM}{R}} \tag{5.19}$$

If the field point lies inside the shell, then

$$\begin{aligned}\Phi(R < b) &= -\frac{2\pi\rho G}{R}\int_b^a r'dr'\int_{r'-R}^{r'+R} dr \\ &= -4\pi\rho G\int_b^a r'dr' \\ &= -2\pi\rho G(a^2 - b^2)\end{aligned} \tag{5.20}$$

The potential is therefore constant and independent of position inside the shell.

Finally, if we wish to calculate the potential for points within the shell, we need only replace the lower limit of integration in the expression for $\Phi(R < b)$ by the variable R replace the upper limit of integration in the expression for $\Phi(R > a)$ by R and add the results. We find

$$\begin{aligned}\Phi(b < R < a) &= -\frac{4\pi\rho G}{3R}(R^3 - b^3) - 2\pi\rho G(a^2 - R^2) \\ &= -4\pi\rho G\left(\frac{a^2}{2} - \frac{b^3}{3R} - \frac{R^2}{6}\right)\end{aligned} \tag{5.21}$$

We see that if $R \to a$, then Equation 5.21 yields the same result as Equation 5.19 for the same limit. Similarly, Equations 5.21 and 5.20 produce the same result for the limit $R \to b$. The potential is therefore *continuous*.

If the potential were not continuous at some point, the gradient of the potential - and hence, the force - would be infinite at that point. Because infinite forces do not represent physical reality, we conclude that realistic potential functions must always be continuous.

Note that we treated the mass shell as homogeneous. In order to perform calculations for a solid, massive body like a planet that has a spherically symmetric mass distribution, we could add up a number of shells or, if we choose, we could allow the density to change as a function of radius.

The results of Example 5.1 are very important. Equation 5.19 states that the potential at any point outside of a spherically symmetric distribution of matter (shell or solid, because solids are composed of many shells) is independent of the size of the distribution. Therefore, to calculate the external potential (or the force), we consider all the mass to be concentrated at the center. Equation 5.20 indicates that the potential is constant (and the force zero) anywhere inside a spherically symmetric mass shell. And finally, at points within the mass shell, the potential given by Equation 5.21 is consistent with both of the previous results.

The magnitude of the field vector $\mathbf{g}$ may be computed from $\mathbf{g} = -d\Phi/dR$ for each of the three regions. The results are

$$\left.\begin{aligned} g(R < b) &= 0 \\ g(b < R < a) &= \frac{4\pi\rho G}{3}\left(\frac{b^3}{R^2} - R\right) \\ g(R > a) &= -\frac{GM}{R^2} \end{aligned}\right\} \tag{5.22}$$

We see that not only the potential but also the field vector (and hence, the force) are continuous. The derivative of the field vector, however, is not continuous across the outer and inner surfaces of the shell.

All these results for the potential and the field vector can be summarized as in Figure 5-4.

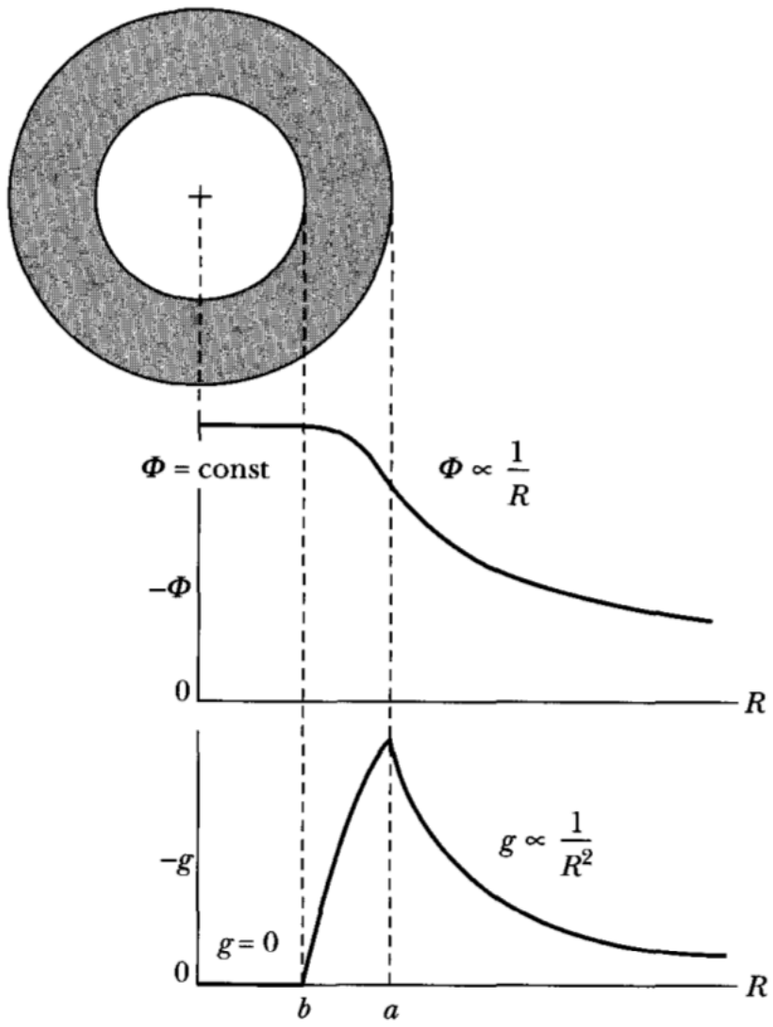


FIGURE 5-4 The results of Example 5.1 indicating the gravitational potential and magnitude of the field vector **g** (actually $-\mathbf{g}$) as a function of radial distance.

EXAMPLE 5.2

Astronomical measurements indicate that the orbital speed of masses in many spiral galaxies rotating about their centers is approximately constant as a function of distance from the center of the galaxy (like our own Milky Way and our nearest neighbor Andromeda) as shown in Figure 5-5. Show that this experimental result is inconsistent with the galaxy having its mass concentrated near the center of the galaxy and can be explained if the mass of the galaxy increases with distance R.

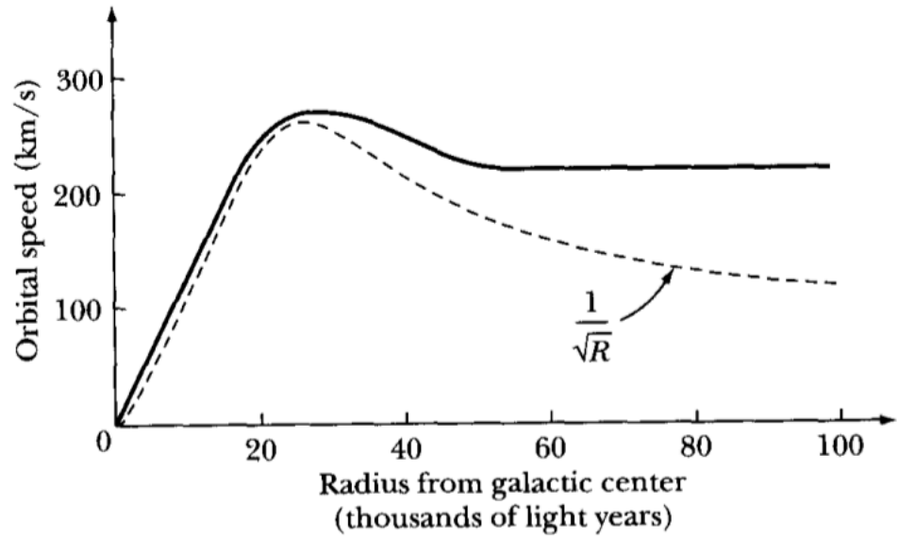


FIGURE 5-5 Example 5.2. The solid line represents data for the orbital speed of mass as a function of distance from the center of the Andromeda galaxy. The dashed line represents the $1/\sqrt{R}$ behavior expected from the Keplerian result of Newton's laws.

Solution. We can find the expected orbital speed v due to the galaxy mass M that is within the radius R. In this case, however, the distance R may be hundreds of light years. We only assume the mass distribution is spherically symmetric. The gravitational force in this case is equal to the centripetal force due to the mass m having orbital speed v.

$$\frac{GMm}{r^2} = \frac{mv^2}{R}$$

We solve this equation for v.

$$v = \sqrt{\frac{GM}{R}}$$

If this were the case, we would expect the orbital speed to decrease as $1/\sqrt{R}$ as shown by the dashed line in Figure 5-5, whereas what is found experimentally is that v is constant as a function of R. This can only happen in the previous equation if the mass M of the galaxy itself is a linear function of R, $M(R) \propto R$. Astrophysicists conclude from this result that for many galaxies there must be matter other than that observed, and that this unobserved matter, often called "dark matter," must account for more than 90 percent of the known mass in the universe. This area of research is at the forefront of astrophysics today.

Digression — An Alternative

The spiral galaxy observations necessitate at least one of the following:

> (1) There exists in galaxies large quantities of unseen matter which boosts the stars' velocities beyond what would be expected on the basis of the visible mass alone, or
>
> (2) Newton's Laws do not apply to galaxies.

Option (1) leads to the dark matter hypothesis; option (2) leads to MOND.

The basic premise of MOND is that while Newton's laws have been extensively tested in high-acceleration environments (in the Solar System and on Earth), they have not been verified for objects with extremely low acceleration, such as stars in the outer parts of galaxies. This led Milgrom to postulate a new effective gravitational force law (sometimes referred to as "Milgrom's law") that relates the true acceleration of an object to the acceleration that would be predicted for it on the basis of Newtonian mechanics. This law, the keystone of MOND, is chosen to reproduce the Newtonian result at high acceleration but leads to different ("deep-MOND") behavior at low acceleration:

$$F_{\mathrm{N}} = m\,\mu\!\left(\frac{a}{a_0}\right) a$$

Here F_{N} is the Newtonian force, m is the object's (gravitational) mass, a is its acceleration, $\mu(x)$ is an as-yet unspecified function (called the *interpolating function*), and a_0 is a new fundamental constant which marks the transition between the Newtonian and deep-MOND regimes. Agreement with Newtonian mechanics requires

$$\mu(x) \longrightarrow 1 \qquad \text{for } x \gg 1$$

and consistency with astronomical observations requires

$$\mu(x) \longrightarrow x \qquad \text{for } x \ll 1$$

Beyond these limits, the interpolating function is not specified by the hypothesis, although it is possible to weakly constrain it empirically.

Two common choices are the "simple interpolating function":

$$\mu\left(\frac{a}{a_0}\right) = \frac{1}{1+\frac{a_0}{a}}$$

and the "standard interpolating function":

$$\mu\left(\frac{a}{a_0}\right) = \sqrt{\frac{1}{1+\left(\frac{a_0}{a}\right)^2}}$$

Thus, in the deep-MOND regime ($a \ll a_0$):

$$F_{\mathrm{N}} = m\frac{a^2}{a_0}$$

Applying this to an object of mass m in circular orbit around a point mass M (a crude approximation for a star in the outer regions of a galaxy), we find:

$$\frac{GMm}{r^2} = m\frac{\left(\frac{v^2}{r}\right)^2}{a_0} \quad \Longrightarrow \quad v^4 = GMa_0$$

that is, the star's rotation velocity is independent of r, its distance from the centre of the galaxy – the rotation curve is flat, as required.

By fitting his law to rotation curve data, Milgrom found $a_0 \approx 1.2 \times 10^{-10} ms^{-2}$ to be optimal. This simple law is sufficient to make predictions for a broad range of galactic phenomena.

Milgrom's law can be interpreted in two different ways:

One possibility is to treat it as a modification to Newton's Second Law, so that the force on an object is not proportional to the particle's acceleration a but rather to

$$\mu\left(\frac{a}{a_0}\right)a$$

in this case, the modified dynamics would apply not only to gravitational phenomena, but also to those generated by other forces.

Alternatively, Milgrom's law can be viewed as leaving Newton's Second Law intact and instead modifying the inverse-square law of gravity, so that the true gravitational force on an object of mass m due to another of mass M is roughly of the form

$$\frac{GMm}{\mu\left(\frac{a}{a_0}\right)r^2}$$

In this interpretation, Milgrom's modification would apply exclusively to gravitational phenomena.

Back to standard Classical Mechanics….

EXAMPLE 5.3

Consider a thin uniform circular ring of radius a and mass M. A mass m is placed in the plane of the ring. Find a position of equilibrium and determine whether it is stable.

Solution. From symmetry, we might believe that the mass m placed in the center of the ring (Figure 5-6) should be in equilibrium because it is uniformly surrounded by mass. Put mass m at a distance r' from the center of the ring, and place the x-axis along this direction.

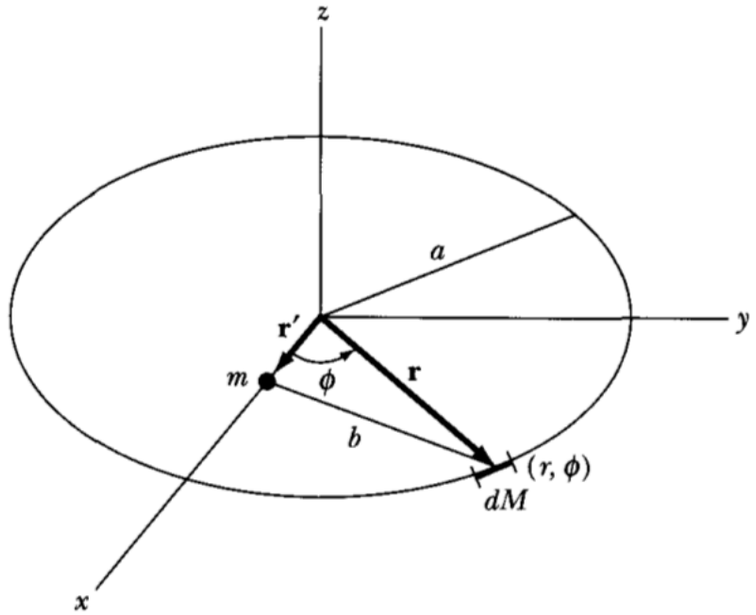


FIGURE 5-6 Example 5.3. The geometry of the point mass m and ring of mass M.

The potential is given by Equation 5.7 where $\rho = M/2\pi a$:

$$d\Phi = -G\frac{dM}{b} = -\frac{Ga\rho}{b}\,d\phi \tag{5.23}$$

where b is the distance between dM and m, and dM = $\rho a d\phi$. Let **r** and **r'** be the position vectors to dM and m, respectively.

$$\begin{aligned} b &= |\mathbf{r} - \mathbf{r}'| = |a\cos\phi\mathbf{e}_1 + a\sin\phi\mathbf{e}_2 - r'\mathbf{e}_1| \\ &= |(a\cos\phi - r')\mathbf{e}_1 + a\sin\phi\mathbf{e}_2| = [(a\cos\phi - r')^2 + a^2\sin^2\phi]^{1/2} \\ &= (a^2 + r'^2 - 2ar'\cos\phi)^{1/2} = a\left[1 + \left(\frac{r'}{a}\right)^2 - \frac{2r'}{a}\cos\phi\right]^{1/2} \end{aligned} \tag{5.24}$$

Integrating Equation 5.23 gives

$$\begin{aligned} \Phi(r') &= -G\int\frac{dM}{b} = -\rho aG\int_0^{2\pi}\frac{d\phi}{b} \\ &= -\rho G\int_0^{2\pi}\frac{d\phi}{\left[1 + \left(\frac{r'}{a}\right)^2 - \frac{2r'}{a}\cos\phi\right]^{1/2}} \end{aligned} \tag{5.25}$$

The integral in Equation 5.25 is difficult, so let us consider positions close to the equilibrium point, r' = 0. If r' << a, we can expand the denominator in Equation 5.25.

$$\left[1 + \left(\frac{r'}{a}\right)^2 - \frac{2r'}{a}\cos\phi\right]^{-1/2} = 1 - \frac{1}{2}\left[\left(\frac{r'}{a}\right)^2 - \frac{2r'}{a}\cos\phi\right] + \frac{3}{8}\left[\left(\frac{r'}{a}\right)^2 - \frac{2r'}{a}\cos\phi\right]^2 + \cdots \tag{5.26}$$

$$= 1 + \frac{r'}{a}\cos\phi + \frac{1}{2}\left(\frac{r'}{a}\right)^2(3\cos^2\phi - 1) + \cdots$$

Equation 5.25 becomes

$$\Phi(r') = -\rho G\int_0^{2\pi}\left\{1 + \frac{r'}{a}\cos\phi + \frac{1}{2}\left(\frac{r'}{a}\right)^2(3\cos^2\phi - 1) + \cdots\right\}d\phi \tag{5.27}$$

which is easily integrated with the result

$$\Phi(r') = -\frac{MG}{a}\left[1 + \frac{1}{4}\left(\frac{r'}{a}\right)^2 + \cdots\right] \tag{5.28}$$

The potential energy U(r') is from Equation 5.11, simply

$$U(r') = m\Phi(r') = -\frac{mMG}{a}\left[1 + \frac{1}{4}\left(\frac{r'}{a}\right)^2 + \cdots\right] \tag{5.29}$$

The position of equilibrium is found (from Equation 2.100) by

$$\frac{dU(r')}{dr'} = 0 = -\frac{mMG}{a}\frac{1}{2}\frac{r'}{a^2} + \cdots \tag{5.30}$$

so r' = 0 is an equilibrium point. We use Equation 2.103 to determine the stability:

$$\frac{d^2U(r')}{dr'^2} = -\frac{mMG}{2a^3} + \cdots < 0 \tag{5.31}$$

so the equilibrium point is unstable.

This last result is not obvious, because we might be led to believe that a small displacement from r' = 0 might still be returned to r' = 0 by the gravitational forces from all the mass in the ring surrounding it.

Poisson's Equation

It is useful to compare these properties of gravitational fields with some of the familiar results from electrostatics that were determined in the formulation of Maxwell's equations. Consider an arbitrary surface as in Figure 5-7 with a mass m placed somewhere inside. Similar to electric flux, let's find the gravitational flux Φ_m emanating from mass m through the arbitrary surface S.

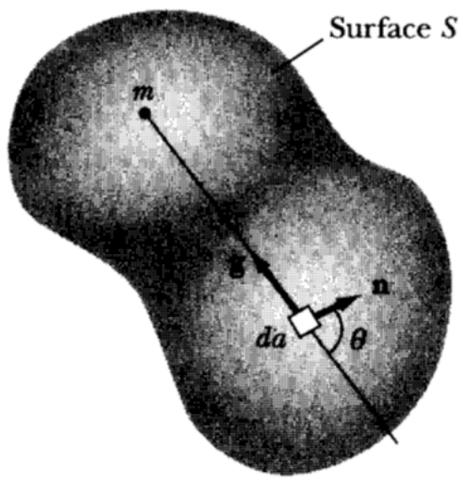


FIGURE 5-7 An arbitrary surface with a mass *m* placed inside. The unit vector **n** is normal to the surface at the differential area *da*.

$$\Phi_m = \int_S \mathbf{n} \cdot \mathbf{g}\, da \tag{5.32}$$

where the integral is over the surface S and the unit vector **n** is normal to the surface at the differential area da. If we substitute **g** from Equation 5.3 for the gravitational field vector for a body of mass m, we have for the scalar product $\mathbf{n} \cdot \mathbf{g}$,

$$\mathbf{n} \cdot \mathbf{g} = -Gm \frac{\cos \theta}{r^2}$$

where θ is the angle between **n** and **g**. We substitute this into Equation 5.32 and obtain

$$\Phi_m = -Gm \int_S \frac{\cos \theta}{r^2} da$$

The integral is over the solid angle of the arbitrary surface and has the value 4π steradians, which gives for the mass flux

$$\Phi_m = \int_S \mathbf{n} \cdot \mathbf{g} \, da = -4\pi Gm \tag{5.33}$$

Note that it is immaterial where the mass is located inside the surface S. We can generalize this result for many masses m_i inside the surface S by summing over the masses.

$$\int_S \mathbf{n} \cdot \mathbf{g} \, da = -4\pi G \sum_i m_i \tag{5.34}$$

If we change to a continuous mass distribution within surface S, we have

$$\int_S \mathbf{n} \cdot \mathbf{g} \, da = -4\pi G \int_V \rho \, dv \tag{5.35}$$

where the integral on the right-hand side is over the volume V enclosed by S, ρ is the mass density, and dv is the differential volume. We use Gauss's divergence theorem to rewrite this result. Gauss's divergence theorem, Equation 1.130 where d**a** = **n** da, is

$$\int_S \mathbf{n}\cdot\mathbf{g}\,da = \int_V \nabla\cdot\mathbf{g}\,dv \tag{5.36}$$

If we set the right-hand sides of Equations 5.35 and 5.36 equal, we have

$$\int_V (-4\pi G)\rho\,dv = \int_V \nabla\cdot\mathbf{g}\,dv$$

and because the surface S, and its volume V, is completely arbitrary, the two integrands must be equal.

$$\nabla\cdot\mathbf{g} = -4\pi G\rho \tag{5.37}$$

This result is similar to the differential form of Gauss's law for electric field, $\nabla\cdot\mathbf{E} = \rho/\varepsilon$, where ρ in this case is the charge density.

We insert $\mathbf{g} = -\nabla\Phi$ from Equation 5.5 into the left-hand side of Equation 5.37 and obtain

$$\nabla\cdot\mathbf{g} = -\nabla\cdot\nabla\Phi = -\nabla^2\Phi$$

Equation 5.37 becomes

$$\nabla^2\Phi = 4\pi G\rho$$

which is known as *Poisson's equation* and is useful in a number of potential theory applications. When the right-hand side of Equation 5.38 is zero, the result $\nabla^2\Phi = 0$ is an even better known equation called *Laplace's equation*. Poisson's equation is useful in developing Green's functions, whereas we often encounter Laplace's equation when dealing with various coordinate systems.

5.3 Lines of Force and Equipotential Surfaces

Let us consider a mass that gives rise to a gravitational field that can be described by a field vector **g**. Let us draw a line outward from the surface of the mass such that the direction of the line at every point is the same as the direction of **g** at that point. This line will extend from the surface of the mass to infinity. Such a line is called a **line of force**.

By drawing similar lines from every small increment of surface area of the mass, we can indicate the direction of the force field at any arbitrary point in space. The lines of force for a single point mass are all straight lines extending from the mass to infinity. Defined in this way, the lines of force are related only to the *direction* of the force field at any point. We may consider, however, that the *density* of such lines - that is, the number of lines passing through a unit area oriented perpendicular to the lines—is proportional to the *magnitude* of the force at that area. The lines-of-force picture is thus a convenient way to visualize both the magnitude and the direction (i.e., the *vector* property) of the field.

The potential function is defined at every point in space (except at the position of a point mass). Therefore, the equation

$$\Phi = \Phi(x_1, x_2, x_3) = \text{constant} \tag{5.39}$$

defines a surface on which the potential is constant. Such a surface is called an equipotential surface. The field vector **g** is equal to the gradient of Φ, so **g** can have no component along an equipotential surface. It therefore follows that every line of force must be normal to every equipotential surface. Thus, the field does no work on a body moving along an equipotential surface. Because the potential function is single valued, no two equipotential surfaces can intersect or touch. The surfaces of equal potential that surround a single, isolated point mass (or any spherically symmetric mass) are all spheres.

Consider two point masses M that are separated by a certain distance. If r_1 is the distance from one mass to some point in space and if r_2 is the distance from the other mass to the same point, then

$$\Phi = -GM\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = \text{constant} \tag{5.40}$$

defines the equipotential surfaces. Several of these surfaces are shown in Figure 5-8 for this two-particle system. In three dimensions, the surfaces are generated by rotating this diagram around the line connecting the two masses.

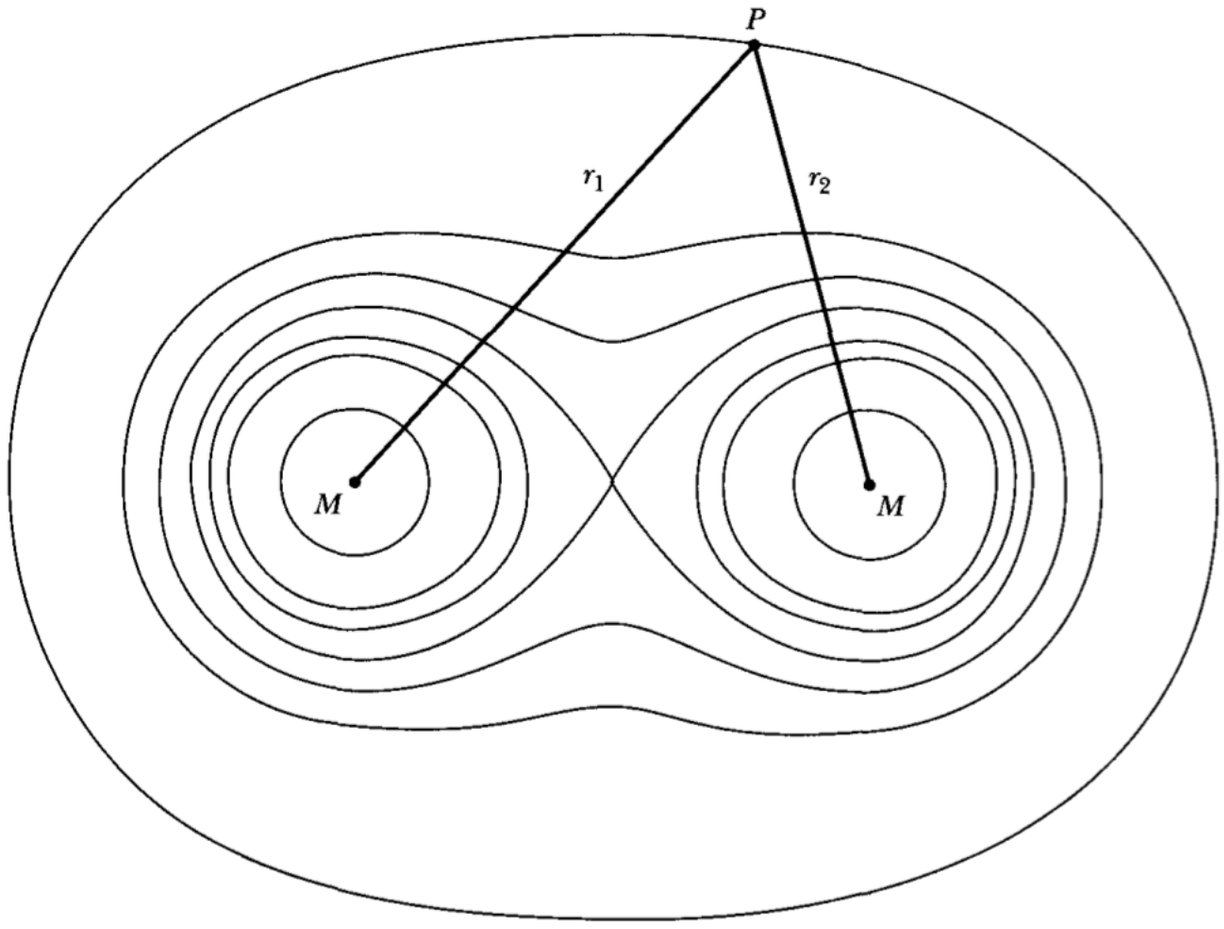


FIGURE 5-8 The equipotential surfaces due to two point masses *M*.

5.4 When Is the Potential Concept Useful?

The use of potentials to describe the effects of "action-at-a-distance" forces is an extremely important and powerful technique. We should not, however, lose sight of the fact that the ultimate justification for using a potential is to provide a convenient means of calculating the force on a body (or the energy for the body in the field) - for it is the force (and energy) and not the potential that is the physically meaningful quantity.

Thus, in some problems, it may be easier to calculate the force directly, rather than computing a potential and then taking the gradient. The advantage of using the potential method is that the potential is a scalar quantity: We need not deal with the added complication of sorting out the components of a vector until the gradient operation is performed. In direct calculations of the force, the components must be carried through the entire computation. Some skill, then, is necessary in choosing the particular approach to use. For example, if a problem has a particular symmetry that, from physical considerations, allows us to determine that the force has a certain direction, then the choice of that direction as one of the coordinate directions reduces the vector calculation to a simple scalar calculation. In such a case, the direct calculation of the force may be sufficiently straightforward to obviate the necessity of using the potential method. Every problem requiring a force must be examined to discover the easiest method of computation.

EXAMPLE 5.4

Consider a thin uniform disk of mass M and radius a. Find the force on a mass m located along the axis of the disk.

Solution. We solve this problem by using both the potential and direct force approaches. Consider Figure 5.9.

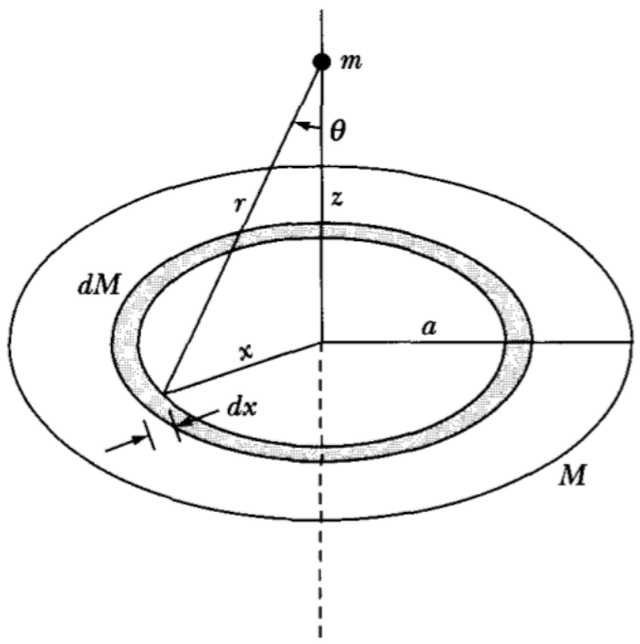


FIGURE 5-9 Example 5.4. We use the geometry shown here to find the gravitational force on a point mass m due to a thin uniform disk of mass M.

The differential potential $d\Phi$ at a distance z is given by

$$d\Phi = -G\frac{dM}{r} \tag{5.41}$$

The differential mass dM is a thin ring of width dx, because we have azimuthal symmetry.

$$dM = \rho dA = \rho 2\pi x\, dx \tag{5.42}$$

$$\begin{aligned} d\Phi &= -2\pi\rho G\frac{x\,dx}{r} = -2\pi\rho G\frac{x\,dx}{(x^2+z^2)^{1/2}} \\ \Phi(z) &= -\pi\rho G\int_0^a \frac{2x\,dx}{(x^2+z^2)^{1/2}} = -2\pi\rho G(x^2+z^2)^{1/2}\Big|_0^a \\ &= -2\pi\rho G[(a^2+z^2)^{1/2} - z] \end{aligned} \tag{5.43}$$

We find the force from

$$\mathbf{F} = -\nabla U = -m\nabla\Phi \tag{5.44}$$

From symmetry, we have only a force in the z direction

$$F_z = -m\frac{\partial\Phi(z)}{\partial z} = +2\pi m\rho G\left[\frac{z}{(a^2+z^2)^{1/2}} - 1\right] \tag{5.45}$$

In our second method, we compute the force directly using Equation 4.2:

$$d\mathbf{F} = -Gm\frac{dM'}{r^2}\mathbf{e}_r \tag{5.46}$$

where dM refers to the mass of a small differential area more like a square than a thin ring.

The vectors complicate matters. How can symmetry help? For every small dM′on one side of the thin ring of width dx, another dM′exists on the other side that exactly cancels the horizontal component of d**F** on m. Similarly, all horizontal components cancel, and we need only consider the vertical component of d**F** along z.

$$dF_z = \cos\theta\,|d\mathbf{F}| = -mG\frac{\cos\theta\ dM'}{r^2}$$

and, because cos θ = z/r,

$$dF_z = -mG\frac{z\ dM'}{r^3}$$

Now we integrate over the mass dM′= $\rho 2\pi$x dx around the ring and obtain

$$dF_z = -mG\rho\frac{2\pi xz\ dx}{r^3}$$

and

$$\begin{aligned} F_z &= -\pi m\rho Gz\int_0^a \frac{2x\,dx}{(z^2+x^2)^{3/2}} \\ &= -\pi m\rho Gz\left[\frac{-2}{(z^2+x^2)^{1/2}}\right]\Bigg|_0^a \\ &= 2\pi m\rho G\left[\frac{z}{(a^2+z^2)^{1/2}} - 1\right] \end{aligned} \tag{5.47}$$

which is identical to Equation 5.45. Notice that the value of F_z is negative, indicating that the force is downward in Figure 5-9 and attractive.

5.5 Ocean Tides

The ocean tides have long been of interest to humans. Galileo tried unsuccessfully to explain ocean tides but could not account for the timing of the approximately two high tides each day. Newton finally gave an adequate explanation. The tides are caused by the gravitational attraction of the ocean to both the Moon and the Sun, but there are several complicating factors.

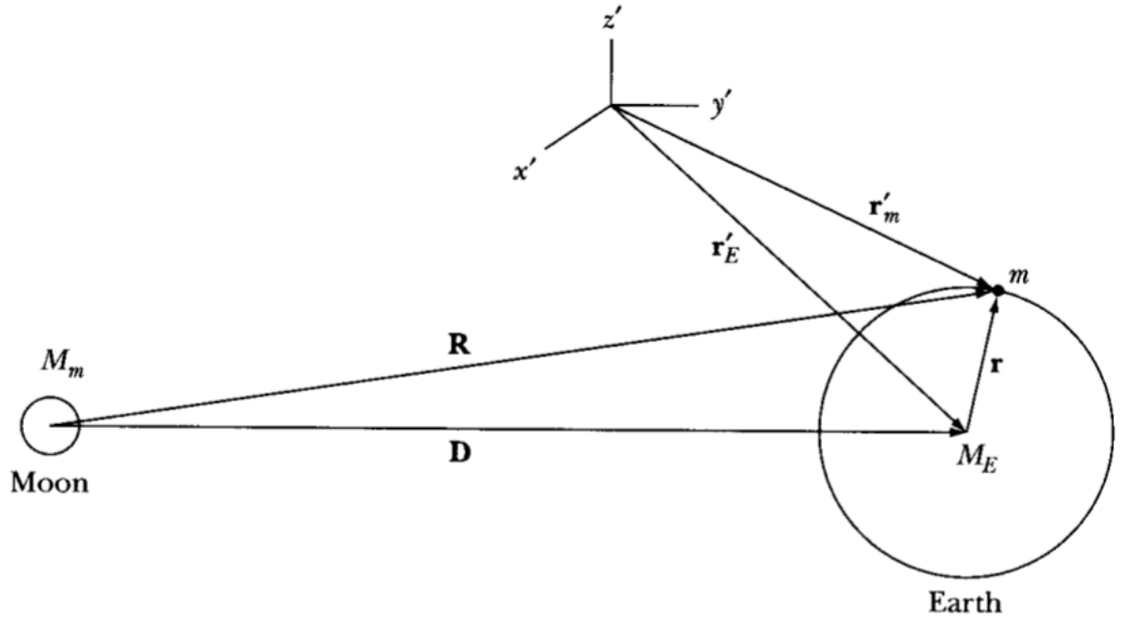
The calculation is complicated by the fact that the surface of Earth is not an inertial system. Earth and Moon rotate about their center of mass (and move about the Sun), so we may regard the water nearest the Moon as being pulled away from Earth, and Earth as being pulled away from the water farthest from the Moon. However, Earth rotates while the Moon rotates about Earth. Let's first consider only the effect of the Moon, adding the effect of the Sun later. We will assume a simple model whereby Earth's surface is completely covered with water, and we shall add the effect of Earth's rotation at an appropriate time. We set up an inertial frame of reference x'y'z' as shown in Figure 5.10a.

We let M_m be the mass of the Moon, r the radius of a circular Earth, and D the distance from the center of the Moon to the center of Earth. We consider the effect of both the Moon's and Earth's gravitational attraction on a small mass m placed on the surface of Earth. As displayed in Figure 5-10a, the position vector of the mass m from the Moon is **R**, from the center of Earth is **r**, and from our inertial system $\mathbf{r}'_m$. The position vector from the inertial system to the center of Earth is $\mathbf{r}'_E$. As measured from the inertial system, the force on m, due to the earth and the Moon, is

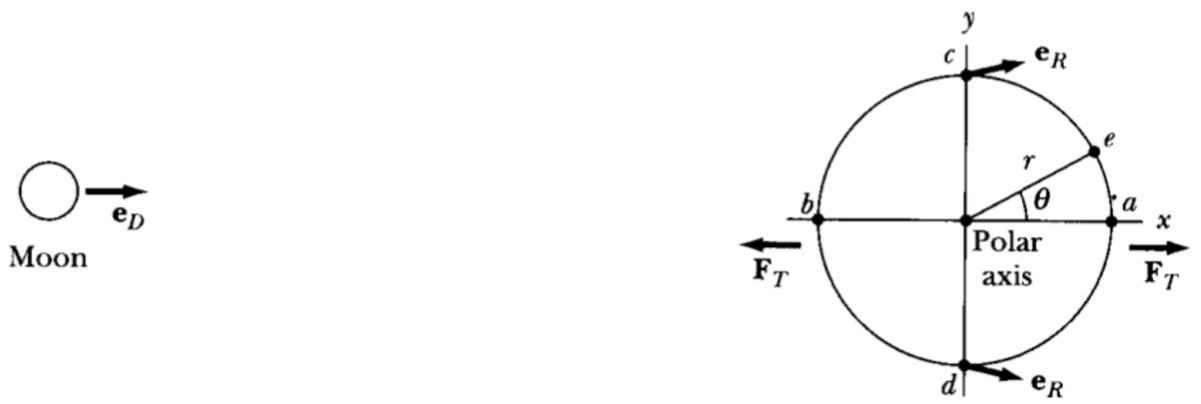
$$m\ddot{\mathbf{r}}'_m = -\frac{GmM_E}{r^2}\mathbf{e}_r - \frac{GmM_m}{R^2}\mathbf{e}_R \tag{5.48}$$

Similarly, the force on the center of mass of Earth caused by the Moon is

$$M_E\ddot{\mathbf{r}}'_E = -\frac{GM_EM_m}{D^2}\mathbf{e}_D \tag{5.49}$$



(a)



(b)

FIGURE 5-10 (a) Geometry to find ocean tides on Earth due to the Moon. (b) Polar view with the polar axis along the z-axis.

We want to find the acceleration $\ddot{r}$ as measured in the noninertial system placed at the center of Earth. Therefore, we want

$$\begin{aligned}\ddot{\mathbf{r}} &= \ddot{\mathbf{r}}'_m - \ddot{\mathbf{r}}'_E = \frac{m\ddot{\mathbf{r}}'_m}{m} - \frac{M_E\ddot{\mathbf{r}}'_E}{M_E} \\ &= -\frac{GM_E}{r^2}\mathbf{e}_r - \frac{GM_m}{R^2}\mathbf{e}_R + \frac{GM_m}{D^2}\mathbf{e}_D \\ &= -\frac{GM_E}{r^2}\mathbf{e}_r - GM_m\left(\frac{\mathbf{e}_R}{R^2} - \frac{\mathbf{e}_D}{D^2}\right)\end{aligned} \tag{5.50}$$

The first part is due to Earth, and the second part is the acceleration from the **tidal** force, which is responsible for producing the ocean tides. It is due to the difference between the Moon's gravitational pull at the center of Earth and on Earth's surface.

We next find the effect of the tidal force at various points on Earth as noted in Figure 5-10b. We show a polar view of Earth with the polar axis along the z-axis. The tidal force $\mathbf{F}_T$ on the mass m on Earth's surface is

$$\mathbf{F}_T = -GmM_m\left(\frac{\mathbf{e}_R}{R^2} - \frac{\mathbf{e}_D}{D^2}\right) \tag{5.51}$$

where we have used only the second part of Equation 5.50. We look first at point a, the farthest point on Earth from the Moon. Both unit vectors $\mathbf{e}_R$ and $\mathbf{e}_D$ are pointing in the same direction away from the Moon along the x-axis. Because R > D, the second term in Equation 5.51 predominates, and the tidal force is along the +x-axis as shown in Figure 5-l0b. For point b, R < D and the tidal force has approximately the same magnitude as at point a because R/D << 1, but is along the -x-axis. The magnitude of the tidal force along the x-axis, F_{Tx}, is

$$F_{Tx} = -GmM_m\left(\frac{1}{R^2} - \frac{1}{D^2}\right) = \blacktriangledown GmM_m\left(\frac{1}{(D+r)^2} - \frac{1}{D^2}\right)$$

$$= -\frac{GmM_m}{D^2}\left(\frac{1}{\left(1+\frac{r}{D}\right)^2} - 1\right)$$

We expand the first term in brackets using the $(1 + x)^{-2}$ expansion in Equation D.9.

$$F_{Tx} = -\frac{GmM_m}{D^2}\left[1 - 2\frac{r}{D} + 3\left(\frac{r}{D}\right)^2 - \cdots - 1\right] = +\frac{2GmM_m r}{D^3} \tag{5.52}$$

where we have kept only the largest nonzero term in the expansion, because r/D = 0.02.

For point c, the unit vector $\mathbf{e}_R$ (Figure 5-l0b) is not quite exactly along $\mathbf{e}_D$, but the x-axis components approximately cancel, because $R \simeq D$ and the x-components of $\mathbf{e}_R$ and $\mathbf{e}_D$ are similar, There will be a small component of $\mathbf{e}_R$ along the y-axis. We approximate the y-component of $\mathbf{e}_R$ by (r/D))**j**, and the tidal force at point c, call it $\mathbf{F}_{Ty}$, is along the y-axis and has the magnitude

$$F_{Ty} = -GmM_m\left(\frac{1}{D^2}\frac{r}{D}\right) = -\frac{GmM_m r}{D^3} \tag{5.53}$$

Note that this force is along the -y-axis toward the center of Earth at point c. We find similarly at point D the same magnitude, but the component of $\mathbf{e}_R$ will be along the -y-axis, so the force itself, with the sign of Equation 5.53, will be along the +y-axis toward the center of Earth. We indicate the tidal forces at points a, b, c, and don Figure 5-11 a.

We determine the force at an arbitrary point e by noting that the x- and y-components of the tidal force can be found by substituting x and y for r in F_{Tx} and F_{Ty}, respectively, in Equations 5.52 and 5.53.

$$F_{Tx} = \frac{2GmM_m x}{D^3}$$

$$F_{Ty} = -\frac{GmM_m y}{D^3}$$

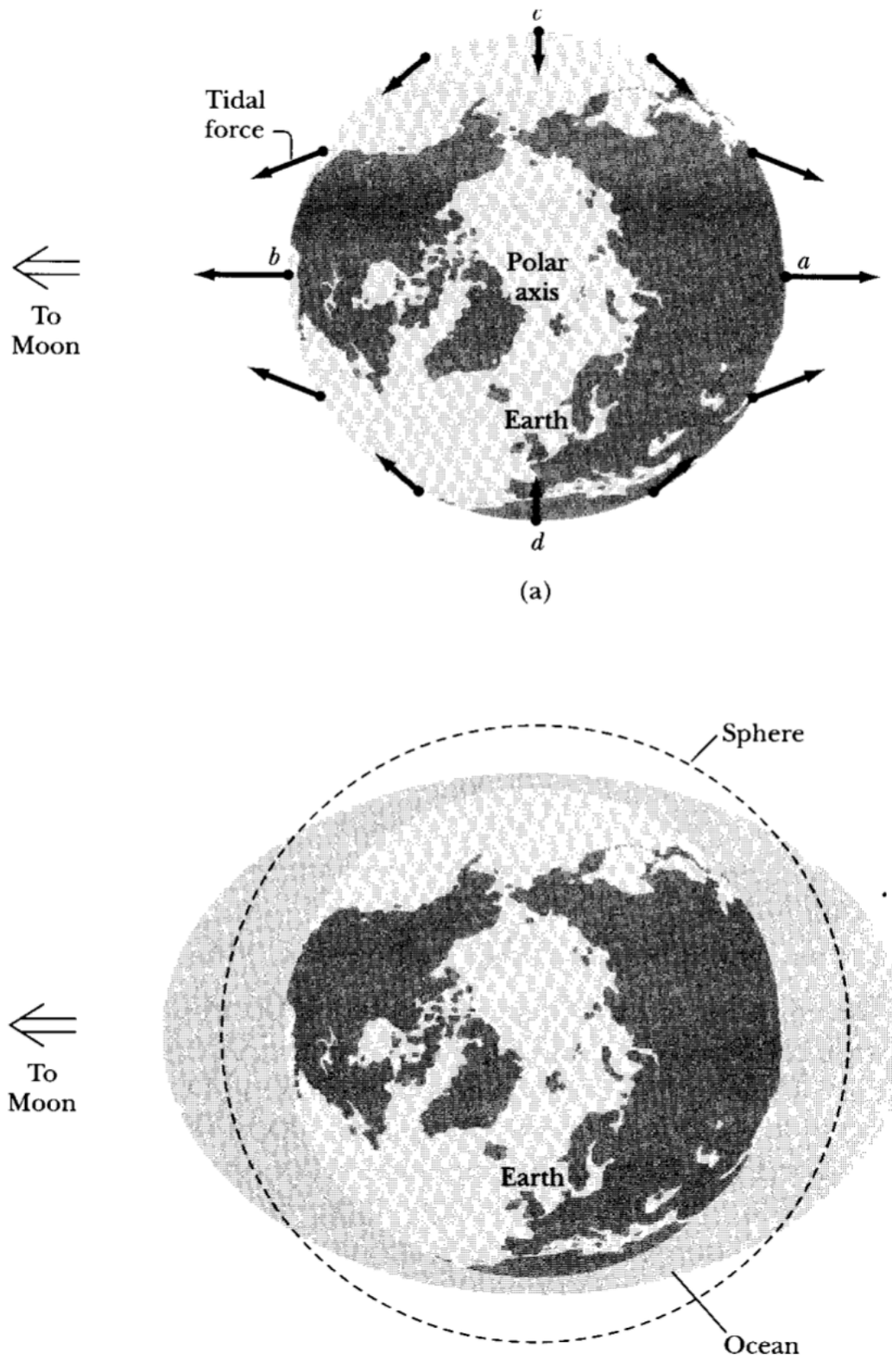


FIGURE 5-11 (a) The tidal forces are shown at various places on Earth's surface including the points *a*, *b*, *c*, and *d* of Figure 5-10. (b) An exaggerated view of Earth's ocean tides.

Then at an arbitrary point such as e, we let x = rcos θ and y = r sin θ , so we have

$$F_{Tx} = \frac{2GmM_m r\cos\theta}{D^3} \tag{5.54a}$$

$$F_{Ty} = -\frac{GmM_m r\sin\theta}{D^3} \tag{5.54b}$$

Equations 5.54a and b give the tidal force around Earth for all angles θ . Note that they give the correct result at points a, b, c, and d.

Figure 5-11a gives a representation of the tidal forces. For our simple model, these forces lead to the water along the y-axis being more shallow than along the x-axis. We show an exaggerated result in Figure 5-11b. As Earth makes a revolution about its own axis every 24 hours, we will observe two high tides a day.

A quick calculation shows that the Sun's gravitational attraction is about 175 times stronger than the Moon's on Earth's surface, so we would expect tidal forces from the Sun as well. The tidal force calculation is similar to the one we have just performed for the Moon. The result is that the tidal force due to the Sun is 0.46 that of the Moon, a sizable effect. Despite the stronger attraction due to the Sun, the gravitational force gradient over the surface of Earth is much smaller, because of the much larger distance to the Sun

EXAMPLE 5.5

Calculate the maximum height change in the ocean tides caused by the Moon.

Solution. We continue to use our simple model of the ocean surrounding Earth. Newton proposed a solution to this calculation by imagining that two wells be dug, one along the direction of high tide (our x-axis) and one along the direction of low tide (our y-axis). If the tidal height change we want to determine is h, then the difference in potential energy of mass m due to the height difference is mgh.

Let's calculate the difference in work if we move the mass m from point c in Figure 5-12 to the center of Earth and then to point a.

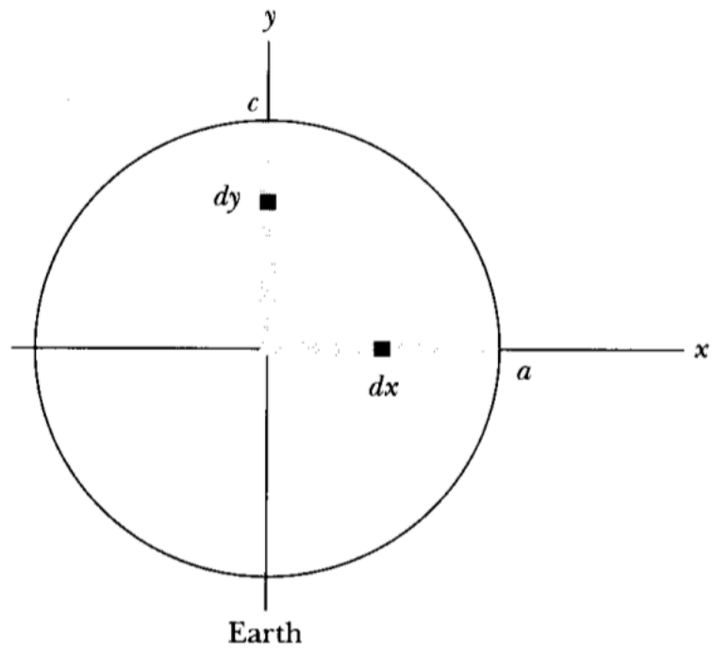


FIGURE 5-12 Example 5.5. We calculate the work done to move a point mass m from point c to the center of Earth and then to point a.

This work W done by gravity must equal the potential energy change mgh. The work W is

$$W = \int_{r+\delta_1}^{0} F_{Ty}\,dy + \int_{0}^{r+\delta_2} F_{Tx}\,dx$$

where we use the tidal forces F_{Ty} and F_{Tx} of Equations 5.54. The small distances δ_1 and δ_2 are to account for the small variations from a spherical Earth, but these values are so small they can be henceforth neglected. The value for W becomes

$$W = \frac{GmM_m}{D^3}\left[\int_r^0 (-y)dy + \int_0^r 2x\,dx\right]$$
$$= \frac{GmM_m}{D^3}\left(\frac{r^2}{2} + r^2\right) = \frac{3GmM_m r^2}{2D^3}$$

Because this work is equal to mgh, we have

$$mgh = \frac{3GmM_m r^2}{2D^3}$$

$$h = \frac{3GM_m r^2}{2gD^3} \tag{5.55}$$

Note that the mass m cancels, and the value of h does not depend on m. Nor does it depend on the substance, so to the extent Earth is plastic, similar tidal effects should be (and are) observed for the surface land. If we insert the known values of the constants into Equation 5.55, we find

$$h = \frac{3(6.67 \times 10^{-11}\,\mathrm{m^3/kg \cdot s^2})(7.350 \times 10^{22}\,\mathrm{kg})(6.37 \times 10^{6}\,\mathrm{m})^2}{2(9.80\,\mathrm{m/s^2})(3.84 \times 10^{8}\,\mathrm{m})^3} = 0.54\ \mathrm{m}$$

The highest tides (called *spring* tides) occur when Earth, the Moon, and the Sun are lined up (new moon and full moon), and the smallest tides (called *neap* tides) occur for the first and third quarters of the Moon when the Sun and Moon are at right angles to each other, partially cancelling their effects. The maximum tide, which occurs every 2 weeks, should be 1.46h = 0.83 m for the spring tides.

An observer who has spent much time near the ocean has noticed that typical ocean shore tides are greater than those calculated in Example 5.5. Several other effects come into play. Earth is not covered completely with water, and the continents play a significant role, especially the shelfs and narrow estuaries. Local effects can be dramatic, leading to tidal changes of several meters. The tides in midocean, however, are similar to what we have calculated. Resonances can affect the natural oscillation of the bodies of water and cause tidal changes.

Tidal friction between water and Earth leads to a significant amount of energy loss on Earth. Earth is not rigid, and it is also distorted by tidal forces.

In addition to the effects just discussed, remember that as Earth rotates, the Moon is also orbiting Earth. This leads to the result that there are not quite exactly two high tides per day, because they occur once every 12 h and 26 min. The plane of the moon's orbit about Earth is also not perpendicular to Earth's rotation axis. This causes one high tide each day to be slightly higher than the other. The tidal friction between water and land mentioned previously also results in Earth "dragging" the ocean with it as Earth rotates. This causes the high tides to be not quite along the Earth-Moon axis, but rather several degrees apart as shown in Figure 5-13.

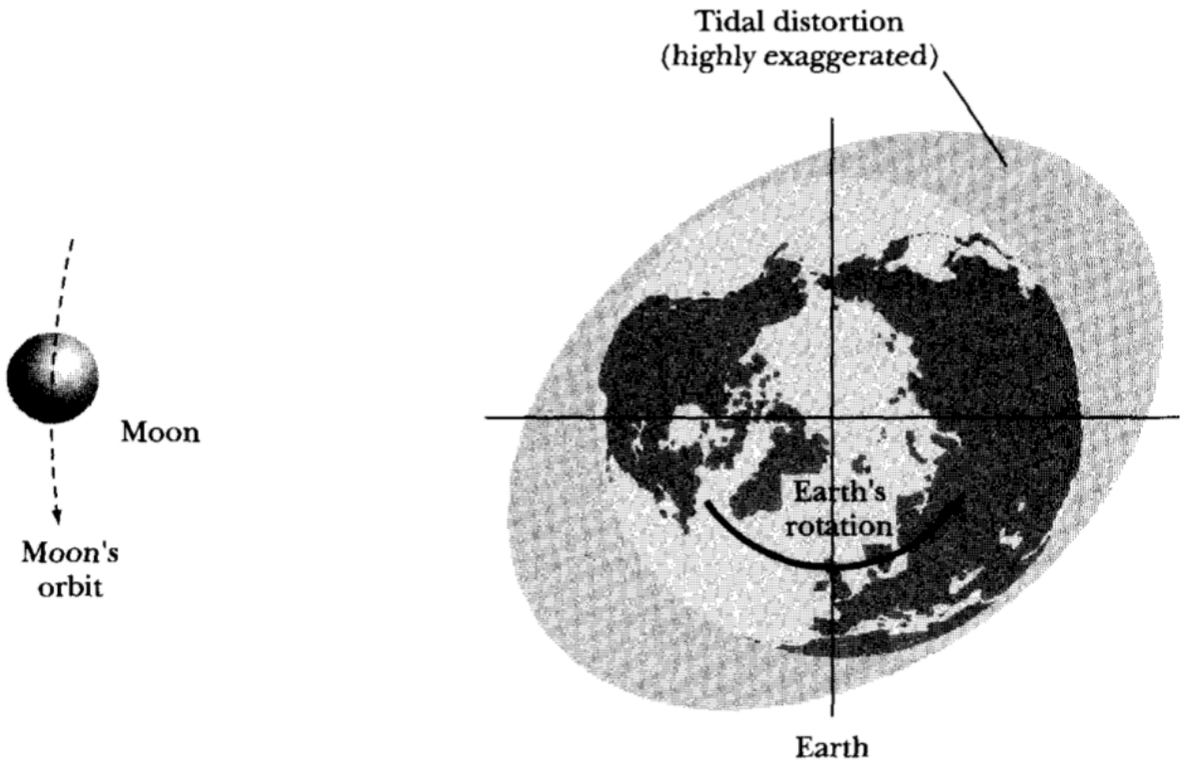


FIGURE 5-13 Some effects cause the high tides to not be exactly along the Earth-Moon axis.

CHAPTER 6 - Some Methods in the Calculus of Variations

6.1 Introduction

Many problems in Newtonian mechanics are more easily analyzed by means of alternative statements of the laws, including Lagrange's equation and Hamilton's principle. As a prelude to these techniques, we consider in this chapter some general principles of the techniques of the calculus of variations.

Emphasis will be placed on those aspects of the theory of variations that have a direct bearing on classical systems, omitting some existence proofs. Our primary interest here is in determining the path that gives extremum solutions, for example, the shortest distance (or time) between two points. A well-known example of the use of the theory of variations is Fermat's principle: Light travels by the path that takes the least amount of time.

6.2 Statement of the Problem

The basic problem of the calculus of variations is to determine the function y(x) such that the integral

$$J = \int_{x_1}^{x_2} f\{y(x), y'(x); x\}\, dx \tag{6.1}$$

is an *extremum* (i.e., either a maximum or a minimum).

In Equation 6.1, y'(x) = dy/dx, and the semicolon in/separates the independent variable x from the dependent variable y(x) and its derivative y'(x). The functional J depends on the function y(x), and the limits of integration are fixed. The function y(x) is then to be varied until an extreme value of J is found. By this we mean that if a function y = y(x) gives the integral J a minimum value, then any neighboring function, no matter how close to y(x), must make J increase. The definition of a neighboring function may be made as follows. We give all possible functions y a parametric representation y = y(α , x) such that, for $\alpha = 0$, y = y(0, x) = y(x) is the function that yields an extremum for/. We can then write

$$y(\alpha, x) = y(0, x) + \alpha\eta(x) \tag{6.2}$$

where η(x) is some function of x that has a continuous first derivative and that vanishes at x_1 and x_2, because the varied function y(α , x) must be identical with y(x) at the endpoints of the path:

$$\eta(x_1) = \eta(x_2) = 0$$

The situation is depicted schematically in Figure 6-1.

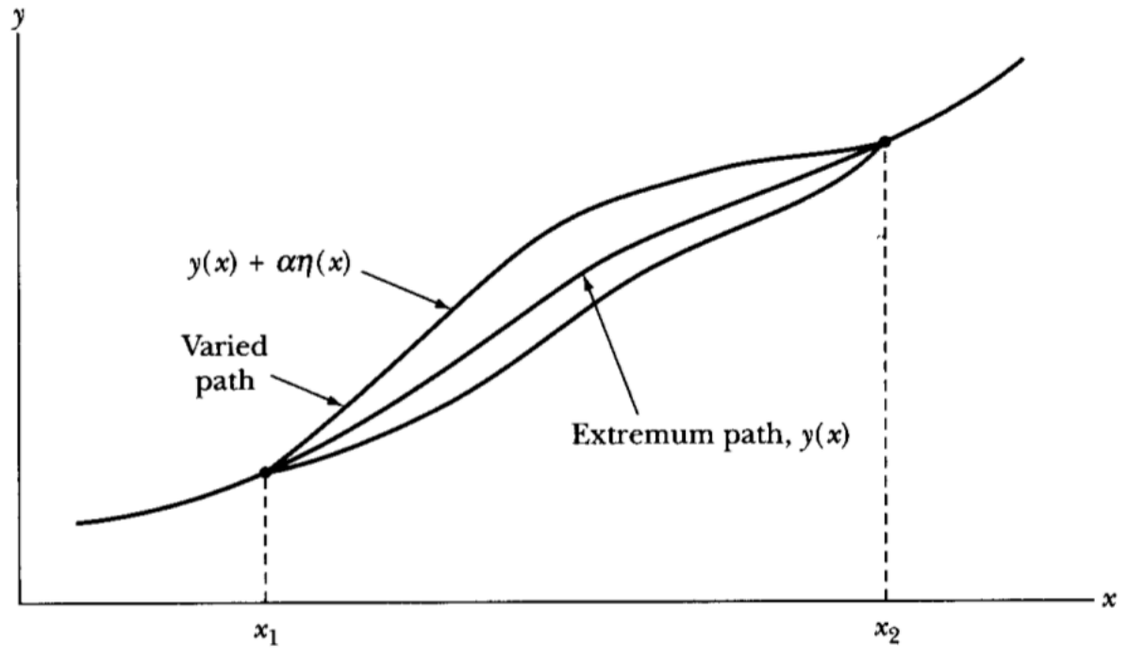


FIGURE 6-1 The function $y(x)$ is the path that makes the functional J an extremum. The neighboring functions $y(x) + \alpha\eta(x)$ vanish at the endpoints and may be close to $y(x)$, but are not the extremum.

If functions of the type given by Equation 6.2 are considered, the integral J becomes a functional of the parameter α :

$$J(\alpha) = \int_{x_1}^{x_2} f\{y(\alpha, x), y'(\alpha, x); x\}\, dx \tag{6.3}$$

The condition that the integral have a stationary value (i.e., that an extremum results) is that J be independent of α in first order along the path giving the extremum ($\alpha = 0$), or, equivalently, that

$$\left.\frac{\partial J}{\partial \alpha}\right|_{\alpha=0} = 0 \tag{6.4}$$

for all functions $\eta(x)$. This is only a necessary condition; it is not sufficient.

EXAMPLE 6.1

Consider the function f = $(dy/dx)^2$, where y(x) = x. Add to y(x) the function η(x) = sin x, and find J(α) between the limits of x = 0 and x = 2 π . Show that the stationary value of J(α) occurs for $\alpha = 0$.

Solution. We may construct neighboring varied paths by adding to y(x),

$$y(x) = x \tag{6.5}$$

the sinusoidal variation α sin x.

$$y(\alpha, x) = x + \alpha \sin x \tag{6.6}$$

These paths are illustrated in Figure 6-2 for $\alpha = 0$ and for two different nonvanishing values of α .

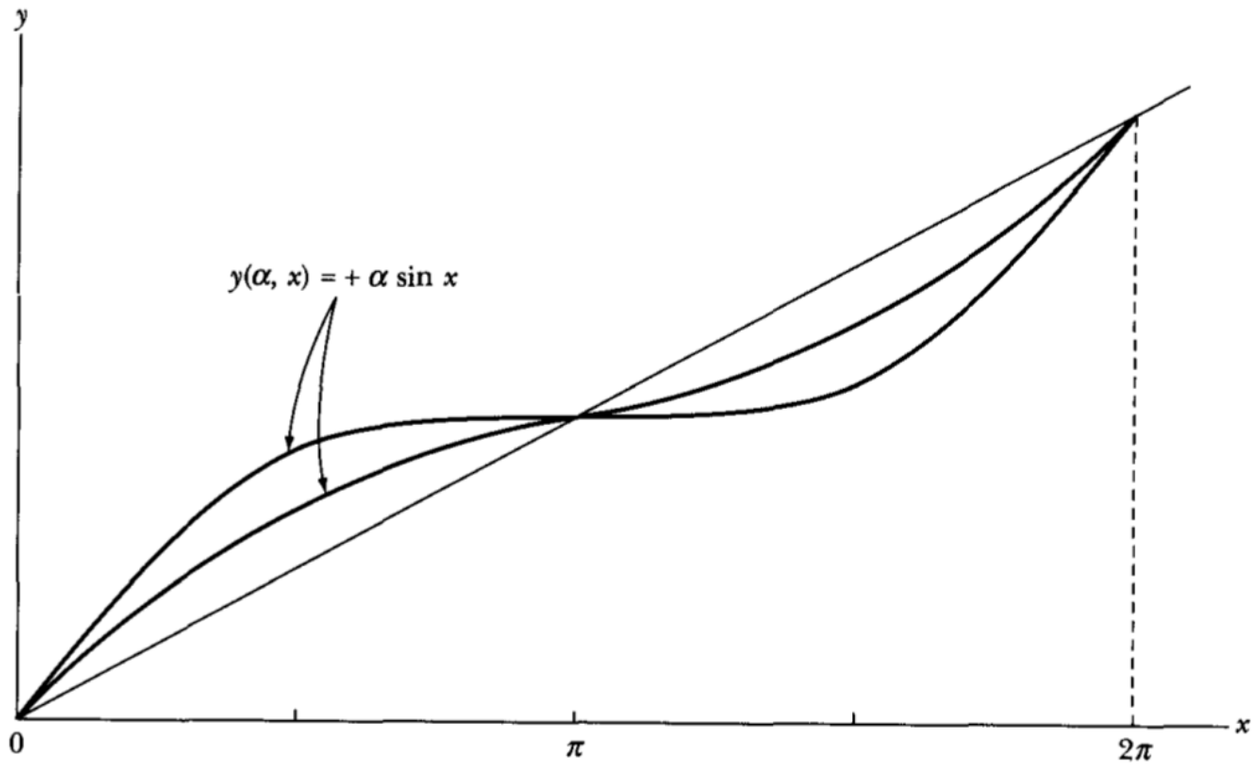


FIGURE 6-2 Example 6.1. The various paths $y(\alpha, x) = x + \alpha \sin x$. The extremum path occurs for $\alpha = 0$.

Clearly, the function η (x) = sin x obeys the endpoint conditions, that is, $\eta(0) = 0 = \eta(2\pi)$. To determine f(y, y'; x) we first determine.

$$\frac{dy(\alpha, x)}{dx} = 1 + \alpha \cos x \tag{6.7}$$

then

$$f = \left(\frac{dy(\alpha, x)}{dx}\right)^2 = 1 + 2\alpha \cos x + \alpha^2 \cos^2 x \tag{6.8}$$

Equation 6.3 now becomes

$$J(\alpha) = \int_0^{2\pi} (1 + 2\alpha \cos x + \alpha^2 \cos^2 x)\, dx \tag{6.9}$$

$$= 2\pi + \alpha^2 \pi \tag{6.10}$$

Thus we see the value of J(α) is always greater than J(0), no matter what value (positive or negative) we choose for α . The condition of Equation 6.4 is also satisfied.

6.3 Euler's Equation

To determine the result of the condition expressed by Equation 6.4, we perform the indicated differentiation in Equation 6.3:

$$\frac{\partial J}{\partial \alpha} = \frac{\partial}{\partial \alpha} \int_{x_1}^{x_2} f\{y, y'; x\}\, dx \tag{6.11}$$

Because the limits of integration are fixed, the differential operation affects only the integrand. Hence,

$$\frac{\partial J}{\partial \alpha} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y}\frac{\partial y}{\partial \alpha} + \frac{\partial f}{\partial y'}\frac{\partial y'}{\partial \alpha}\right) dx \tag{6.12}$$

From Equation 6.2, we have

$$\frac{\partial y}{\partial \alpha} = \eta(x); \qquad \frac{\partial y'}{\partial \alpha} = \frac{d\eta}{dx} \tag{6.13}$$

Equation 6.12 becomes

$$\frac{\partial J}{\partial \alpha} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} \eta(x) + \frac{\partial f}{\partial y'} \frac{d\eta}{dx} \right) dx \tag{6.14}$$

The second term in the integrand can be integrated by parts:

$$\int u\, dv = uv - \int v\, du \tag{6.15}$$

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial y'} \frac{d\eta}{dx}\, dx = \frac{\partial f}{\partial y'} \eta(x) \Big|_{x_1}^{x_2} - \int_{x_1}^{x_2} \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \eta(x)\, dx \tag{6.16}$$

The integrated term vanishes because $\eta(x_1) = \eta(x_2) = 0$. Therefore, Equation 6.12 becomes

$$\begin{aligned} \frac{\partial J}{\partial \alpha} &= \int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y} \eta(x) - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \eta(x) \right] dx \\ &= \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y'} \right) \eta(x)\, dx \end{aligned} \tag{6.17}$$

The integral in Equation 6.17 now appears to be independent of α. But the functions y and y' with respect to which the derivatives of f are taken are still functions of α. Because $(\partial J/\partial \alpha)|_{\alpha=0}$ must vanish for the extremum value and because η (x) is an arbitrary function (subject to the conditions already stated), the integrand in Equation 6.17 must itself vanish for $\alpha = 0$:

$$\boxed{\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} = 0} \qquad \textbf{Euler's equation} \tag{6.18}$$

where now y and y' are the original functions, independent of α . This result is known as **Euler's equation**, which is a necessary condition for J to have an extremum value.

EXAMPLE 6.2

We can use the calculus of variations to solve a classic problem in the history of physics: the *brachistochrone*. Consider a particle moving in a constant force field starting at rest from some point (x_1,y_1) to some lower point (x_2, y_2). Find the path that allows the particle to accomplish the transit in the least possible time.

Solution. The coordinate system may be chosen so that the point (x_1,y_1) is at the origin. Further, let the force field be directed along the positive x-axis as in Figure 6-3.

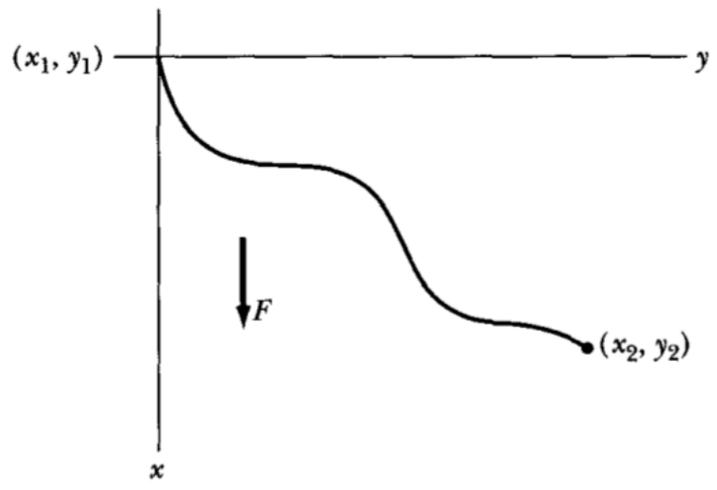


FIGURE 6-3 Example 6.2. The *brachistochrone* problem is to find the path of a particle moving from (x_1, y_1) to (x_2, y_2) that occurs in the least possible time. The force field acting on the particle is *F*, which is down and constant.

Because the force on the particle is constant - and if we ignore the possibility of friction - the field is conservative, and the total energy of the particle is T + U = constant.

If we measure the potential from the point x = 0 [i.e., U(x = 0) =0], then, because the particle starts from rest, T+U= 0. The kinetic energy is $T = \frac{1}{2}mv^2$, and the potential energy is U = -Fx = -mgx, where g is the acceleration imparted by the force. Thus

$$v = \sqrt{2gx} \tag{6.19}$$

The time required for the particle to make the transit from the origin to (x_2, y_2) is

$$\begin{aligned} t &= \int_{(x_1,y_1)}^{(x_2,y_2)} \frac{ds}{v} = \int \frac{(dx^2 + dy^2)^{1/2}}{(2gx)^{1/2}} \\ &= \int_{x_1=0}^{x_2} \left(\frac{1 + y'^2}{2gx}\right)^{1/2} dx \end{aligned} \tag{6.20}$$

The time of transit is the quantity for which a minimum is desired. Because the constant $(2g)^{-1/2}$ does not affect the final equation, the function f may be identified as

$$f = \left(\frac{1 + y'^2}{x}\right)^{1/2} \tag{6.21}$$

And, because $\partial f / \partial y = 0$, the Euler equation (Equation 6.18) becomes

$$\frac{d}{dx}\frac{\partial f}{\partial y'} = 0$$

or

$$\frac{\partial f}{\partial y'} = \text{constant} \equiv (2a)^{-1/2}$$

where a is a new constant.

Performing the differentiation $\partial f / \partial u'$ on Equation 6.21 and squaring the result, we have

$$\frac{y'^2}{x(1 + y'^2)} = \frac{1}{2a} \tag{6.22}$$

This may be put in the form

$$y = \int \frac{x\,dx}{(2ax - x^2)^{1/2}} \tag{6.23}$$

We now make the following change of variable:

$$\begin{aligned} x &= a(1 - \cos\theta) \\ dx &= a \sin\theta\, d\theta \end{aligned} \tag{6.24}$$

The integral in Equation 6.23 then becomes

$$y = \int a(1 - \cos\theta)\, d\theta$$

and

$$y = a(\theta - \sin\theta) + \text{constant} \tag{6.25}$$

The parametric equations for a cycloid passing through the origin are

$$\left.\begin{aligned} x &= a(1 - \cos\theta) \\ y &= a(\theta - \sin\theta) \end{aligned}\right\} \tag{6.26}$$

which is just the solution found, with the constant of integration set equal to zero to conform with the requirement that $(0, 0)$ is the starting point of the motion.

The path is then as shown in Figure 6-4, and the constant a must be adjusted to allow the cycloid to pass through the specified point (x_2, y_2).

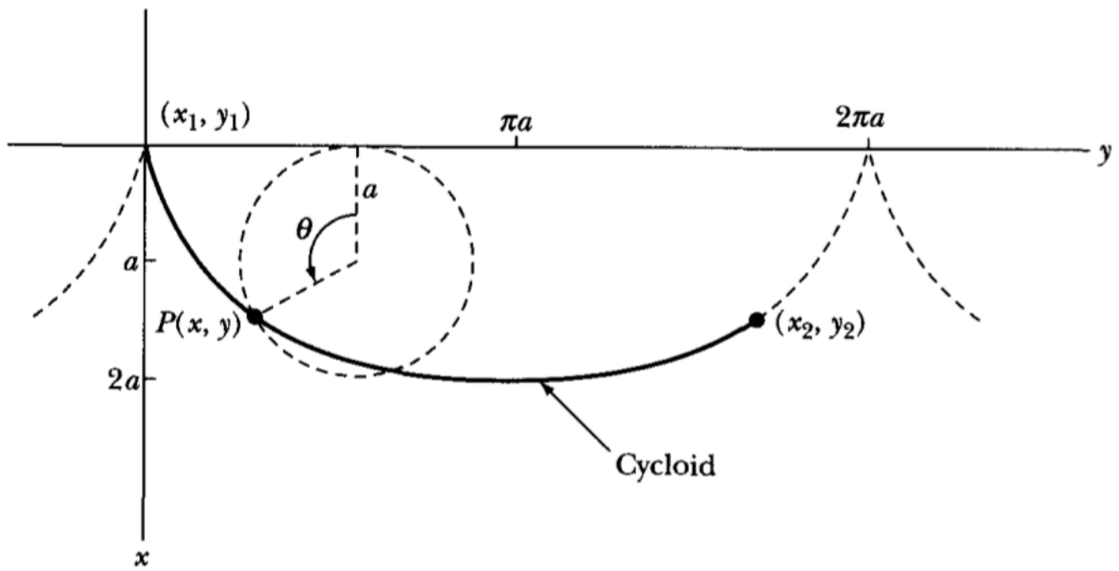


FIGURE 6-4 Example 6.2. The solution of the brachistochrone problem is a cycloid.

Solving the problem of the brachistochrone does indeed yield a path the particle traverses in a minimum time. But the procedures of variational calculus are designed only to produce an extremum - either a minimum or a maximum. It is almost always the case in dynamics that we desire (and find) a minimum for the problem.

EXAMPLE 6.3

Consider the surface generated by revolving a line connecting two fixed points (x_1, y_1) and (x_2, y_2) about an axis coplanar with the two points. Find the equation of the line connecting the points such that the surface area generated by the revolution (i.e., the area of the surface of revolution) is a minimum.

Solution. We assume that the curve passing through (x_1, y_1) and (x_2, y_2) is revolved about the y-axis, coplanar with the two points. To calculate the total area of the surface of revolution, we first find the area dA of a strip. Refer to Figure 6-5.

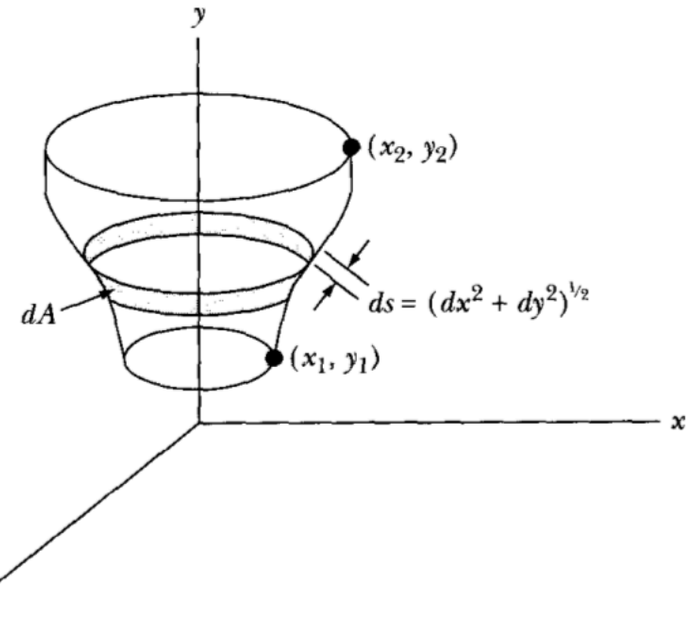


FIGURE 6-5 Example 6.3. The geometry of the problem and area dA are indicated to minimize the surface of revolution around the y-axis.

$$dA = 2\pi x\, ds = 2\pi x(dx^2 + dy^2)^{1/2} \tag{6.27}$$

$$A = 2\pi \int_{x_1}^{x_2} x(1 + y'^2)^{1/2}\, dx \tag{6.28}$$

where y' = dy/dx. To find the extremum value we let

$$f = x(1 + y'^2)^{1/2} \tag{6.29}$$

and insert into Equation 6.18:

$$\frac{\partial f}{\partial y} = 0$$

$$\frac{\partial f}{\partial y'} = \frac{xy'}{(1 + y'^2)^{1/2}}$$

therefore,

$$\frac{d}{dx}\left[\frac{xy'}{(1+y'^2)^{1/2}}\right] = 0$$

$$\frac{xy'}{(1+y'^2)^{1/2}} = \text{constant} \equiv a \tag{6.30}$$

From Equation 6.30, we determine

$$y' = \frac{a}{(x^2 - a^2)^{1/2}} \tag{6.31}$$

$$y = \int \frac{a\,dx}{(x^2 - a^2)^{1/2}} \tag{6.32}$$

The solution of this integration is

$$y = a \cos h^{-1}\left(\frac{x}{a}\right) + b \tag{6.33}$$

where a and b are constants of integration determined by requiring the curve to pass through the points (x_1, y_1) and (x_2, y_2). Equation 6.33 can also be written as

$$x = a \cos h\left(\frac{y - b}{a}\right) \tag{6.34}$$

which is more easily recognized as the equation of a catenary, the curve of a flexible cord hanging freely between two points of support.

Choose two points located at (x_1, y_1) and (x_2, y_2) joined by a curve y(x). We want to find y(x) such that if we revolve the curve around the x-axis, the surface area of the revolution is a minimum. This is the "soap film" problem, because a soap film suspended between two wire circular rings takes this shape (Figure 6-6).

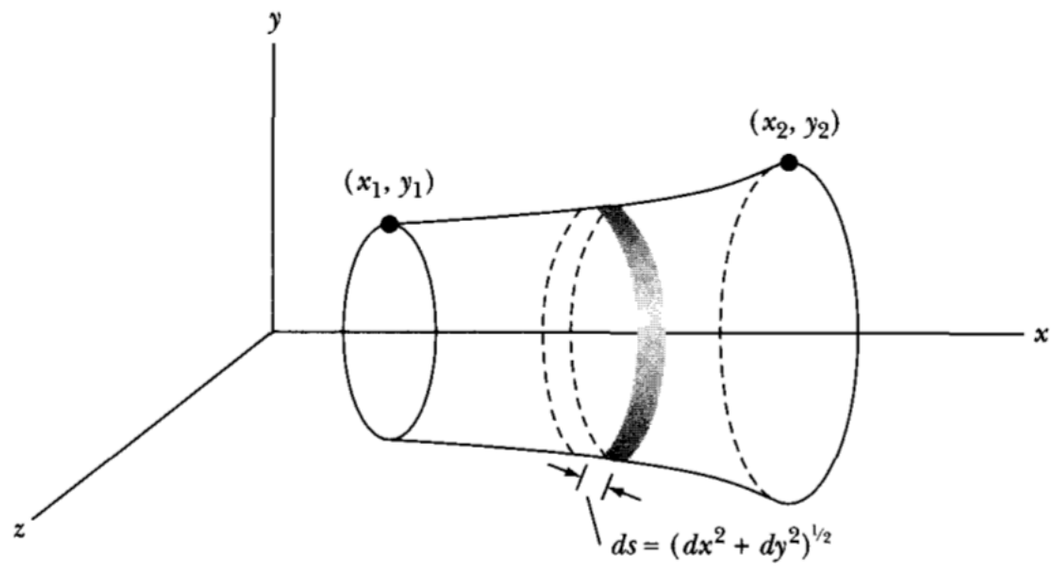


FIGURE 6-6 The "soap film" problem in which we want to minimize the surface area of revolution around the *x*-axis.

We want to minimize the integral of the area $dA = 2\pi y\, ds$ where $ds = y\sqrt{1 + y'^2}\, dx$ and $y' = dy/dx$.

$$A = 2\pi \int_{x_1}^{x_2} y\sqrt{1 + y'^2}\, dx \tag{6.35}$$

We find the extremum by setting $f = y\sqrt{1 + y'^2}$ and inserting into Equation 6.18. The derivatives we need are

$$\frac{\partial f}{\partial y} = \sqrt{1 + y'^2}$$

$$\frac{\partial f}{\partial y'} = \frac{yy'}{\sqrt{1 + y'^2}}$$

Equation 6.18 becomes

$$\sqrt{1 + y'^2} = \frac{d}{dx}\left[\frac{yy'}{\sqrt{1 + y'^2}}\right] \tag{6.36}$$

Equation 6.36 does not appear to be a simple equation to solve for y(x). Let's stop and think about whether there might be an easier method of solution. You may have noticed that this problem is just like Example 6.3, but in that case we were minimizing a surface of revolution about the y-axis rather than around the x-axis. The solution to the soap film problem should be identical to Equation 6.34 if we interchange x and y. But how did we end up with such a complicated equation as Equation 6.36? We blindly chose x as the independent variable and decided to find the function y(x). In fact, in general, we can choose the independent variable to be anything we want: x, θ , t, or even y. If we choose y as the independent variable, we would need to interchange x and y in many of the previous equations that led up to Euler's equation (Equation 6.18). It might be easier in the beginning to just interchange the variables that we started with (i.e., call the horizontal axis y in Figure 6-6 and let the independent variable be x). (In a right-handed coordinate system, the x-direction would be down, but that presents no difficulty in this case because of symmetry.) No matter what we do, the solution of our present problem would just parallel Example 6.3. Unfortunately, it is not always possible to look ahead to make the best choice of independent variable. Sometimes we just have to proceed by trial and error

6.4 The "Second Form" of the Euler Equation

A second equation may be derived from Euler's equation that is convenient for functions that do not explicitly depend on x: $\partial f/\partial x = 0$. We first note that for any function f{y, y'; x) the derivative is a sum of terms

$$\begin{aligned}\frac{df}{dx} = \frac{d}{dx}f\{y, y'; x\} &= \frac{\partial f}{\partial y}\frac{dy}{dx} + \frac{\partial f}{\partial y'}\frac{dy'}{dx} + \frac{\partial f}{\partial x} \\ &= y'\frac{\partial f}{\partial y} + y''\frac{\partial f}{\partial y'} + \frac{\partial f}{\partial x}\end{aligned} \tag{6.37}$$

Also

$$\frac{d}{dx}\left(y'\frac{\partial f}{\partial y'}\right) = y''\frac{\partial f}{\partial y'} + y'\frac{d}{dx}\frac{\partial f}{\partial y'}$$

or, substituting from Equation 6.37 for $y''(\partial f/\partial y')$,

$$\frac{d}{dx}\left(y'\frac{\partial f}{\partial y'}\right) = \frac{df}{dx} - \frac{\partial f}{\partial x} - y'\frac{\partial f}{\partial y} + y'\frac{d}{dx}\frac{\partial f}{\partial y'} \tag{6.38}$$

The last two terms in Equation 6.38 may be written as

$$y'\left(\frac{d}{dx}\frac{\partial f}{\partial y'} - \frac{\partial f}{\partial y}\right)$$

which vanishes in view of the Euler equation (Equation 6.18). Therefore,

$$\boxed{\frac{\partial f}{\partial x} - \frac{d}{dx}\left(f - y'\frac{\partial f}{\partial y'}\right) = 0} \tag{6.39}$$

We can use this so-called "second form" of the Euler equation in cases in which f does not depend explicitly on x, and $\partial f/\partial x = 0$. Then,

$$f - y'\frac{\partial f}{\partial y'} = \text{constant} \qquad \left(\text{for } \frac{\partial f}{\partial x} = 0\right) \tag{6.40}$$

EXAMPLE 6.4

A *geodesic* is a line that represents the shortest path between any two points when the path is restricted to a particular surface. Find the geodesic on a sphere.

Solution. The element of length on the surface of a sphere of radius ρ is given (see Equation F.15 with dr = 0) by

$$ds = \rho(d\theta^2 + \sin^2\theta\, d\phi^2)^{1/2} \tag{6.41}$$

The distance s between points 1 and 2 is therefore

$$s = \rho\int_1^2\left[\left(\frac{d\theta}{d\phi}\right)^2 + \sin^2\theta\right]^{1/2} d\phi \tag{6.42}$$

and, if s is to be a minimum, f is identified as

$$f = (\theta'^2 + \sin^2\theta)^{1/2} \tag{6.43}$$

where $\theta' = d\theta/d\phi$. Because $\partial f/\partial\phi = 0$, we may use the second form of the Euler equation (Equation 6.40), which yields

$$(\theta'^2 + \sin^2\theta)^{1/2} - \theta'\cdot\frac{\partial}{\partial\theta'}(\theta'^2 + \sin^2\theta)^{1/2} = \text{constant} \equiv a \tag{6.44}$$

Differentiating and multiplying through by f we have

$$\sin^2\theta = a(\theta'^2 + \sin^2\theta)^{1/2} \tag{6.45}$$

This may be solved for $d\phi/d\theta = \theta'^{-1}$, with the result

$$\frac{d\phi}{d\theta} = \frac{a\csc^2\theta}{(1 - a^2\csc^2\theta)^{1/2}} \tag{6.46}$$

Solving for ϕ , we obtain

$$\phi = \sin^{-1}\left(\frac{\cot\theta}{\beta}\right) + \alpha \tag{6.47}$$

where α is the constant of integration and $\beta^2 = (1 - a^2)/a^2$. Rewriting Equation 6.47 produces

$$\cot\theta = \beta \sin(\phi - \alpha) \tag{6.48}$$

To interpret this result, we convert the equation to rectangular coordinates by multiplying through by $\rho \sin\theta$ to obtain, on expanding $\sin(\phi - \alpha)$,

$$(\beta \cos\alpha)\rho \sin\theta \sin\phi - (\beta \sin\alpha)\rho \sin\theta \cos\phi = \rho \cos\theta \tag{6.49}$$

Because α and β are constants, we may write them a

$$\beta \cos\alpha \equiv A, \qquad \beta \sin\alpha \equiv B \tag{6.50}$$

Then Equation 6.49 becomes

$$A(\rho \sin\theta \sin\phi) - B(\rho \sin\theta \cos\phi) = (\rho \cos\theta) \tag{6.51}$$

The quantities in the parentheses are just the expressions for y, x, and z, respectively, in spherical coordinates (see Figure F-3, Appendix F); therefore Equation 6.51 may be written as

$$Ay - Bx = z \tag{6.52}$$

which is the equation of a plane passing through the center of the sphere.

Hence the geodesic on a sphere is the path that the plane forms at the intersection with the surface of the sphere - a *great circle*. Note that the great circle is the maximum as well as the minimum "straight-line" distance between two points on the surface of a sphere.

6.5 Functions with Several Dependent Variables

The Euler equation derived in the preceding section is the solution of the variational problem in which it was desired to find the single function y(x) such that the integral of the functional/was an extremum. The case more commonly encountered in mechanics is that in which/is a functional of several dependent variables:

$$f = f\{y_1(x), y_1'(x), y_2(x), y_2'(x), \cdots; x\} \tag{6.53}$$

or simply

$$f = f\{y_i(x), y_i'(x); x\}, \quad i = 1, 2, \cdots, n \tag{6.54}$$

In analogy with Equation 6.2, we write

$$y_i(\alpha, x) = y_i(0, x) + \alpha\eta_i(x) \tag{6.55}$$

The development proceeds analogously (cf. Equation 6.17), resulting in

$$\frac{\partial J}{\partial \alpha} = \int_{x_1}^{x_2} \sum_i \left(\frac{\partial f}{\partial y_i} - \frac{d}{dx}\frac{\partial f}{\partial y_i'} \right) \eta_i(x)\, dx \tag{6.56}$$

Because the individual variations - the $\eta(x)$ - are all independent, the vanishing of Equation 6.56 when evaluated at $\alpha = 0$ requires the separate vanishing of each expression in the brackets:

$$\boxed{\frac{\partial f}{\partial y_i} - \frac{d}{dx}\frac{\partial f}{\partial y_i'} = 0, \qquad i = 1, 2, \cdots, n} \tag{6.57}$$

6.6 Euler's Equations When Auxiliary Conditions Are Imposed

Suppose we want to find, for example, the shortest path between two points on a surface. Then, in addition to the conditions already discussed, there is the condition that the path must satisfy the equation of the surface, say, $g\{y_i; x\} = 0$. Such an equation was implicit in the solution of Example 6.4 for the geodesic on a sphere where the condition was

$$g = \sum_i x_i^2 - \rho^2 = 0 \tag{6.58}$$

that is.

$$r = \rho = \text{constant} \tag{6.59}$$

But in the general case, we must make explicit use of the auxiliary equation or equations. These equations are also called **equations of constraint**. Consider the case in which

$$f = f\{y_i, y_i'; x\} = f\{y, y', z, z'; x\} \tag{6.60}$$

The equation corresponding to Equation 6.17 for the case of *two* variables is

$$\frac{\partial J}{\partial \alpha} = \int_{x_1}^{x_2} \left[\left(\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y'} \right) \frac{\partial y}{\partial \alpha} + \left(\frac{\partial f}{\partial z} - \frac{d}{dx} \frac{\partial f}{\partial z'} \right) \frac{\partial z}{\partial \alpha} \right] dx \tag{6.61}$$

But now there also exists an equation of constraint of the form

$$g\{y_i; x\} = g\{y, z; x\} = 0 \tag{6.62}$$

and the variations $\partial y/\partial \alpha$ and $\partial z/\partial \alpha$ are no longer independent, so the expressions in parentheses in Equation 6.61 do not separately vanish at $\alpha = 0$.

Differentiating g from Equation 6.62, we have

$$dg = \left(\frac{\partial g}{\partial y} \frac{\partial y}{\partial \alpha} + \frac{\partial g}{\partial z} \frac{\partial z}{\partial \alpha} \right) d\alpha = 0 \tag{6.63}$$

where no term in x appears since $\partial x/\partial\alpha = 0$. Now

$$\left.\begin{aligned} y(\alpha, x) &= y(x) + \alpha\eta_1(x) \\ z(\alpha, x) &= z(x) + \alpha\eta_2(x) \end{aligned}\right\} \tag{6.64}$$

Therefore, by determining $\partial y/\partial\alpha$ and $\partial z/\partial\alpha$ from Equation 6.64 and inserting into the term in parentheses of Equation 6.63, which, in general, must be zero, we obtain

$$\frac{\partial g}{\partial y}\eta_1(x) = -\frac{\partial g}{\partial z}\eta_2(x) \tag{6.65}$$

Equation 6.61 becomes

$$\frac{\partial J}{\partial\alpha} = \int_{x_1}^{x_2}\left[\left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right)\eta_1(x) + \left(\frac{\partial f}{\partial z} - \frac{d}{dx}\frac{\partial f}{\partial z'}\right)\eta_2(x)\right]dx$$

Factoring $\eta_1(x)$ out of the square brackets and writing Equation 6.65 as

$$\frac{\eta_2(x)}{\eta_1(x)} = -\frac{\partial g/\partial y}{\partial g/\partial z}$$

we have

$$\frac{\partial J}{\partial\alpha} = \int_{x_1}^{x_2}\left[\left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right) - \left(\frac{\partial f}{\partial z} - \frac{d}{dx}\frac{\partial f}{\partial z'}\right)\left(\frac{\partial g/\partial y}{\partial g/\partial z}\right)\right]\eta_1(x)\,dx \tag{6.66}$$

This latter equation now contains the single arbitrary function $\eta_1(x)$, which is not in any way restricted by Equation 6.64, and on requiring the condition of Equation 6.4, the expression in the brackets must vanish. Thus we have

$$\left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right)\left(\frac{\partial g}{\partial y}\right)^{-1} = \left(\frac{\partial f}{\partial z} - \frac{d}{dx}\frac{\partial f}{\partial z'}\right)\left(\frac{\partial g}{\partial z}\right)^{-1} \tag{6.67}$$

The left-hand side of this equation involves only derivatives of f and g with respect to y and y', and the right-hand side involves only derivatives with respect to z and z'. Because y and z are both functions of x, the two sides of Equation 6.67 may be set equal to a function of x, which we write as $-\lambda(x)$:

$$\left.\begin{aligned}\frac{\partial f}{\partial y}-\frac{d}{dx}\frac{\partial f}{\partial y'}+\lambda(x)\frac{\partial g}{\partial y}=0\\ \frac{\partial f}{\partial z}-\frac{d}{dx}\frac{\partial f}{\partial z'}+\lambda(x)\frac{\partial g}{\partial z}=0\end{aligned}\right\} \tag{6.68}$$

The complete solution to the problem now depends on finding three functions: y(x), z(x), and $\lambda(x)$. But there are three relations that may be used: the two equations (Equation 6.68) and the equation of constraint (Equation 6.62). Thus, there is a sufficient number of relations to allow a complete solution. Note that here $\lambda(x)$ is considered to be *undetermined* and is obtained as a part of the solution. The function $\lambda(x)$ is known as a **Lagrange undetermined multiplier**.

For the general case of several dependent variables and several auxiliary conditions, we have the following set of equations:

$$\frac{\partial f}{\partial y_i}-\frac{d}{dx}\frac{\partial f}{\partial y_i'}+\sum_j \lambda_j(x)\frac{\partial g_j}{\partial y_i}=0 \tag{6.69}$$

$$g_j\{y_i;\,x\}=0 \tag{6.70}$$

If i = 1, 2, … ,m, and j = 1, 2, …. , n, Equation 6.69 represents m equations in m + n unknowns, but there are also the n equations of constraint (Equation 6.70). Thus, there are m + n equations in m + n unknowns, and the system is soluble.

Equation 6.70 is equivalent to the set of n differential equations

$$\sum_i \frac{\partial g_j}{\partial y_i}\, dy_i = 0, \quad \begin{cases} i = 1, 2, \cdots, m \\ j = 1, 2, \cdots, n \end{cases} \tag{6.71}$$

In problems in mechanics, the constraint equations are frequently differential equations rather than algebraic equations. Therefore, equations such as Equation 6.71 are sometimes more useful than the equations represented by Equation 6.70.

EXAMPLE 6.5

Consider a disk rolling without slipping on an inclined plane (Figure 6-7).

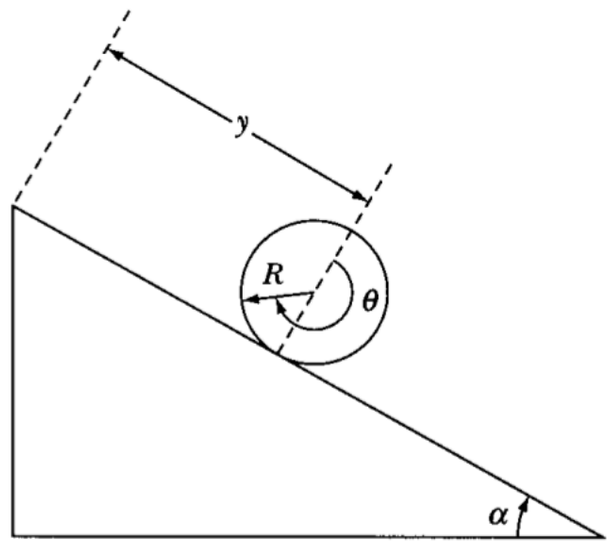


FIGURE 6-7 Example 6.5. A disk rolls down an inclined plane without slipping.

Determine the equation of constraint in terms of the "coordinates" y and θ .

Solution. The relation between the coordinates (which are not independent) is

$$y = R\theta \tag{6.72}$$

where R is the radius of the disk. Hence the equation of constraint is

$$g(y, \theta) = y - R\theta = 0 \tag{6.73}$$

and

$$\frac{\partial g}{\partial y} = 1, \qquad \frac{\partial g}{\partial \theta} = -R \tag{6.74}$$

are the quantities associated with λ , the single undetermined multiplier for this case.

The constraint equation can also appear in an integral form. Consider the isoperimetric problem that is stated as finding the curve y = y(x) for which the functional

$$J[y] = \int_a^b f\{y, y'; x\}dx \tag{6.75}$$

has an extremum, and the curve y(x) satisfies boundary conditions y(a) = A and y(b) = B as well as the second functional

$$K[y] = \int_a^b g\{y, y'; x\}dx \tag{6.76}$$

that has a fixed value for the length of the curve (ℓ). This second functional represents an integral constraint.

$$\int_a^b (f + \lambda g)\, dx. \tag{6.77}$$

Similarly to what we have done previously, there will be a constant λ such that y(x) is the extremal solution of the functional

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} + \lambda\left(\frac{\partial g}{\partial y} - \frac{d}{dx}\frac{\partial g}{\partial y'}\right) = 0 \tag{6.78}$$

subject to the constraints y(a) = A, y(b) = B, and K[y] = ℓ . We will work an example for this so-called *Dido Problem*.

EXAMPLE 6.6

One version of the Dido Problem is to find the curve y(x) of length ℓ bounded by the x-axis on the bottom that passes through the points (-a, 0) and (a, 0) and encloses the largest area. The value of the endpoints a is determined by the problem.

Solution. We can use the equations just developed to solve this problem. We show in Figure 6-8 that the differential area dA = y dx.

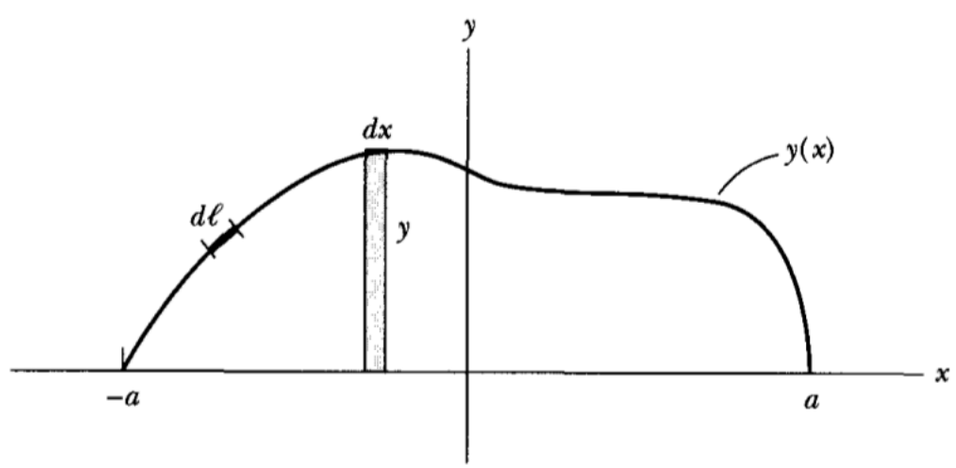


FIGURE 6-8 Example 6.6. We want to find the curve $y(x)$ that maximizes the area above the $y = 0$ line consistent with a fixed perimeter length. The curve must go through $x = -a$ and a. The differential area $dA = ydx$, and the differential length along the curve is $d\ell$.

We want to maximize the area, so we want to find the extremum solution for Equation 6.75, which becomes

$$J = \int_{-a}^{a} y\,dx \tag{6.79}$$

The constraint equations are

$$y(x): y(-a) = 0, y(a) = 0 \quad \text{and} \quad K = \int d\ell = \ell. \tag{6.80}$$

The differential length along the curve $d\ell = (dx^2 + dy^2)^{1/2} = (1 + y'^2)^{1/2}$ where y' = dy/dx. The constraint functional becomes

$$K = \int_{-a}^{a} [1 + y'^2]^{1/2} dx = \ell. \tag{6.81}$$

We now have y(x) = y and g(x) = $\sqrt{1 + y'^2}$, and we use these functions in Equation 6.78.

$$\frac{\partial f}{\partial y} = 1, \quad \frac{\partial f}{\partial y'} = 0, \quad \frac{\partial g}{\partial y} = 0, \quad \frac{\partial g}{\partial y'} = \frac{y'}{(1 + y'^2)^{1/2}}$$

Equation 6.78 becomes

$$1 - \lambda \frac{d}{dx}\left[\frac{y'}{(1 + y'^2)^{1/2}}\right] = 0 \tag{6.82}$$

We manipulate Equation 6.82 to find

$$\frac{d}{dx}\left[\frac{y'}{(1 + y'^2)^{1/2}}\right] = \frac{1}{\lambda} \tag{6.83}$$

We integrate over x to find

$$\frac{\lambda y'}{\sqrt{(1 + y'^2)}} = x - C_1$$

where C_1 is an integration constant. This can be rearranged to be

$$dy = \frac{\pm (x - C_1)\, dx}{\sqrt{\lambda^2 - (x - C_1)^2}}$$

This equation is integrated to find

$$y = \mp \sqrt{\lambda^2 - (x - C_1)^2} + C_2 \tag{6.84}$$

where C_2 is another integration constant. We can rewrite this as the equation of a circle of radius λ .

$$(x - C_1)^2 + (y - C_2)^2 = \lambda^2 \tag{6.85}$$

The maximum area is a semicircle bounded by the y = 0 line. The semicircle must go through (x, y) points of (-a, 0) and (a, 0), which means the circle must be centered at the origin, so that $C_1 = 0 = C_2$, and the radius = a = λ . The perimeter of the top half of the semicircle is what we called ℓ , and the perimeter length of a half circle is it π a. Therefore, we have $\pi a = \ell$, and a = ℓ/π .

6.7 The δ Notation

In analyses that use the calculus of variations, we customarily use a shorthand notation to represent the variation. Thus, Equation 6.17, which can be written as

$$\frac{\partial J}{\partial \alpha}\, d\alpha = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} \right) \frac{\partial y}{\partial \alpha}\, d\alpha\, dx \tag{6.86}$$

may be expressed as

$$\delta J = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} \right) \delta y\, dx \tag{6.87}$$

where

$$\left.\begin{aligned} \frac{\partial J}{\partial \alpha} d\alpha &\equiv \delta J \\ \frac{\partial y}{\partial \alpha} d\alpha &\equiv \delta y \end{aligned}\right\} \tag{6.88}$$

The condition of extremum then becomes

$$\delta J = \delta \int_{x_1}^{x_2} f\{y, y'; x\} dx = 0 \tag{6.89}$$

Taking the variation symbol δ inside the integral (because, by hypothesis, the limits of integration are not affected by the variation), we have

$$\begin{aligned}\delta J &= \int_{x_1}^{x_2} \delta f \, dx \\ &= \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} \delta y + \frac{\partial f}{\partial y'} \delta y' \right) dx \end{aligned} \tag{6.90}$$

But

$$\delta y' = \delta \left(\frac{dy}{dx} \right) = \frac{d}{dx} (\delta y) \tag{6.91}$$

so

$$\delta J = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} \delta y + \frac{\partial f}{\partial y'} \frac{d}{dx} \delta y \right) dx \tag{6.92}$$

Integrating the second term by parts as before, we find

$$\delta J = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial y'} \right) \delta y \, dx \tag{6.93}$$

Because the variation δy is arbitrary, the extremum condition $\delta J = 0$ requires the integrand to vanish, thereby yielding the Euler equation (Equation 6.18).

Although the δ notation is frequently used, it is important to realize that it is only a shorthand expression of the more precise differential quantities. The varied path represented by δy can be thought of physically as a virtual displacement from the actual path consistent with all the forces and constraints (see Figure 6-9).

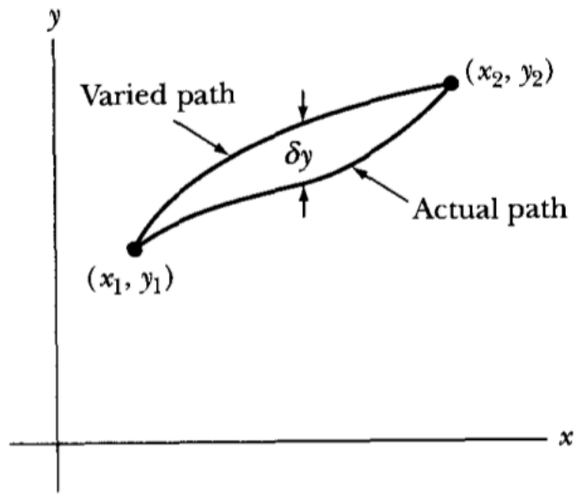


FIGURE 6-9 The varied path is a virtual displacement δy from the actual path consistent with all the forces and constraints.

This variation δy is distinguished from an actual differential displacement dy by the condition that dt = 0 - that is, that time is fixed. The varied path δy, in fact, need not even correspond to a possible path of motion. The variation must vanish at the endpoints.

CHAPTER 7 - Hamilton's Principle - Lagrangian and Hamiltonian Dynamics

7.1 Introduction

Experience has shown that a particle's motion in an inertial reference frame is correctly described by the Newtonian equation $\mathbf{F} = \dot{\mathbf{p}}$. If the particle is not required to move in some complicated manner and if rectangular coordinates are used to describe the motion, then usually the equations of motion are relatively simple. But if either of these restrictions is removed, the equations can become quite complex and difficult to manipulate. For example, if a particle is constrained to move on the surface of a sphere, the equations of motion result from the projection of the Newtonian vector equation onto that surface. The representation of the acceleration vector in spherical coordinates is a formidable expression.

Moreover, if a particle is constrained to move on a given surface, certain forces must exist (called **forces of constraint**) that maintain the particle in contact with the specified surface. For a particle moving on a smooth horizontal surface, the force of constraint is simply $\mathbf{F}_c = -m\mathbf{g}$. But, if the particle is, say, a bead sliding down a curved wire, the force of constraint can be quite complicated. Indeed, in particular situations it may be difficult or even impossible to obtain explicit expressions for the forces of constraint. But in solving a problem by using the Newtonian procedure, we must know *all* the forces, because the quantity **F** that appears in the fundamental equation is the *total* force acting on a body.

To circumvent some of the practical difficulties that arise in attempts to apply Newton's equations to particular problems, alternate procedures may be developed. All such approaches are in essence *a posteriori*, because we know beforehand that a result equivalent to the Newtonian equations must be obtained. Thus, to effect a simplification we need not formulate a *new* theory of mechanics - the Newtonian theory is quite correct - but only devise an alternate method of dealing with complicated problems in a general manner. Such a method is contained in **Hamilton's Principle**, and the equations of motion resulting from the application of this principle are called **Lagrange's equations**.

If Lagrange's equations are to constitute a proper description of the dynamics of particles, they must be equivalent to Newton's equations. On the other hand, Hamilton's Principle can be applied to a wide range of physical phenomena (particularly those involving *fields*) not usually associated with Newton's equations. To be sure, each of the results that can be obtained from Hamilton's Principle was first obtained, as were Newton's equations, by the correlation of experimental facts. Hamilton's Principle has not provided us with any new physical theories, but it has allowed a satisfying unification of many individual theories by a single basic postulate. This is not an idle exercise in hindsight, because it is the goal of physical theory not only to give precise mathematical formulation to observed phenomena but also to describe these effects with an economy of fundamental postulates and in the most unified manner possible. Indeed, Hamilton's Principle is one of the most elegant and far-reaching principles of physical theory.

In view of its wide range of applicability (even though this is an after-the-fact discovery), it is not unreasonable to assert that Hamilton's Principle is more "fundamental" than Newton's equations. Therefore, we proceed by first postulating Hamilton's Principle; we then obtain Lagrange's equations and show that these are equivalent to Newton's equations.

Because we have already discussed (in Chapters 2, 3, and 4) dissipative phenomena at some length, we henceforth confine our attention to conservative systems. Consequently, we do not discuss the more general set of Lagrange's equations, which take into account the effects of nonconservative forces. The reader is referred to the literature for these details.

7.2 Hamilton's Principle

Minimal principles in physics have a long and interesting history. The search for such principles is predicated on the notion that nature always minimizes certain important quantities when a physical process takes place. The first such minimum principles were developed in the field of optics.

Hero of Alexandria, in the second century B.C., found that the law governing the reflection of light could be obtained by asserting that a light ray, traveling from one point to another by a reflection from a plane mirror, always takes the shortest possible path. A simple geometric construction verifies that this minimum principle does indeed lead to the equality of the angles of incidence and reflection for a light ray reflected from a plane mirror. Hero's principle of the shortest path cannot, however, yield a correct law for refraction. In 1657, Fermat reformulated the principle by postulating that a light ray always travels from one point to another in a medium by a path that requires the least time. Fermat's principle of least time leads immediately, not only to the correct law of reflection, but also to Snell's law of refraction.

Minimum principles continued to be sought, and in the latter part of the seventeenth century the beginnings of the calculus of variations were developed by Newton, Leibniz, and the Bernoullis when such problems as the brachistochrone and the shape of a hanging chain (a catenary) were solved.

The first application of a general minimum principle in mechanics was made in 1747 by Maupertuis, who asserted that dynamical motion takes place with minimum action. Maupertuis's principle of least action was based on theological grounds (action is minimized through the "wisdom of God"), and his concept of "action" was rather vague. (*Action*, as we will see, is a quantity with the dimensions of *length* x *momentum* or *energy* x *time*.) Only later was a firm mathematic foundation of the principle given by Lagrange. Although it is a useful form from which to make the transition from classical mechanics to optics and to quantum mechanics, the principle of least action is less general than Hamilton's Principle and, indeed, can be derived from it.

In 1828, Gauss developed a method of treating mechanics by his principle of **least constraint**; a modification was later made by Hertz and embodied in his principle of **least curvature**. These principles are closely related to Hamilton's Principle and add nothing to the content of Hamilton's more general formulation; their mention only emphasizes the continual concern with minimal principles in physics.

In two papers published in 1834 and 1835, Hamilton announced the dynamical principle on which it is possible to base all of mechanics and, indeed, most of classical physics. Hamilton's Principle may be stated as follows:

> *Of all the possible paths along which a dynamical system may move from one point to another within a specified time interval (consistent with any constraints), the actual path followed is that which minimizes the time integral of the difference between the kinetic and potential energies.*

In terms of the calculus of variations, Hamilton's Principle becomes

$$\delta \int_{t_1}^{t_2} (T - U)\, dt = 0 \tag{7.1}$$

where the symbol δ is a shorthand notation to describe the variation discussed in Sections 6.3 and 6.7.

This variational statement of the principle requires only that the integral of T - U be an *extremum*, not necessarily a *minimum*. But in almost all important applications in dynamics, the *minimum* condition occurs.

The kinetic energy of a particle expressed in fixed, rectangular coordinates is a function only of the $\dot{x}_i$, and if the particle moves in a conservative force field, the potential energy is a function only of the x_i:

$$T = T(\dot{x}_i), \quad U = U(x_i)$$

If we define the difference of these quantities to be

$$L \equiv T - U = L(x_i, \dot{x}_i) \tag{7.2}$$

then Equation 7.1 becomes

$$\boxed{\delta \int_{t_1}^{t_2} L(x_i, \dot{x}_i)\, dt = 0} \tag{7.3}$$

The function L appearing in this expression may be identified with the function f of the variational integral (see Section 6.5),

$$\delta \int_{x_1}^{x_2} f\{y_i(x), y_i'(x); x\}\, dx$$

if we make the transformations

$$x \rightarrow t$$
$$y_i(x) \rightarrow x_i(t)$$
$$y_i'(x) \rightarrow \dot{x}_i(t)$$
$$f\{y_i(x), y_i'(x); x\} \rightarrow L(x_i, \dot{x}_i)$$

The Euler-Lagrange equations (Equation 6.57) corresponding to Equation 7.3 are therefore

$$\boxed{\frac{\partial L}{\partial x_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}_i} = 0, \quad i = 1, 2, 3} \qquad \text{Lagrange equations of motion} \tag{7.4}$$

These are the **Lagrange equations of motion** for the particle, and the quantity L is called the **Lagrange function** or **Lagrangian** for the particle.

By way of example, let us obtain the Lagrange equation of motion for the one-dimensional harmonic oscillator. With the usual expressions for the kinetic and potential energies, we have

$$L = T - U = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2$$

$$\frac{\partial L}{\partial x} = -kx$$

$$\frac{\partial L}{\partial \dot{x}} = m\dot{x}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = m\ddot{x}$$

Substituting these results into Equation 7.4 leads to

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

which is identical with the equation of motion obtained using Newtonian mechanics.

The Lagrangian procedure seems needlessly complicated if it can only duplicate the simple results of Newtonian theory. However, let us continue illustrating the method by considering the plane pendulum (see Section 4.4). Using Equation 4.23 for T and U, we have, for the Lagrangian function

$$L = \frac{1}{2} ml^2\dot{\theta}^2 - mgl(1 - \cos\theta)$$

We now treat 6 as if it were a rectangular coordinate and apply the operations specified in Equation 7.4; we obtain

$$\frac{\partial L}{\partial \theta} = -mgl \sin\theta$$

$$\frac{\partial L}{\partial \dot{\theta}} = ml^2\dot{\theta}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) = ml^2\ddot{\theta}$$

$$\ddot{\theta} + \frac{g}{l}\sin\theta = 0$$

which again is identical with the Newtonian result (Equation 4.21). This is a remarkable result; it has been obtained by calculating the kinetic and potential energies in terms of θ rather than x and then applying a set of operations designed for use with rectangular rather than angular coordinates. We are therefore led to suspect that the Lagrange equations are more general and useful than the form of Equation 7.4 would indicate. We pursue this matter in Section 7.4.

Another important characteristic of the method used in the two preceding simple examples is that nowhere in the calculations did there enter any statement regarding force. The equations of motion were obtained only by specifying certain properties associated with the particle (the kinetic and potential energies), and without the necessity of explicitly taking into account the fact that there was an external agency acting on the particle (the force). Therefore, insofar as energy can be defined independently of Newtonian concepts, Hamilton's Principle allows us to calculate the equations of motion of a body completely without recourse to Newtonian theory. We shall return to this important point in Sections 7.5 and 7.7.

7.3 Generalized Coordinates

We now seek to take advantage of the flexibility in specifying coordinates that the two examples of the preceding section have suggested is inherent in Lagrange's equations.

We consider a general mechanical system consisting of a collection of n discrete point particles, some of which may be connected to form rigid bodies. We discuss such systems of particles in Chapter 9 and rigid bodies in Chapter 11. To specify the state of such a system at a given time, it is necessary to use n radius vectors. Because each radius vector consists of three numbers (e.g., the rectangular coordinates), 3n quantities must be specified to describe the positions of all the particles. If there exist equations of constraint that relate some of these coordinates to others (as would be the case, for example, if some of the particles were connected to form rigid bodies or if the motion were constrained to lie along some path or on some surface), then not all the Sn coordinates are independent.

In fact, if there are m equations of constraint, then 3n - m coordinates are independent, and the system is said to possess 3n- m *degrees of freedom*.

It is important to note that if s = 3n - m coordinates are required in a given case, we need not choose s rectangular coordinates or even s curvilinear coordinates (e.g., spherical, cylindrical). We can choose *any* s independent parameters, as long as they completely specify the state of the system. These s quantities need not even have the dimensions of length. Depending on the problem at hand, it may prove more convenient to choose some of the parameters with dimensions of *energy*, some with dimensions of $(length)^2$, some that are *dimensionless*, and so forth. In Example 6.5, we described a disk rolling down an inclined plane in terms of one coordinate that was a length and one that was an angle. We give the name **generalized coordinates** to any set of quantities that completely specifies the state of a system. The generalized coordinates are customarily written as $q_1, q_2, \ldots$, or simply as the q_j. A set of independent generalized coordinates whose number equals the number s of degrees of freedom of the system and not restricted by the constraints is called a *proper* set of generalized coordinates. In certain instances, it may be advantageous to use generalized coordinates whose number exceeds the number of degrees of freedom and to explicitly take into account the constraint relations through the use of the Lagrange undetermined multipliers. Such would be the case, for example, if we desired to calculate the forces of constraint.

The choice of a set of generalized coordinates to describe a system is not unique; there are in general many sets of quantities (in fact, an infinite number!) that completely specify the state of a given system. For example, in the problem of the disk rolling down the inclined plane, we might choose as coordinates the height of the center of mass of the disk above some reference level and the distance through which some point on the rim has traveled since the start of the motion. The ultimate test of the "suitability" of a particular set of generalized coordinates is whether the resulting equations of motion are sufficiently simple to allow a straightforward interpretation.

Unfortunately, we can state no general rules for selecting the "most suitable" set of generalized coordinates for a given problem. A certain skill must be developed through experience, and we present many examples in this chapter.

In addition to the generalized coordinates, we may define a set of quantities consisting of the time derivatives of $\dot{q}_1, \dot{q}_2, \dot{q}_3, \ldots\ldots\ldots$, or simply $\dot{q}_j$. In analogy with the nomenclature for rectangular coordinates, we call $\dot{q}_j$ the **generalized velocities**.

If we allow for the possibility that the equations connecting $x_{\alpha,i}$ and q_j explicitly contain the time, then the set of transformation equations is given by

$$\begin{aligned} x_{\alpha,i} &= x_{\alpha,i}(q_1, q_2, \ldots, q_s, t), \quad \begin{cases} \alpha = 1, 2, \ldots, n \\ i = 1, 2, 3 \end{cases} \\ &= x_{\alpha,i}(q_j, t), \qquad\qquad j = 1, 2, \ldots, s \end{aligned} \tag{7.5}$$

I note that in this chapter, we attempt to simplify the notation by reserving the subscript i to designate rectangular axes; therefore, we always have i = 1, 2, 3.

In general, the rectangular components of the velocities depend on the generalized coordinates, the generalized velocities, and the time:

$$\dot{x}_{\alpha,i} = \dot{x}_{\alpha,i}(q_j, \dot{q}_j, t) \tag{7.6}$$

We may also write the inverse transformations as

$$q_j = q_j(x_{\alpha,i}, t) \tag{7.7}$$

$$\dot{q}_j = \dot{q}_j(x_{\alpha,i}, \dot{x}_{\alpha,i}, t) \tag{7.8}$$

Also, there are m = 2n - s equations of constraint of the form

$$f_k(x_{\alpha,i}, t) = 0, \quad k = 1, 2, \ldots, m \tag{7.9}$$

EXAMPLE 7.1

Find a suitable set of generalized coordinates for a point particle moving on the surface of a hemisphere of radius R whose center is at the origin.

Solution. Because the motion always takes place on the surface, we have

$$x^2 + y^2 + z^2 - R^2 = 0, \quad z \geq 0 \tag{7.10}$$

Let us choose as our generalized coordinates the cosines of the angles between the x-, y-, and z-axes and the line connecting the particle with the origin.

Therefore,

$$q_1 = \frac{x}{R}, \qquad q_2 = \frac{y}{R}, \qquad q_3 = \frac{z}{R} \tag{7.11}$$

But the sum of the squares of the direction cosines of a line equals unity. Hence,

$$q_1^2 + q_2^2 + q_3^2 = 1 \tag{7.12}$$

This set of q_j does not constitute a proper set of generalized coordinates, because we can write q_3 as a function of q_1 and q_2.

$$q_3 = \sqrt{1 - q_1^2 - q_2^2} \tag{7.13}$$

We may, however, choose $q_1 = x/R$ and $q_2 = y/R$ as proper generalized coordinates, and these quantities, together with the equation of constraint (Equation 7.13)

$$z = \sqrt{R^2 - x^2 - y^2} \tag{7.14}$$

are sufficient to uniquely specify the position of the particle. This should be an obvious result, because only two coordinates (e.g., latitude and longitude) are necessary to specify a point on the surface of a sphere. But the example illustrates the fact that the equations of constraint can always be used to reduce a trial set of coordinates to a proper set of generalized coordinates.

EXAMPLE 7.2

Use the (x, y) coordinate system of Figure 7-1 to find the kinetic energy T, potential energy U, and the Lagrangian L for a simple pendulum (length ℓ , mass bob m) moving in the x, y plane.

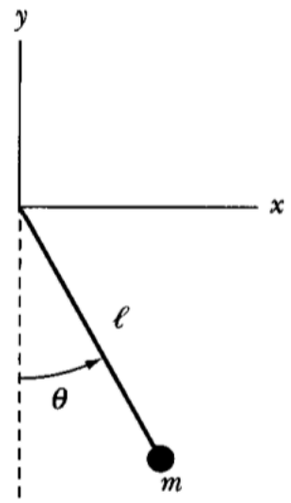


FIGURE 7-1 Example 7.2. A simple pendulum of length ℓ and bob of mass m.

Determine the transformation equations from the (x, y) rectangular system to the coordinate θ . Find the equation of motion.

Solution. We have already examined this general problem in Sections 4.4 and 7.1. When using the Lagrangian method, it is often useful to begin with rectangular coordinates and transform to the most obvious system with the simplest generalized coordinates. In this case, the kinetic and potential energies and the Lagrangian become

$$T = \frac{1}{2} m\dot{x}^2 + \frac{1}{2} m\dot{y}^2$$

$$U = mgy$$

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

Inspection of Figure 7-1 reveals that the motion can be better described by using θ and $\dot{\theta}$. Let's transform x and y into the coordinate θ and then find L in terms of θ.

$$x = \ell \sin\theta$$
$$y = -\ell \cos\theta$$

We now find for $\dot{x}$ and $\dot{y}$

$$\dot{x} = \ell\dot{\theta}\cos\theta$$

$$\dot{y} = \ell\dot{\theta}\sin\theta$$

$$L = \frac{m}{2}(\ell^2\dot{\theta}^2\cos^2\theta + \ell^2\dot{\theta}^2\sin^2\theta) + mg\ell\cos\theta = \frac{m}{2}\ell^2\dot{\theta}^2 + mg\ell\cos\theta$$

The only generalized coordinate in the case of the pendulum is the angle θ, and we have expressed the Lagrangian in terms of θ by following a simple procedure of finding L in terms of $\dot{x}$ and $\dot{y}$, finding the transformation equations, and then inserting them into the expression for L. If we do as we did in the previous section and treat θ *as if it were a rectangular coordinate*, we can find the equation of motion as follows:

$$\frac{\partial L}{\partial\theta} = -mg\ell\sin\theta$$

$$\frac{\partial L}{\partial\dot{\theta}} = m\ell^2\dot{\theta}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{\theta}}\right) = m\ell^2\ddot{\theta}$$

We insert these relations into Equation 7.4 to find the same equation of motion as found previously.

$$\ddot{\theta} + \frac{g}{\ell}\sin\theta = 0$$

The state of a system consisting of n particles and subject to m constraints that connect some of the 3n rectangular coordinates is completely specified by s = 3n - m generalized coordinates. We may therefore represent the state of such a system by a point in an s-dimensional space called **configuration space**. Each dimension of this space corresponds to one of the q_j coordinates. We may represent the time history of a system by a curve in configuration space, each point specifying the configuration of the system at a particular instant. Through each such point passes an infinity of curves representing possible motions of the system; each curve corresponds to a particular set of initial conditions. We may therefore speak of the "path" of a system as it "moves" through configuration space. But we must be careful not to confuse this terminology with that applied to the motion of a particle along a path in ordinary three-dimensional space.

We should also note that a dynamical path in a configuration space consisting of proper generalized coordinates is automatically consistent with the constraints on the system, because the coordinates are chosen to correspond only to realizable motions of the system.

7.4 Lagrange's Equations of Motion in Generalized Coordinates

In view of the definitions in the preceding sections, we may now restate Hamilton's Principle as follows:

> *Of all the possible paths along which a dynamical system may move from one point to another in configuration space within a specified time interval, the actual path followed is that which minimizes the time integral of the Lagrangian function for the system.*

To set up the variational form of Hamilton's Principle in generalized coordinates, we may take advantage of an important property of the Lagrangian we have not so far emphasized. The Lagrangian for a system is defined to be the difference between the kinetic and potential energies. But *energy* is a scalar quantity and so the *Lagrangian is a scalar function*. Hence the Lagrangian *must be invariant with respect to coordinate transformations*.

However, certain transformations that change the Lagrangian *but leave the equations of motion unchanged* are allowed. For example, equations of motion are unchanged if L is replaced by L + d/dt f(q_i, t) for a function f(q_i, t) with continuous second partial derivatives. As long as we define the Lagrangian to be the difference between the kinetic and potential energies, we may use different generalized coordinates. (The Lagrangian is, however, indefinite to an additive constant in the potential energy U.) It is therefore immaterial whether we express the Lagrangian in terms of $x_{\alpha,i}$ and $\dot{x}_{\alpha,i}$ or q_j and $\dot{q}_j$.

$$\begin{aligned} L &= T(\dot{\boldsymbol{x}}_{\alpha,i}) - U(\boldsymbol{x}_{\alpha,i}) \\ &= T(q_j, \dot{q}_j, t) - U(q_j, t) \end{aligned} \tag{7.15}$$

that is,

$$\begin{aligned} L &= L(q_1, q_2, \ldots, q_s; \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_s; t) \\ &= L(q_j, \dot{q}_j, t) \end{aligned} \tag{7.16}$$

Thus, Hamilton's Principle becomes

$$\boxed{\delta \int_{t_1}^{t_2} L(q_j, \dot{q}_j, t)\, dt = 0} \quad \textbf{Hamilton's Principle} \tag{7.17}$$

If we refer to the definitions of the quantities in Section 6.5 and make the identifications

$$\begin{aligned} x &\rightarrow t \\ y_i(x) &\rightarrow q_j(t) \\ y_i'(x) &\rightarrow \dot{q}_j(t) \\ f\{y_i, y_i'; x\} &\rightarrow L(q_j, \dot{q}_j, t) \end{aligned}$$

then the Euler equations (Equation 6.57) corresponding to the variational problem stated in Equation 7.17 become

$$\boxed{\frac{\partial L}{\partial q_j} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j} = 0, \qquad j = 1, 2, \ldots, s} \tag{7.18}$$

These are the Euler-Lagrange equations of motion for the system (usually called simply **Lagrange's equations**). There are s of these equations, and together with the m equations of constraint and the initial conditions that are imposed, they completely describe the motion of the system.

Because there are s second-order differential equations, 2s initial conditions must be supplied to determine the motion uniquely.

It is important to realize that the validity of Lagrange's equations requires the following two conditions:

1. The forces acting on the system (apart from any forces of constraint) must be derivable from a potential (or several potentials).
2. The equations of constraint must be relations that connect the *coordinates* of the particles and may be functions of the time—that is, we must have constraint relations of the form given by Equation 7.9.

If the constraints can be expressed as in condition 2, they are termed **holonomic** constraints. If the equations do not explicitly contain the time, the constraints are said to be **fixed** or **scleronomic**; moving constraints are **rheonomic**.

Here we consider only the motion of systems subject to conservative forces. Such forces can always be derived from potential functions, so that condition 1 is satisfied. This is not a necessary restriction on either Hamilton's Principle or Lagrange's equations; the theory can readily be extended to include nonconservative forces.

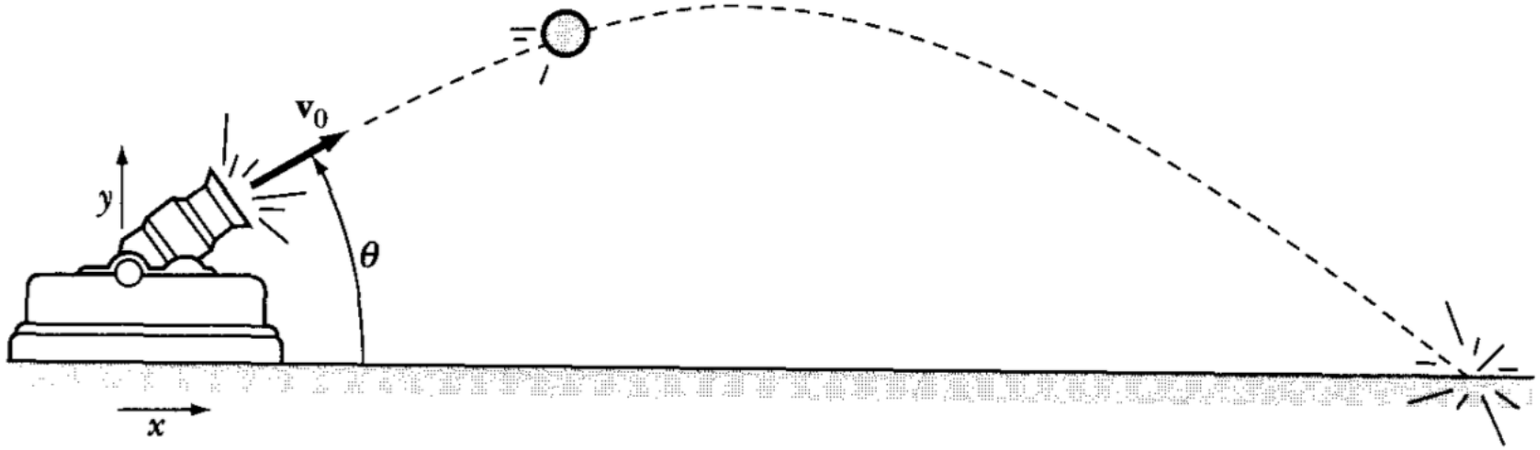
Similarly, we can formulate Hamilton's Principle to include certain types of nonholonomic constraints, but the treatment here is confined to holonomic systems. We return to nonholonomic constraints in Section 7.5.

We now want to work several examples using Lagrange's equations. Experience is the best way to determine a set of generalized coordinates, realize the constraints, and set up the Lagrangian. Once this is done, the remainder of the problem is for the most part mathematical.

EXAMPLE 7.3

Consider the case of projectile motion under gravity in two dimensions as was discussed in Example 2.6. Find the equations of motion in both Cartesian and polar coordinates.

Solution. We use Figure 2-7 to describe the system.



In Cartesian coordinates, we use x (horizontal) and y (vertical). In polar coordinates we use r (in radial direction) and θ (elevation angle from horizontal). First, in Cartesian coordinates we have

$$\left.\begin{aligned} T &= \frac{1}{2} m\dot{x}^2 + \frac{1}{2} m\dot{y}^2 \\ U &= mgy \end{aligned}\right\} \tag{7.19}$$

where U= 0 at y = 0.

$$L = T - U = \frac{1}{2} m\dot{x}^2 + \frac{1}{2} m\dot{y}^2 - mgy \tag{7.20}$$

We find the equations of motion by using Equation 7.18:

x:

$$\begin{aligned} \frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} &= 0 \\ 0 - \frac{d}{dt} m\dot{x} &= 0 \\ \ddot{x} &= 0 \end{aligned} \tag{7.21}$$

y:

$$\begin{aligned} \frac{\partial L}{\partial y} - \frac{d}{dt}\frac{\partial L}{\partial \dot{y}} &= 0 \\ -mg - \frac{d}{dt}(m\dot{y}) &= 0 \\ \ddot{y} &= -g \end{aligned} \tag{7.22}$$

By using the initial conditions, Equations 7.21 and 7.22 can be integrated to determine the appropriate equations of motion. In polar coordinates, we have

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

$$U = mgr \sin\theta$$

where U= 0 at $\theta = 0$.

$$L = T - U = \frac{1}{2}m\dot{r}^2 + \frac{1}{2}mr^2\dot{\theta}^2 - mgr\sin\theta \tag{7.23}$$

r:

$$\frac{\partial L}{\partial r} - \frac{d}{dt}\frac{\partial L}{\partial \dot{r}} = 0$$

$$mr\dot{\theta}^2 - mg\sin\theta - \frac{d}{dt}(m\dot{r}) = 0 \tag{7.24}$$

$$r\dot{\theta}^2 - g\sin\theta - \ddot{r} = 0$$

θ:

$$\frac{\partial L}{\partial \theta} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} = 0$$

$$-mgr\cos\theta - \frac{d}{dt}(mr^2\dot{\theta}) = 0 \tag{7.25}$$

$$-gr\cos\theta - 2r\dot{r}\dot{\theta} - r^2\ddot{\theta} = 0$$

The equations of motion expressed by Equations 7.21 and 7.22 are clearly simpler than those of Equations 7.24 and 7.25. We should choose Cartesian coordinates as the generalized coordinates to solve this problem. The key in recognizing this was that the potential energy of the system only depended on the y coordinate. In polar coordinates, the potential energy depended on both r and θ .

EXAMPLE 7.4

A particle of mass m is constrained to move on the inside surface of a smooth cone of half-angle α (see Figure 7-2).

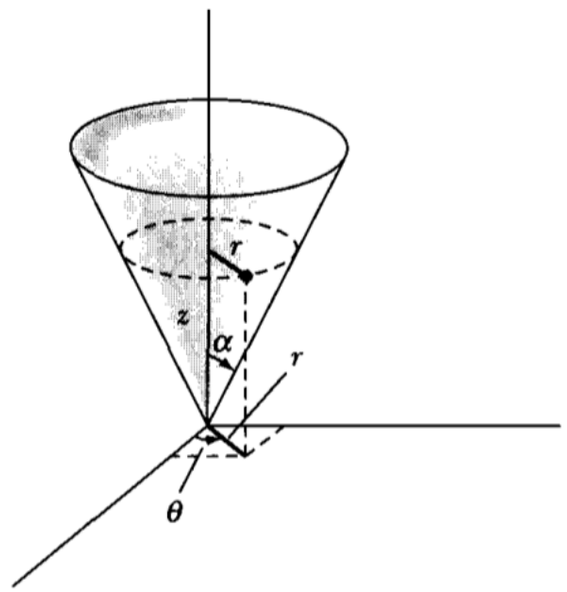


FIGURE 7-2 Example 7.4. A smooth cone of half-angle α. We choose r, θ, and z as the generalized coordinates.

The particle is subject to a gravitational force. Determine a set of generalized coordinates and determine the constraints. Find Lagrange's equations of motion, Equation 7.18

Solution. Let the axis of the cone correspond to the z-axis and let the apex of the cone be located at the origin. Since the problem possesses cylindrical symmetry, we choose r, θ , and z as the generalized coordinates. We have, however, the equation of constraint

$$z = r \cot \alpha \tag{7.26}$$

so there are only two degrees of freedom for the system, and therefore only two proper generalized coordinates. We may use Equation 7.26 to eliminate either the coordinate z or r; we choose to do the former. Then the square of the velocity is

$$\begin{aligned} v^2 &= \dot{r}^2 + r^2\dot{\theta}^2 + \dot{z}^2 \\ &= \dot{r}^2 + r^2\dot{\theta}^2 + \dot{r}^2 \cot^2 \alpha \\ &= \dot{r}^2 \csc^2 \alpha + r^2\dot{\theta}^2 \end{aligned} \tag{7.27}$$

The potential energy (if we choose U = 0 at z = 0) is

$$U = mgz = mgr \cot \alpha$$

so the Langrangian is

$$L = \frac{1}{2}m\,(\dot{r}^2 \csc^2\alpha + r^2\dot{\theta}^2) - mgr \cot \alpha \tag{7.28}$$

We note first that L does not explicitly contain θ . Therefore $\partial L/\partial\theta = 0$, and the Lagrange equation for the coordinate θ is

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

Hence

$$\frac{\partial L}{\partial \dot{\theta}} = mr^2\dot{\theta} = \text{constant} \tag{7.29}$$

But $mr^2\dot{\theta} = mr^2\omega$) is just the angular momentum about the z-axis. Therefore, Equation 7.29 expresses the conservation of angular momentum about the axis of symmetry of the system.

The Lagrange equation for r is

$$\frac{\partial L}{\partial r} - \frac{d}{dt}\frac{\partial L}{\partial \dot{r}} = 0 \tag{7.30}$$

Calculating the derivatives. we find

$$\ddot{r} - r\dot{\theta}^2 \sin^2\alpha + g \sin\alpha \cos\alpha = 0 \tag{7.31}$$

which is the equation of motion for the coordinate r.

We shall return to this example in Section 8.10 and examine the motion in more detail.

EXAMPLE 7.5

The point of support of a simple pendulum of length b moves on a massless rim of radius a rotating with constant angular velocity ω . Obtain the expression for the Cartesian components of the velocity and acceleration of the mass m. Obtain also the angular acceleration for the angle θ shown in Figure 7-3.

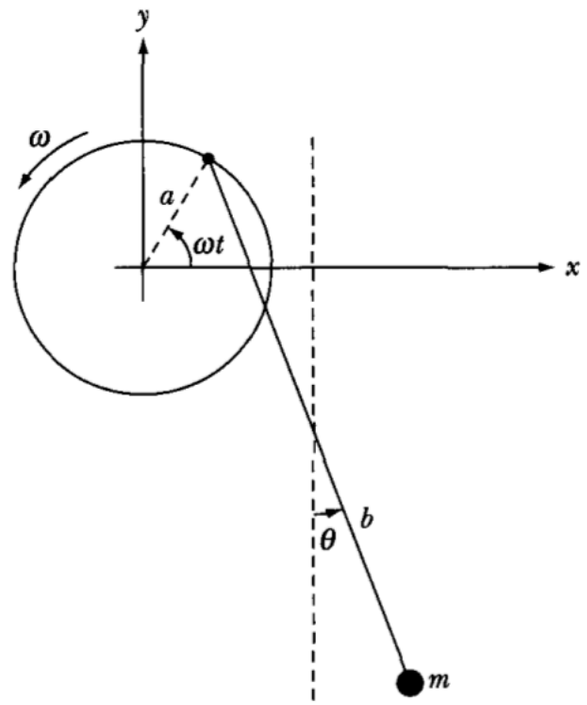


FIGURE 7-3 Example 7.5. A simple pendulum is attached to a rotating rim.

Solution. We choose the origin of our coordinate system to be at the center of the rotating rim. The Cartesian components of mass m become

$$\left.\begin{aligned} x &= a\cos\omega t + b\sin\theta \\ y &= a\sin\omega t - b\cos\theta \end{aligned}\right\} \tag{7.32}$$

The velocities are

$$\left.\begin{aligned} \dot{x} &= -a\omega\sin\omega t + b\dot{\theta}\cos\theta \\ \dot{y} &= a\omega\cos\omega t + b\dot{\theta}\sin\theta \end{aligned}\right\} \tag{7.33}$$

Taking the time derivative once again gives the acceleration:

$$\ddot{x} = -a\omega^2 \cos \omega t + b(\ddot{\theta} \cos \theta - \dot{\theta}^2 \sin \theta)$$

$$\ddot{y} = -a\omega^2 \sin \omega t + b(\ddot{\theta} \sin \theta + \dot{\theta}^2 \cos \theta)$$

It should now be clear that the single generalized coordinate is θ . The kinetic and potential energies are

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

$$U = mgy$$

where U = 0 at y = 0. The Lagrangian is

$$L = T - U = \frac{m}{2}[a^2\omega^2 + b^2\dot{\theta}^2 + 2b\dot{\theta}a\omega \sin(\theta - \omega t)]$$
$$- mg(a \sin \omega t - b \cos \theta) \tag{7.34}$$

The derivatives for the Lagrange equation of motion for θ are

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} = mb^2\ddot{\theta} + mba\omega(\dot{\theta} - \omega) \cos(\theta - \omega t)$$

$$\frac{\partial L}{\partial \theta} = mb\dot{\theta}a\omega \cos(\theta - \omega t) - mgb \sin \theta$$

which results in the equation of motion (after solving for $\ddot{\theta}$)

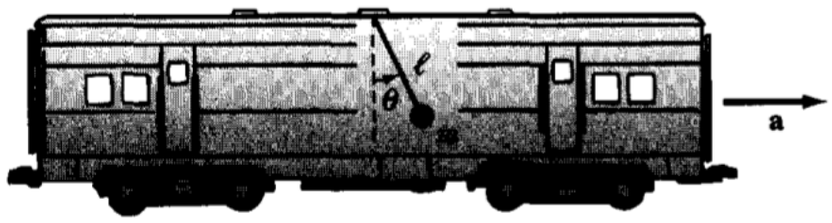
$$\ddot{\theta} = \frac{\omega^2 a}{b} \cos(\theta - \omega t) - \frac{g}{b} \sin \theta \tag{7.35}$$

Notice that this result reduces to the well-known equation of motion for a simple pendulum if $\omega = 0$.

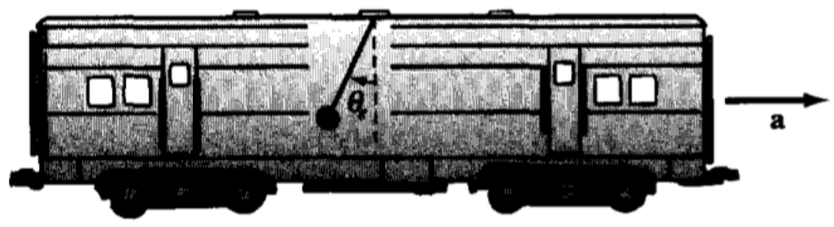
EXAMPLE 7.6

Find the frequency of small oscillations of a simple pendulum placed in a railroad car that has a constant acceleration a in the x-direction.

Solution. A schematic diagram is shown in Figure 7-4a for the pendulum of length ℓ , mass m, and displacement angle θ .



(a)



(b)

FIGURE 7-4 Example 7.6. (a) A simple pendulum swings in an accelerating railroad car. (b) The angle θ_e is the equilibrium angle due to the car's acceleration a and acceleration of gravity g.

We choose a fixed cartesian coordinate system with x = 0 and $\dot{x}$ = v_0 at t = 0. The position and velocity of m become

$$x = v_0 t + \frac{1}{2} a t^2 + \ell \sin\theta$$

$$y = -\ell \cos\theta$$

$$\dot{x} = v_0 + at + \ell\dot{\theta}\cos\theta$$

$$\dot{y} = \ell\dot{\theta}\sin\theta$$

The kinetic and potential energies are

$$T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) \quad U = -mg\ell\cos\theta$$

and the Lagrangian is

$$L = T - U = \frac{1}{2}m(v_0 + at + \ell\dot{\theta}\cos\theta)^2 + \frac{1}{2}m(\ell\dot{\theta}\sin\theta)^2 + mg\ell\cos\theta$$

The angle θ is the only generalized coordinate, and after taking the derivatives for Lagrange's equations and suitable collection of terms, the equation of motion becomes

$$\ddot{\theta} = -\frac{g}{\ell}\sin\theta - \frac{a}{\ell}\cos\theta \tag{7.36}$$

We determine the equilibrium angle $\theta = \theta_e$ setting $\ddot{\theta} = 0$,

$$0 = g\sin\theta_e + a\cos\theta_e \tag{7.37}$$

The equilibrium angle θ_e, shown in Figure 7-4b, is obtained by

$$\tan\theta_e = -\frac{a}{g} \tag{7.38}$$

Because the oscillations are small and are about the equilibrium angle, let $\theta = \theta_e + \eta$ where η is a small angle.

$$\ddot{\theta} = \ddot{\eta} = -\frac{g}{\ell}\sin(\theta_e + \eta) - \frac{a}{\ell}\cos(\theta_e + \eta) \tag{7.39}$$

We expand the sine and cosine terms and use the small angle approximation for $\sin\eta$ and $\cos\eta$, keeping only the first terms in the Taylor series expansions.

$$\ddot{\eta} = -\frac{g}{\ell}(\sin\theta_e \cos\eta + \cos\theta_e \sin\eta) - \frac{a}{\ell}(\cos\theta_e \cos\eta - \sin\theta_e \sin\eta)$$

$$= -\frac{g}{\ell}(\sin\theta_e + \eta\cos\theta_e) - \frac{a}{\ell}(\cos\theta_e - \eta\sin\theta_e)$$

$$= -\frac{1}{\ell}[(g\sin\theta_e + a\cos\theta_e) + \eta(g\cos\theta_e - a\sin\theta_e)]$$

The first term in the brackets is zero because of Equation 7.37, which leaves

$$\ddot{\eta} = -\frac{1}{\ell}(g\cos\theta_e - a\sin\theta_e)\eta \tag{7.40}$$

We use Equation 7.38 to determine $\sin\theta_e$ and $\cos\theta_e$ and after a little manipulation, Equation 7.40 becomes

$$\ddot{\eta} = -\frac{\sqrt{a^2 + g^2}}{\ell}\eta \tag{7.41}$$

Because this equation now represents simple harmonic motion, the frequency ω is determined to be

$$\omega^2 = \frac{\sqrt{a^2 + g^2}}{\ell} \tag{7.42}$$

This result seems plausible, because $\omega \to \sqrt{g/\ell}$ for a = 0 when the railroad car is at rest.

EXAMPLE 7.7

A bead slides along a smooth wire bent in the shape of a parabola $z = cr^2$ (Figure 7-5).

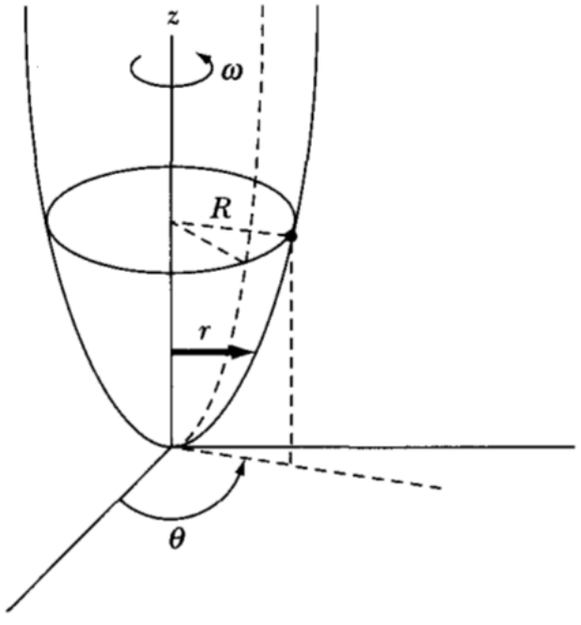


FIGURE 7-5 Example 7.7. A bead slides along a smooth wire that rotates about the z-axis.

The bead rotates in a circle of radius R when the wire is rotating about its vertical symmetry axis with angular velocity ω . Find the value of c.

Solution. Because the problem has cylindrical symmetry, we choose r, θ , and z as the generalized coordinates. The kinetic energy of the bead is

$$T = \frac{m}{2}[\dot{r}^2 + \dot{z}^2 + (r\dot{\theta}^2)] \tag{7.43}$$

If we choose U = 0 at z = 0, the potential energy term is

$$U = mgz \tag{7.44}$$

But r, z, and θ are not independent. The equation of constraint for the parabola is

$$z = cr^2 \tag{7.45}$$

$$\dot{z} = 2c\dot{r}r \tag{7.46}$$

We also have an explicit time dependence of the angular rotation

$$\theta = \omega t$$
$$\dot{\theta} = \omega \tag{7.47}$$

We can now construct the Lagrangian as being dependent only on r, because there is no direct θ dependence.

$$L = T - U$$
$$= \frac{m}{2}(\dot{r}^2 + 4c^2r^2\dot{r}^2 + r^2\omega^2) - mgcr^2 \tag{7.48}$$

The problem stated that the bead moved in a circle of radius R. The reader might be tempted at this point to let r = R = const, and $\dot{r} = 0$. It would be a mistake to do this now in the Lagrangian. First, we should find the equation of motion for the variable r and then let r = R as a condition of the particular motion. This determines the particular value of c needed for r = R.

$$\frac{\partial L}{\partial \dot{r}} = \frac{m}{2}(2\dot{r} + 8c^2r^2\dot{r})$$

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{r}} = \frac{m}{2}(2\ddot{r} + 16c^2r\dot{r}^2 + 8c^2r^2\ddot{r})$$

$$\frac{\partial L}{\partial r} = m(4c^2r\dot{r}^2 + r\omega^2 - 2gcr)$$

Lagrange's equation of motion becomes

$$\ddot{r}(1 + 4c^2r^2) + \dot{r}^2(4c^2r) + r(2gc - \omega^2) = 0 \tag{7.49}$$

which is a complicated result.

If, however, the bead rotates with r = R= constant, then $\dot{r} = \ddot{r} = 0$, and Equation 7.49 becomes

$$R(2gc - \omega^2) = 0$$

and

$$c = \frac{\omega^2}{2g} \tag{7.50}$$

is the result we wanted.

EXAMPLE 7.8

Consider the double pulley system shown in Figure 7-6.

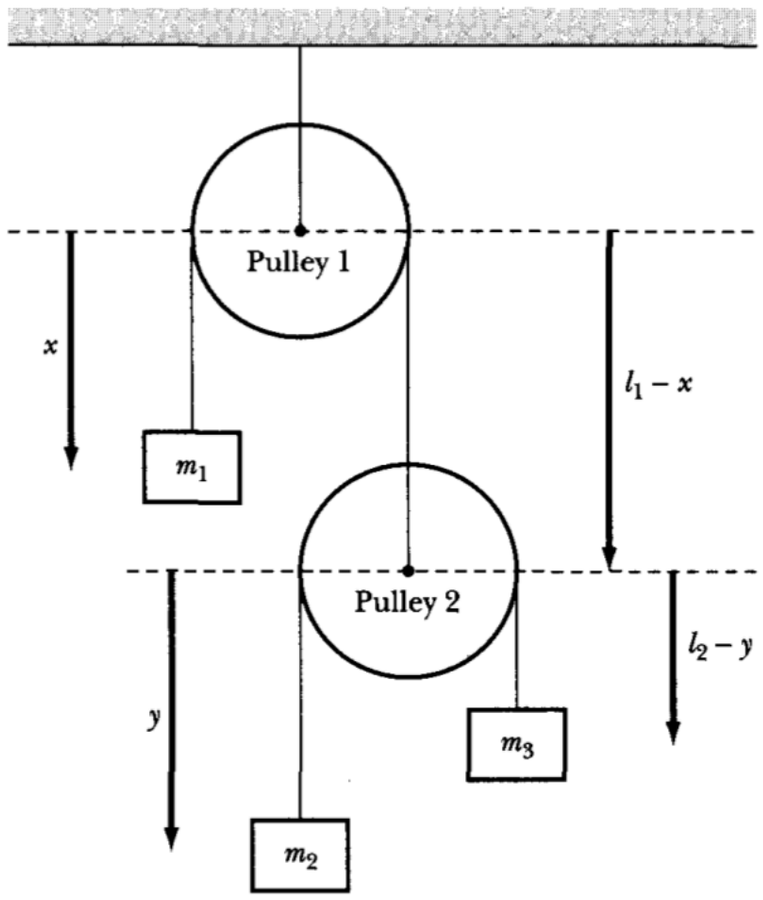


FIGURE 7-6 Example 7.8. The double pulley system.

Use the coordinates indicated, and determine the equations of motion.

Solution. Consider the pulleys to be massless, and let l_1 and l_2 be the lengths of rope hanging freely from each of the two pulleys. The distances x and y are measured from the center of the two pulleys.

m_1:

$$v_1 = \dot{x} \tag{7.51}$$

m_2:

$$v_2 = \frac{d}{dt}(l_1 - x + y) = -\dot{x} + \dot{y} \tag{7.52}$$

m_3:

$$v_3 = \frac{d}{dt}(l_1 - x + l_2 - y) = -\dot{x} - \dot{y} \tag{7.53}$$

$$T = \frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2 + \frac{1}{2}m_3 v_3^2$$

$$= \frac{1}{2}m_1\dot{x}^2 + \frac{1}{2}m_2(\dot{y} - \dot{x})^2 + \frac{1}{2}m_3(-\dot{x} - \dot{y})^2 \tag{7.54}$$

Let the potential energy U = 0 at x = 0.

$$U = U_1 + U_2 + U_3$$

$$= -m_1 gx - m_2 g(l_1 - x + y) - m_3 g(l_1 - x + l_2 - y) \tag{7.55}$$

Because T and U have been determined, the equations of motion can be obtained using Equation 7.18. The results are

$$m_1\ddot{x} + m_2(\ddot{x} - \ddot{y}) + m_3(\ddot{x} + \ddot{y}) = (m_1 - m_2 - m_3)g \tag{7.56}$$

$$-m_2(\ddot{x} - \ddot{y}) + m_3(\ddot{x} + \ddot{y}) = (m_2 - m_3)g \tag{7.57}$$

Equations 7.56 and 7.57 can be solved for $\ddot{x}$ and $\ddot{y}$.

Examples 7.2-7.8 indicate the ease and usefulness of using Lagrange's equations. It has been said, probably unfairly, that Lagrangian techniques are simply recipes to follow. The argument is that we lose track of the "physics" by their use. Lagrangian methods, on the contrary, are extremely powerful and allow us to solve problems that otherwise would lead to severe complications using Newtonian methods. Simple problems can perhaps be solved just as easily using Newtonian methods, but the Lagrangian techniques can be used to attack a wide range of complex physical situations (including those occurring in quantum mechanics).

7.5 Lagrange's Equations with Undetermined Multipliers

Constraints that can be expressed as algebraic relations among the coordinates are holonomic constraints. If a system is subject only to such constraints, we can always find a proper set of generalized coordinates in terms of which the equations of motion are free from explicit reference to the constraints.

Any constraints that must be expressed in terms of the velocities of the particles in the system are of the form

$$f(x_{\alpha,i}, \dot{x}_{\alpha,i}, t) = 0 \tag{7.58}$$

and constitute nonholonomic constraints unless the equations can be integrated to yield relations among the coordinates.

Consider a constraint relation of the form

$$\sum_i A_i \dot{x}_i + B = 0, \quad i = 1, 2, 3 \tag{7.59}$$

In general, this equation is nonintegrable, and therefore the constraint is nonholonomic. But if A_i and B have the forms

$$A_i = \frac{\partial f}{\partial x_i}, \quad B = \frac{\partial f}{\partial t}, \quad f = f(x_i, t) \tag{7.60}$$

then Equation 7.59 may be written as

$$\sum_i \frac{\partial f}{\partial x_i}\frac{dx_i}{dt} + \frac{\partial f}{\partial t} = 0 \tag{7.61}$$

But this is just

$$\frac{df}{dt} = 0$$

which can be integrated to yield

$$f(x_i, t) - \text{constant} = 0 \tag{7.62}$$

so the constraint is actually holonomic.

From the preceding discussion, we conclude that constraints expressible in differential form as

$$\sum_j \frac{\partial f_k}{\partial q_j} dq_j + \frac{\partial f_k}{\partial t} dt = 0 \tag{7.63}$$

are equivalent to those having the form of Equation 7.9.

If the constraint relations for a problem are given in differential form rather than as algebraic expressions, we can incorporate them directly into Lagrange's equations by using the Lagrange undetermined multipliers (see Section 6.6) without first performing the integrations; that is, for constraints expressible as in Equation 6.71,

$$\sum_j \frac{\partial f_k}{\partial q_j} dq_j = 0 \quad \begin{cases} j = 1, 2, \ldots, s \\ k = 1, 2, \ldots, m \end{cases} \tag{7.64}$$

the Lagrange equations (Equation 6.69) are

$$\boxed{\frac{\partial L}{\partial q_j} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j} + \sum_k \lambda_k(t) \frac{\partial f_k}{\partial q_j} = 0} \tag{7.65}$$

In fact, because the variation process involved in Hamilton's Principle holds the time constant at the endpoints, we could add to Equation 7.64 a term ($\partial f_k / \partial t$)dt without affecting the equations of motion.

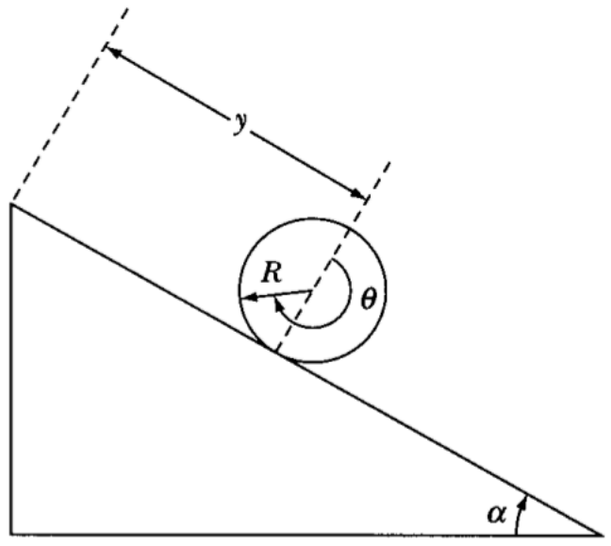
Thus constraints expressed by Equation 7.63 also lead to the Lagrange equations given in Equation 7.65.

The great advantage of the Lagrangian formulation of mechanics is that the explicit inclusion of the forces of constraint is not necessary; that is, the emphasis is placed on the dynamics of the system rather than the calculation of the forces acting on each component of the system. In certain instances, however, it might be desirable to know the forces of constraint. For example, from an engineering standpoint, it would be useful to know the constraint forces for design purposes. It is therefore worth pointing out that in Lagrange's equations expressed as in Equation 7.65, **the undetermined multipliers $\lambda_j(t)$ are closely related to the forces of constraint**. The generalized forces of constraint Q_j are given by

$$Q_j = \sum_k \lambda_k \frac{\partial f_k}{\partial q_j} \tag{7.66}$$

EXAMPLE 7.9

Let us consider again the case of the disk rolling down an inclined plane (see Example 6.5 and Figure 6-7).



Find the equations of motion, the force of constraint, and the angular acceleration.

Solution. The kinetic energy may be separated into translational and rotational terms (We anticipate here a well-known result from rigid-body dynamics discussed in Chapter 11.)

$$T = \frac{1}{2} M\dot{y}^2 + \frac{1}{2} I\dot{\theta}^2$$

$$= \frac{1}{2} M\dot{y}^2 + \frac{1}{4} MR^2 \dot{\theta}^2$$

where M is the mass of the disk and R is the radius; $I = \frac{1}{2}MR^2$ is the moment of inertia of the disk about a central axis. The potential energy is

$$U = Mg(l - y) \sin \alpha \tag{7.67}$$

where l is the length of the inclined surface of the plane and where the disk is assumed to have zero potential energy at the bottom of the plane. The Lagrangian is therefore

$$L = T - U$$

$$= \frac{1}{2}M\dot{y}^2 + \frac{1}{4}MR^2\dot{\theta}^2 + Mg(y - l) \sin \alpha \tag{7.68}$$

The equation of constraint is

$$f(y, \theta) = y - R\theta = 0 \tag{7.69}$$

The system has only one degree of freedom if we insist that the rolling takes place without slipping. We may therefore choose either y or θ as the proper coordinate and use Equation 7.69 to eliminate the other. Alternatively, we may continue to consider *both* y and θ as generalized coordinates and use the method of undetermined multipliers. The Lagrange equations in this case are

$$\left.\begin{aligned} \frac{\partial L}{\partial y} - \frac{d}{dt}\frac{\partial L}{\partial \dot{y}} + \lambda\frac{\partial f}{\partial y} = 0 \\ \frac{\partial L}{\partial \theta} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} + \lambda\frac{\partial f}{\partial \theta} = 0 \end{aligned}\right\} \tag{7.70}$$

Performing the differentiations, we obtain, for the equations of motion,

$$Mg \sin \alpha - M\ddot{y} + \lambda = 0 \tag{7.71a}$$

$$-\frac{1}{2} MR^2\ddot{\theta} - \lambda R = 0 \tag{7.71b}$$

Also, from the constraint equation, we have

$$y = R\theta \tag{7.72}$$

These equations (Equations 7.71 and 7.72) constitute a soluble system for the three unknowns y, θ , λ .
Differentiating the equation of constraint (Equation 7.72), we obtain

$$\ddot{\theta} = \frac{\ddot{y}}{R} \tag{7.73}$$

Combining Equations 7.71b and 7.73, we find

$$\lambda = -\frac{1}{2} M\ddot{y} \tag{7.74}$$

and then using this expression in Equation 7.71a there results

$$\ddot{y} = \frac{2g \sin \alpha}{3} \tag{7.75}$$

with

$$\lambda = -\frac{Mg \sin \alpha}{3} \tag{7.76}$$

so that Equation 7.71b yields

$$\ddot{\theta} = \frac{2g \sin \alpha}{3R} \tag{7.77}$$

Thus, we have three equations for the quantities $\ddot{y}$, $\ddot{\theta}$, and λ that can be immediately integrated.

We note that if the disk were to slide without friction down the plane, we would have $\ddot{y} = g \sin \alpha$. Therefore, the rolling constraint reduces the acceleration to $\frac{2}{3}$ of the value of frictionless sliding. The magnitude of the force of friction producing the constraint is just λ - that is, $(Mg/3) \sin \alpha$.

The generalized forces of constraint. Equation 7.66, are

$$Q_y = \lambda \frac{\partial f}{\partial y} = \lambda = -\frac{Mg \sin \alpha}{3}$$

$$Q_\theta = \lambda \frac{\partial f}{\partial \theta} = -\lambda R = \frac{MgR \sin \alpha}{3}$$

Note that Q_y and Q_β are a force and a torque, respectively, and they are the generalized forces of constraint required to keep the disk rolling down the plane without slipping.

Note that we may eliminate $\dot{\theta}$ from the Lagrangian by substituting $\dot{\theta} = \dot{y}/R$ from the equation of constraint:

$$L = \frac{3}{4} M\dot{y}^2 + Mg(y - l) \sin \alpha \tag{7.78}$$

The Lagrangian is then expressed in terms of only one proper coordinate, and the single equation of motion is immediately obtained from Equation 7.18:

$$Mg \sin \alpha - \frac{3}{2} M\ddot{y} = 0 \tag{7.79}$$

which is the same as Equation 7.75. Although this procedure is simpler, it cannot be used to obtciin the force of constraint.

EXAMPLE 7.10

A particle of mass m starts at rest on top of a smooth fixed hemisphere of radius a. Find the force of constraint, and determine the angle at which the particle leaves the hemisphere.

Solution. See Figure 7-7.

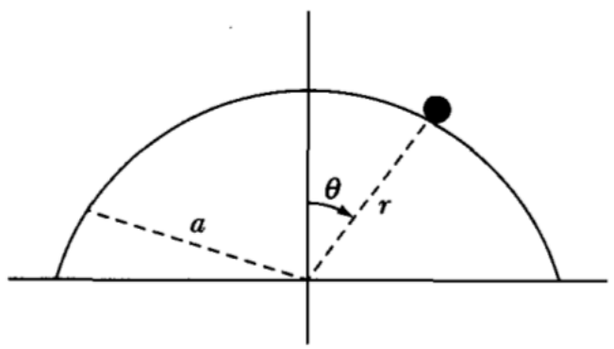


FIGURE 7-7 Example 7.10. A particle of mass m moves on the surface of a fixed smooth hemisphere.

Because we are considering the possibility of the particle leaving the hemisphere, we choose the generalized coordinates to be r and θ . The constraint equation is

$$f(r, \theta) = r - a = 0 \tag{7.80}$$

The Lagrangian is determined from the kinetic and potential energies:

$$T = \frac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2)$$

$$U = mgr\cos\theta$$

$$L = T - U$$

$$L = \frac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2) - mgr\cos\theta \tag{7.81}$$

where the potential energy is zero at the bottom of the hemisphere. The Lagrange equations, Equation 7.65, are

$$\frac{\partial L}{\partial r} - \frac{d}{dt}\frac{\partial L}{\partial \dot{r}} + \lambda\frac{\partial f}{\partial r} = 0 \tag{7.82}$$

$$\frac{\partial L}{\partial \theta} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} + \lambda\frac{\partial f}{\partial \theta} = 0 \tag{7.83}$$

Performing the differentiations on Equation 7.80 gives

$$\frac{\partial f}{\partial r} = 1, \quad \frac{\partial f}{\partial \theta} = 0 \tag{7.84}$$

Equations 7.82 and 7.83 become

$$mr\dot{\theta}^2 - mg\cos\theta - m\ddot{r} + \lambda = 0 \tag{7.85}$$

$$mgr\sin\theta - mr^2\ddot{\theta} - 2mr\dot{r}\dot{\theta} = 0 \tag{7.86}$$

Next, we apply the constraint r = a to these equations of motion:

$$r = a, \quad \dot{r} = 0 = \ddot{r}$$

Equations 7.85 and 7.86 then become

$$ma\dot{\theta}^2 - mg\cos\theta + \lambda = 0 \tag{7.87}$$

$$mga\sin\theta - ma^2\ddot{\theta} = 0 \tag{7.88}$$

From Equation 7.88, we have

$$\ddot{\theta} = \frac{g}{a}\sin\theta \tag{7.89}$$

We can integrate Equation 7.89 to determine $\dot{\theta}^2$.

$$\ddot{\theta} = \frac{d}{dt}\frac{d\theta}{dt} = \frac{d\dot{\theta}}{dt} = \frac{d\dot{\theta}}{d\theta}\frac{d\theta}{dt} = \dot{\theta}\frac{d\dot{\theta}}{d\theta} \tag{7.90}$$

We integrate Equation 7.89,

$$\int \dot{\theta}\, d\dot{\theta} = \frac{g}{a}\int \sin\theta\, d\theta \tag{7.91}$$

which results in

$$\frac{\dot{\theta}^2}{2} = \frac{-g}{a}\cos\theta + \frac{g}{a} \tag{7.92}$$

where the integration constant is g/a, because $\dot{\theta} = 0$ at t = 0 when $\theta = 0$. Substituting $\dot{\theta}^2$ from Equation 7.92 into Equation 7.87 gives, after solving for ,

$$\lambda = mg(3\cos\theta - 2) \tag{7.93}$$

which is the force of constraint. The particle falls off the hemisphere at angle θ_0 when $\lambda = 0$.

$$\lambda = 0 = mg(3\cos\theta_0 - 2) \tag{7.94}$$

$$\theta_0 = \cos^{-1}\left(\frac{2}{3}\right) \tag{7.95}$$

As a quick check, notice that the constraint force is $\lambda = $ mg at $\theta = 0$ when the particle is perched on top of the hemisphere.

The usefulness of the method of undetermined multipliers is twofold:

1. The Lagrange multipliers are closely related to the forces of constraint that are often needed.
2. When a proper set of generalized coordinates is not desired or too difficult to obtain, the method may be used to increase the number of generalized coordinates by including constraint relations between the coordinates.

7.6 Equivalence of Lagrange's and Newton's Equations

As we have emphasized from the outset, the Lagrangian and Newtonian formulations of mechanics are equivalent: The viewpoint is different, but the content is the same. We now explicitly demonstrate this equivalence by showing that the two sets of equations of motion are in fact the same.

In Equation 7.18, let us choose the generalized coordinates to be the rectangular coordinates. Lagrange's equations (for a single particle) then become

$$\frac{\partial L}{\partial x_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}_i} = 0, \qquad i = 1, 2, 3 \tag{7.96}$$

or

$$\frac{\partial (T - U)}{\partial x_i} - \frac{d}{dt}\frac{\partial (T - U)}{\partial \dot{x}_i} = 0$$

But in rectangular coordinates and for a conservative system, we have T = T($\dot{x}_i$) and U= U(x_i), so

$$\frac{\partial T}{\partial x_i} = 0 \quad \text{and} \quad \frac{\partial U}{\partial \dot{x}_i} = 0$$

Lagrange's equations therefore become

$$-\frac{\partial U}{\partial x_i} = \frac{d}{dt}\frac{\partial T}{\partial \dot{x}_i} \tag{7.97}$$

We also have (for a conservative system)

$$-\frac{\partial U}{\partial x_i} = F_i$$

and

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{x}_i} = \frac{d}{dt}\frac{\partial}{\partial \dot{x}_i}\left(\sum_{j=1}^{3}\frac{1}{2}m\dot{x}_j^2\right) = \frac{d}{dt}(m\dot{x}_i) = \dot{p}_i$$

so Equation 7.97 yields the Newtonian equations, as required:

$$F_i = \dot{p}_i \tag{7.98}$$

Thus, the Lagrangian and Newtonian equations are identical if the generalized coordinates are the rectangular coordinates.

Now let us derive Lagrange's equations of motion using Newtonian concepts. Consider only a single particle for simplicity. We need to transform from the x_i-coordinates to the generalized coordinates q_j. From Equation 7.5, we have

$$x_i = x_i(q_j, t) \tag{7.99}$$

$$\dot{x}_i = \sum_j \frac{\partial x_i}{\partial q_j} \dot{q}_j + \frac{\partial x_i}{\partial t} \tag{7.100}$$

and

$$\frac{\partial \dot{x}_i}{\partial \dot{q}_i} = \frac{\partial x_i}{\partial q_i} \tag{7.101}$$

A generalized momentum p_j associated with q_j is easily determined by

$$p_j = \frac{\partial T}{\partial \dot{q}_j} \tag{7.102}$$

For example, for a particle moving in plane polar coordinates, $T = (\dot{r}^2 + r^2\dot{\theta}^2)\, m/2$, we have $p_r = m\dot{r}$ for coordinate r and $p_\theta = mr^2\dot{\theta}$ for coordinate θ. Obviously p_r is a linear momentum and p_θ is an angular momentum, so our generalized momentum definition seems consistent with Newtonian concepts.

We can determine a generalized force by considering the *virtual* work δW done by a varied path δx_i, as described in Section 6.7.

$$\delta W = \sum_i F_i \delta x_i = \sum_{i,j} F_i \frac{\partial x_i}{\partial q_j} \delta q_j \tag{7.103}$$

$$\equiv \sum_j Q_j \delta q_j \tag{7.104}$$

so that the generalized force Q_j associated with q_j is

$$Q_j = \sum_i F_i \frac{\partial x_i}{\partial q_j} \tag{7.105}$$

Just as work is always energy, so is the product of Qq. If q is length, Q is force; if q is an angle, Q is torque. For a conservative system, Q_j is derivable from the potential energy:

$$Q_j = -\frac{\partial U}{\partial q_j} \tag{7.106}$$

Now we are ready to obtain Lagrange's equations:

$$p_j = \frac{\partial T}{\partial \dot{q}_j} = \frac{\partial}{\partial \dot{q}_j}\left(\sum_i \frac{1}{2} m \dot{x}_i^2\right)$$

$$= \sum_i m \dot{x}_i \frac{\partial \dot{x}_i}{\partial \dot{q}_j}$$

$$p_j = \sum_i m \dot{x}_i \frac{\partial x_i}{\partial q_j} \tag{7.107}$$

where we use Equation 7.101 for the last step. Taking the time derivative of Equation 7.107 gives

$$\dot{p}_j = \sum_i \left(m \ddot{x}_i \frac{\partial x_i}{\partial q_j} + m \dot{x}_i \frac{d}{dt} \frac{\partial x_i}{\partial q_j}\right) \tag{7.108}$$

Expanding the last term gives

$$\frac{d}{dt}\frac{\partial x_i}{\partial q_j} = \sum_k \frac{\partial^2 x_i}{\partial q_k \partial q_j}\dot{q}_k + \frac{\partial^2 x_i}{\partial q_j \partial t}$$

and Equation 7.108 becomes

$$\dot{p}_j = \sum_i m\ddot{x}_i \frac{\partial x_i}{\partial q_j} + \sum_{i,k} m\dot{x}_i \frac{\partial^2 x_i}{\partial q_k \partial q_j}\dot{q}_k + \sum_i m\dot{x}_i \frac{\partial^2 x_i}{\partial q_j \partial t} \tag{7.109}$$

The first term on the right side of Equation 7.109 is just Q_j (Fi = m$\ddot{x}_i$ and Equation 7.105). The sum of the other two terms is $\partial T/\partial q_j$.

$$\frac{\partial T}{\partial q_j} = \sum_i m\dot{x}_i \frac{\partial \dot{x}_i}{\partial q_j}$$

$$= \sum_i m\dot{x}_i \frac{\partial}{\partial q_j}\left(\sum_k \frac{\partial x_i}{\partial q_k}\dot{q}_k + \frac{\partial x_i}{\partial t}\right) \tag{7.110}$$

where we have used T = $\sum_i 1/2 m\dot{x}_i^2$ and Equation 7.100.

Equation 7.109 can now be written as

$$\dot{p}_j = Q_j + \frac{\partial T}{\partial q_j} \tag{7.111}$$

or, using Equations 7.102 and 7.106,

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_j}\right) - \frac{\partial T}{\partial q_j} = Q_j = -\frac{\partial U}{\partial q_j} \tag{7.112}$$

Because U does not depend on the generalized velocities q_j, Equation 7.112 can be written

$$\frac{d}{dt}\left[\frac{\partial (T - U)}{\partial \dot{q}_j}\right] - \frac{\partial (T - U)}{\partial q_j} = 0 \tag{7.113}$$

and using L= T— U,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right) - \frac{\partial L}{\partial q_j} = 0 \tag{7.114}$$

which are Lagrange's equations of motion.

7.7 Essence of Lagrangian Dynamics

In the preceding sections, we made several general and important statements concerning the Lagrange formulation of mechanics. Before proceeding further, we should summarize these points to emphasize the differences between the Lagrange and Newtonian viewpoints.

Historically, the Lagrange equations of motion expressed in generalized coordinates were derived before the statement of Hamilton's Principle. We elected to deduce Lagrange's equations by postulating Hamilton's Principle because this is the most straightforward approach and is also the formal method for unifying classical dynamics.

First, we must reiterate that Lagrangian dynamics does not constitute a new theory in any sense of the word. The results of a Lagrangian analysis or a Newtonian analysis must be the same for any given mechanical system. The only difference is the method used to obtain these results.

Whereas the Newtonian approach emphasizes an outside agency acting on a body (the force), the Lagrangian method deals only with quantities associated with the body (the kinetic and potential energies). In fact, nowhere in the Lagrangian formulation does the concept of force enter. This is a particularly important property—and for a variety of reasons. First, because energy is a scalar quantity, the Lagrangian function for a system is invariant to coordinate transformations. Indeed, such transformations are not restricted to be between various orthogonal coordinate systems in ordinary space; they may also be transformations between ordinary coordinates and generalized coordinates.

Thus, it is possible to pass from ordinary space (in which the equations of motion may be quite complicated) to a configuration space that can be chosen to yield maximum simplification for a particular problem. We are accustomed to thinking of mechanical systems in terms of vector quantities such as force, velocity, angular momentum, and torque. But in the Lagrangian formulation, the equations of motion are obtained entirely in terms of scalar operations in configuration space.

Another important aspect of the force-versus-energy viewpoint is that in certain situations it may not even be possible to state explicitly all the forces acting on a body (as is sometimes the case for forces of constraint), whereas it is still possible to give expressions for the kinetic and potential energies. It is just this fact that makes Hamilton's Principle useful for quantum-mechanical systems where we normally know the energies but not the forces.

The differential statement of mechanics contained in Newton's equations or the integral statement embodied in Hamilton's Principle (and the resulting Lagrangian equations) have been shown to be entirely equivalent. Hence, no distinction exists between these viewpoints, which are based on the description of *physical effects*. But from a philosophical standpoint, we can make a distinction. In the Newtonian formulation, a certain force on a body produces a definite motion - that is, we always associate a definite *effect* with a certain *cause*. According to Hamilton's Principle, however, the motion of a body results from the attempt of nature to achieve a certain purpose, namely, to minimize the time integral of the difference between the kinetic and potential energies. The operational solving of problems in mechanics does not depend on adopting one or the other of these views. But historically such considerations have had a profound influence on the development of dynamics (as, for example, in Maupertuis's principle, mentioned in Section 7.2).

7.8 A Theorem Concerning the Kinetic Energy

If the kinetic energy is expressed in fixed, rectangular coordinates, the result is a homogeneous quadratic function of $\dot{x}_{\alpha,i}$.

$$T = \frac{1}{2} \sum_{\alpha=1}^{n} \sum_{i=1}^{3} m_\alpha \dot{x}_{\alpha,i}^2 \tag{7.115}$$

We now wish to consider in more detail the dependence of T on the generalized coordinates and velocities. For many particles. Equations 7.99 and 7.100 become

$$x_{\alpha,i} = x_{\alpha,i}(q_j, t), \quad j = 1, 2, \ldots, s \tag{7.116}$$

$$\dot{x}_{\alpha,i} = \sum_{j=1}^{s} \frac{\partial x_{\alpha,i}}{\partial q_j} \dot{q}_j + \frac{\partial x_{\alpha,i}}{\partial t} \tag{7.117}$$

Evaluating the square of $\dot{x}_{\alpha,i}$, we obtain

$$\dot{x}_{\alpha,i}^2 = \sum_{j,k} \frac{\partial x_{\alpha,i}}{\partial q_j} \frac{\partial x_{\alpha,i}}{\partial q_k} \dot{q}_j \dot{q}_k + 2 \sum_{j} \frac{\partial x_{\alpha,i}}{\partial q_j} \frac{\partial x_{\alpha,i}}{\partial t} \dot{q}_j + \left(\frac{\partial x_{\alpha,i}}{\partial t} \right)^2 \tag{7.118}$$

and the kinetic energy becomes

$$T = \sum_{\alpha} \sum_{i,j,k} \frac{1}{2} m_\alpha \frac{\partial x_{\alpha,i}}{\partial q_j} \frac{\partial x_{\alpha,i}}{\partial q_k} \dot{q}_j \dot{q}_k + \sum_{\alpha} \sum_{i,j} m_\alpha \frac{\partial x_{\alpha,i}}{\partial q_j} \frac{\partial x_{\alpha,i}}{\partial t} \dot{q}_j + \sum_{\alpha} \sum_{i} \frac{1}{2} m_\alpha \left(\frac{\partial x_{\alpha,i}}{\partial t} \right)^2 \tag{7.119}$$

Thus, we have the general result

$$T = \sum_{j,k} a_{jk} \dot{q}_j \dot{q}_k + \sum_{j} b_j \dot{q}_j + c \tag{7.120}$$

A particularly important case occurs when the system is *scleronomic*, so that the time does not appear explicitly in the equations of transformation (Equation 7.116); then the partial time derivatives vanish:

$$\frac{\partial x_{\alpha,i}}{\partial t} = 0, \qquad b_j = 0, \qquad c = 0$$

Therefore, under these conditions, the kinetic energy is a *homogeneous quadratic function* of the generalized velocities:

$$\boxed{T = \sum_{j,k} a_{jk}\dot{q}_j\dot{q}_k} \tag{7.121}$$

Next, we differentiate Equation 7.121 with respect to $\dot{q}_i$:

$$\frac{\partial T}{\partial \dot{q}_l} = \sum_k a_{lk}\dot{q}_k + \sum_j a_{jl}\dot{q}_j$$

Multiplying this equation by $\dot{q}_l$ and summing over l , we have

$$\sum_l \dot{q}_l \frac{\partial T}{\partial \dot{q}_l} = \sum_{k,l} a_{lk}\dot{q}_k\dot{q}_l + \sum_{j,l} a_{jl}\dot{q}_j\dot{q}_l$$

In this case, *all* the indices are dummies, so both terms on the right-hand side are identical:

$$\boxed{\sum_l \dot{q}_l \frac{\partial T}{\partial \dot{q}_l} = 2\sum_{j,k} a_{jk}\dot{q}_j\dot{q}_k = 2T} \tag{7.122}$$

This important result is a special case of *Euler's theorem*, which states that if $f(y_k)$ is a homogeneous function of the y_k, that is of degree n, then

$$\sum_k y_k \frac{\partial f}{\partial y_k} = nf \tag{7.123}$$

7.9 Conservation Theorems Revisited

Conservation of Energy

We saw in our previous arguments(Section 2.3) that time is homogeneous within an inertial reference frame. Therefore, the Lagrangian that describes a closed system (i.e., a system not interacting with anything outside the system) cannot depend explicitly on time, that is,

$$\frac{\partial L}{\partial t} = 0 \tag{7.124}$$

so that the total derivative of the Lagrangian becomes

$$\frac{dL}{dt} = \sum_j \frac{\partial L}{\partial q_j}\dot{q}_j + \sum_j \frac{\partial L}{\partial \dot{q}_j}\ddot{q}_j \tag{7.125}$$

where the usual term, $\partial L/\partial t$, does not now appear. But Lagrange's equations are

$$\frac{\partial L}{\partial q_j} = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j} \tag{7.126}$$

Using Equation 7.126 to substitute for $\partial L/\partial q_j$ in Equation 7.125, we have

$$\frac{dL}{dt} = \sum_j \dot{q}_j \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j} + \sum_j \frac{\partial L}{\partial \dot{q}_j}\ddot{q}_j$$

or

$$\frac{dL}{dt} - \sum_j \frac{d}{dt}\left(\dot{q}_j \frac{\partial L}{\partial \dot{q}_j}\right) = 0$$

so that

$$\frac{d}{dt}\left(L - \sum_j \dot{q}_j \frac{\partial L}{\partial \dot{q}_j}\right) = 0 \tag{7.127}$$

The quantity in the parentheses is therefore constant in time; denote this constant by -H:

$$L - \sum_j \dot{q}_j \frac{\partial L}{\partial \dot{q}_j} = -H = \text{constant} \tag{7.128}$$

If the potential energy U does not depend explicitly on the velocities $\dot{x}_{\alpha,i}$ or the time t, then U= U($x_{\alpha,i}$). The relations connecting the rectangular coordinates and the generalized coordinates are of the form $x_{\alpha,i} = x_{\alpha,i}(q_j)$ or $q_j = q_j(x_{\alpha,i})$, where we exclude the possibility of an explicit time dependence in the transformation equations. Therefore, U = U(q_j), and $\partial U/\partial \dot{q}_j = 0$. Thus

$$\frac{\partial L}{\partial \dot{q}_j} = \frac{\partial (T - U)}{\partial \dot{q}_j} = \frac{\partial T}{\partial \dot{q}_j}$$

Equation 7.128 can then be written as

$$(T - U) - \sum_j \dot{q}_j \frac{\partial T}{\partial \dot{q}_j} = -H \tag{7.129}$$

and, using Equation 7.122, we have

$$(T - U) - 2T = -H$$

or

$$T + U = E = H = \text{constant} \tag{7.130}$$

The total energy E is a constant of the motion for this case.

The function H, called the **Hamiltonian** of the system, may be defined as in Equation 7.128 (but see Section 7.10). It is important to note that the Hamiltonian H is equal to the total energy E only if the following conditions are met:

1. The equations of the transformation connecting the rectangular and generalized coordinates (Equation 7.116) must be independent of the time, thus ensuring that the kinetic energy is a homogeneous quadratic function of the $\dot{q}_j$.
2. The potential energy must be velocity independent, thus allowing the elimination of the terms $\partial U/\partial \dot{q}_j$ from the equation for H (Equation 7.129).

The questions "Does H = E for the system?" and "Is energy conserved for the system?," then, pertain to two different aspects of the problem, and each question must be examined separately. We may, for example, have cases in which the Hamiltonian does not equal the total energy, but nevertheless, the energy is conserved. Thus, consider a conservative system, and let the description be made in terms of generalized coordinates in motion with respect to fixed, rectangular axes. The transformation equations then contain the time, and the kinetic energy is *not* a homogeneous quadratic function of the generalized velocities. The choice of a mathematically convenient set of generalized coordinates cannot alter the physical fact that energy is conserved. But in the moving coordinate system, the Hamiltonian is no longer equal to the total energy.

Conservation of Linear Momentum

Because space is homogeneous in an inertial reference frame, the Lagrangian of a closed system is unaffected by a translation of the entire system in space. Consider an infinitesimal translation of every radius vector r,, such that $\mathbf{r}_\alpha \rightarrow \mathbf{r}_\alpha + \delta\mathbf{r}$; this amounts to translating the entire system by $\delta\mathbf{r}$. For simplicity, let us examine a system consisting of only a single particle (by including a summation over α we could consider an n-particle system in an entirely equivalent manner), and let us write the Lagrangian in terms of rectangular coordinates L= L($x_i, \dot{x}_i$). The change in L caused by the infinitesimal displacement $\delta\mathbf{r} = \sum_i \delta x_i \mathbf{e}_i$ is

$$\delta L = \sum_i \frac{\partial L}{\partial x_i}\delta x_i + \sum_i \frac{\partial L}{\partial \dot{x}_i}\delta \dot{x}_i = 0 \qquad (7.131)$$

We consider only a varied *displacement*, so that the δx_i are not explicit or implicit functions of the time. Thus,

$$\delta \dot{x}_i = \delta \frac{dx_i}{dt} = \frac{d}{dt} \delta x_i \equiv 0 \tag{7.132}$$

Therefore, δL becomes

$$\delta L = \sum_i \frac{\partial L}{\partial x_i} \delta x_i = 0 \tag{7.133}$$

Because each of the δx_i is an independent displacement, δL vanishes identically only if each of the partial derivatives of L vanishes:

$$\frac{\partial L}{\partial x_i} = 0 \tag{7.134}$$

Then, according to Lagrange's equations,

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{x}_i} = 0 \tag{7.135}$$

and

$$\frac{\partial L}{\partial \dot{x}_i} = \text{constant} \tag{7.136}$$

or

$$\frac{\partial (T - U)}{\partial \dot{x}_i} = \frac{\partial T}{\partial \dot{x}_i} = \frac{\partial}{\partial \dot{x}_i} \left(\frac{1}{2} m \sum_j \dot{x}_j^2 \right)$$
$$= m \dot{x}_i = p_i = \text{constant} \tag{7.137}$$

Thus, the homogeneity of space implies that the linear momentum **p** of a closed system is constant in time.

This result may also be interpreted according to the following statement: **If the Lagrangian of a system (not necessarily closed) is invariant with respect to translation in a certain direction, then the linear momentum of the system in that direction is constant in time.**

Conservation of Angular Momentum

We stated in Section 2.3 that one characteristic of an inertial reference frame is that space is *isotropic* - that is, that the mechanical properties of a closed system are unaffected by the orientation of the system. In particular, the Lagrangian of a closed system does not change if the system is rotated through an infinitesimal angle. We limit the rotation to an infinitesimal angle because we wish to be able to represent the rotation by a vector; see Section 1.15.

If a system is rotated about a certain axis by an infinitesimal angle $\delta\theta$ (see Figure 7-8), the radius vector $\mathbf{r}$ to a given point changes to $\mathbf{r} + \delta\mathbf{r}$, where (see Equation 1.106)

$$\delta\mathbf{r} = \delta\boldsymbol{\theta} \times \mathbf{r} \tag{7.138}$$

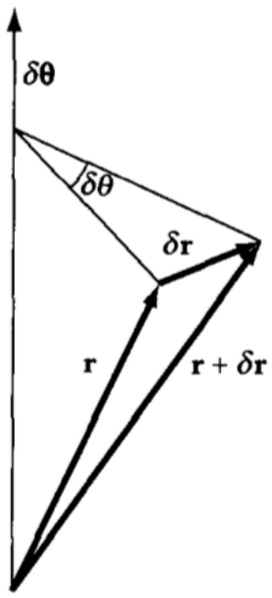


FIGURE 7-8 A system is rotated by an infinitesimal angle $\delta\theta$.

The velocity vectors also change on rotation of the system, and because the transformation equation for all vectors is the same, we have

$$\delta\dot{\mathbf{r}} = \delta\boldsymbol{\theta} \times \dot{\mathbf{r}} \tag{7.139}$$

We consider only a single particle and express the Lagrangian in rectangular coordinates. The change in L caused by the infinitesimal rotation is

$$\delta L = \sum_i \frac{\partial L}{\partial x_i}\delta x_i + \sum_i \frac{\partial L}{\partial \dot{x}_i}\delta \dot{x}_i = 0 \tag{7.140}$$

Equations 7.136 and 7.137 show that the rectangular components of the momentum vector are given by

$$p_i = \frac{\partial L}{\partial \dot{x}_i} \tag{7.141}$$

Lagrange's equations may then be expressed by

$$\dot{p}_i = \frac{\partial L}{\partial x_i} \tag{7.142}$$

Hence, Equation 7.140 becomes

$$\delta L = \sum_i \dot{p}_i \delta x_i + \sum_i p_i \delta \dot{x}_i = 0 \tag{7.143}$$

or

$$\dot{\mathbf{p}} \cdot \delta \mathbf{r} + \mathbf{p} \cdot \delta \dot{\mathbf{r}} = 0 \tag{7.144}$$

Using Equations 7.138 and 7.139, this equation may be written as

$$\dot{\mathbf{p}} \cdot (\delta \boldsymbol{\theta} \times \mathbf{r}) + \mathbf{p} \cdot (\delta \boldsymbol{\theta} \times \dot{\mathbf{r}}) = 0 \tag{7.145}$$

We may permute in cyclic order the factors of a triple scalar product without altering the value. Thus,

$$\delta \boldsymbol{\theta} \cdot (\mathbf{r} \times \dot{\mathbf{p}}) + \delta \boldsymbol{\theta} \cdot (\dot{\mathbf{r}} \times \mathbf{p}) = 0$$

or

$$\delta \boldsymbol{\theta} \cdot [(\mathbf{r} \times \dot{\mathbf{p}}) + (\dot{\mathbf{r}} \times \mathbf{p})] = 0 \tag{7.146}$$

The terms in the brackets are just the factors that result from the differentiation with respect to time of $\mathbf{r} \times \mathbf{p}$:

$$\delta \boldsymbol{\theta} \cdot \frac{d}{dt}(\mathbf{r} \times \mathbf{p}) = 0 \tag{7.147}$$

Because $\delta\boldsymbol{\theta}$ is arbitrary, we must have

$$\frac{d}{dt}(\mathbf{r} \times \mathbf{p}) = 0 \tag{7.148}$$

so

$$\mathbf{r} \times \mathbf{p} = \text{constant} \tag{7.149}$$

But $\mathbf{r} \times \mathbf{p} = \mathbf{L}$; the angular momentum of the particle in a closed system is therefore constant in time.

An important corollary of this theorem is the following. Consider a system in an external force field. If the field possesses an axis of symmetry, then the Lagrangian of the system is invariant with respect to rotations about the symmetry axis. Hence, the angular momentum of the system about the axis of symmetry is constant in time. This is exactly the case discussed in Example 7.4; the vertical direction was an axis of symmetry of the system, and the angular momentum about that axis was conserved.

The importance of the connection between symmetry properties and the *invariance* of physical quantities can hardly be overemphasized. The association goes beyond momentum conservation - indeed beyond classical systems - and finds wide application in modem theories of field phenomena and elementary particles.

We have derived the conservation theorems for a closed system simply by considering the properties of an inertial reference frame. The results, summarized in Table 7-1, are generally credited to Emmy Noether, the originator of Noether's Theorem, which proves a relationship between symmetries and conservation principles.

There are then seven constants (or integrals) of the motion for a closed system: total energy, linear momentum (three components), and angular momentum (three components). These and only these seven integrals have the property that they are *additive* for the particles composing the system; they possess this property whether or not there is an interaction among the particles.

TABLE 7-1

Characteristic of inertial frame	Property of Lagrangian	Conserved quantity
Time homogeneous	Not explicit function of time	Total energy
Space homogeneous	Invariant to translation	Linear momentum
Space isotropic	Invariant to rotation	Angular momentum

7.10 Canonical Equations of Motion - Hamiltonian Dynamics

In the previous section, we found that if the potential energy of a system is velocity independent, then the linear momentum components in rectangular coordinates are given by

$$p_i = \frac{\partial L}{\partial \dot{x}_i} \tag{7.150}$$

By analogy, we extend this result to the case in which the Lagrangian is expressed in generalized coordinates and define the generalized momenta according to

$$\boxed{p_j \equiv \frac{\partial L}{\partial \dot{q}_j}} \tag{7.151}$$

(Unfortunately, the customary notations for ordinary momentum and generalized momentum are the same, even though the two quantities may be quite different.) The Lagrange equations of motion are then expressed by

$$\boxed{\dot{p}_j = \frac{\partial L}{\partial q_j}} \tag{7.152}$$

Using the definition of the generalized momenta, Equation 7.128 for the Hamiltonian may be written as

$$H = \sum_j p_j \dot{q}_j - L \tag{7.153}$$

The Lagrangian is considered to be a function of the generalized coordinates, the generalized velocities, and possibly the time. The dependence of L on the time may arise either if the constraints are time dependent or if the transformation equations connecting the rectangular and generalized coordinates explicitly contain the time. (Recall that we do not consider time-dependent potentials.) We may solve Equation 7.151 for the generalized velocities and express them as

$$\dot{q}_j = \dot{q}_j(q_k, p_k, t) \tag{7.154}$$

Thus, in Equation 7.153, we may make a change of variables from the (q_j , $\dot{q}_j$, t) set to the (q_j, p_j, t) set* and express the Hamiltonian as

$$\boxed{H(q_k, p_k, t) = \sum_j p_j \dot{q}_j - L(q_k, \dot{q}_k, t)} \tag{7.155}$$

This change of variables is similar to that frequently encountered in thermodynamics and falls in the general class of the so-called Legendre transformations.

This equation is written in a manner that stresses the fact that *the Hamiltonian is always considered as a function of the* (q_j, p_j, t) *set, whereas the Lagrangian is a function of the* (q_j , $\dot{q}_j$, t) *set*.

$$\boxed{H = H(q_k, p_k, t), \qquad L = L(q_k, \dot{q}_k, t)} \tag{7.156}$$

The total differential of H is therefore

$$dH = \sum_k \left(\frac{\partial H}{\partial q_k} dq_k + \frac{\partial H}{\partial p_k} dp_k \right) + \frac{\partial H}{\partial t} dt \tag{7.157}$$

According to Equation 7.155, we can also write

$$dH = \sum_k \left(\dot{q}_k \, dp_k + p_k \, d\dot{q}_k - \frac{\partial L}{\partial q_k} dq_k - \frac{\partial L}{\partial \dot{q}_k} d\dot{q}_k \right) - \frac{\partial L}{\partial t} dt \tag{7.158}$$

Using Equations 7.151 and 7.152 to substitute for $\partial L/\partial q_k$, and $\partial L/\partial \dot{q}_k$, the second and fourth terms in the parentheses in Equation 7.158 cancel, and there remains

$$dH = \sum_k (\dot{q}_k dp_k - \dot{p}_k dq_k) - \frac{\partial L}{\partial t} dt \tag{7.159}$$

If we identify the coefficients of dq_k, dp_k, and dt between Equations 7.157 and 7.159, we find

$$\boxed{\dot{q}_k = \frac{\partial H}{\partial p_k}} \tag{7.160}$$

Hamilton's equations of motion

$$\boxed{-\dot{p}_k = \frac{\partial H}{\partial q_k}} \tag{7.161}$$

and

$$-\frac{\partial L}{\partial t} = \frac{\partial H}{\partial t} \tag{7.162}$$

Furthermore, using Equations 7.160 and 7.161 in Equation 7.157, the term in the parentheses vanishes, and it follows that

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} \tag{7.163}$$

Equations 7.160 and 7.161 are **Hamilton's equations of motion**. Because of their symmetric appearance, they are also known as the **canonical equations of motion**. The description of motion by these equations is termed **Hamiltonian dynamics**.

Equation 7.163 expresses the fact that if H does not explicitly contain the time, then the Hamiltonian is a conserved quantity. We have seen previously (Section 7.9) that the Hamiltonian equals the total energy T + U if the potential energy is velocity independent and the transformation equations between $x_{\alpha,i}$ and q_j do not explicitly contain the time. Under these conditions, and if $\partial H/\partial t = 0$, then H = E = constant.

There are 2s canonical equations and they replace the s Lagrange equations. (Recall that s = 3n - m is the number of degrees of freedom of the system.) But the canonical equations are *first-order* differential equations, whereas the Lagrange equations are of **second order**. To use the canonical equations in solving a problem, we must first construct the Hamiltonian as a function of the generalized coordinates and momenta. It may be possible in some instances to do this directly. In more complicated cases, it may be necessary first to set up the Lagrangian and then to calculate the generalized momenta according to Equation 7.151. The equations of motion are then given by the canonical equations.

EXAMPLE 7.11

Use the Hamiltonian method to find the equations of motion of a particle of mass m constrained to move on the surface of a cylinder defined by $x^2 + y^2 = R^2$. The particle is subject to a force directed toward the origin and proportional to the distance of the particle from the origin: $\mathbf{F} = -k\mathbf{r}$.

Solution. The situation is illustrated in Figure 7-9.

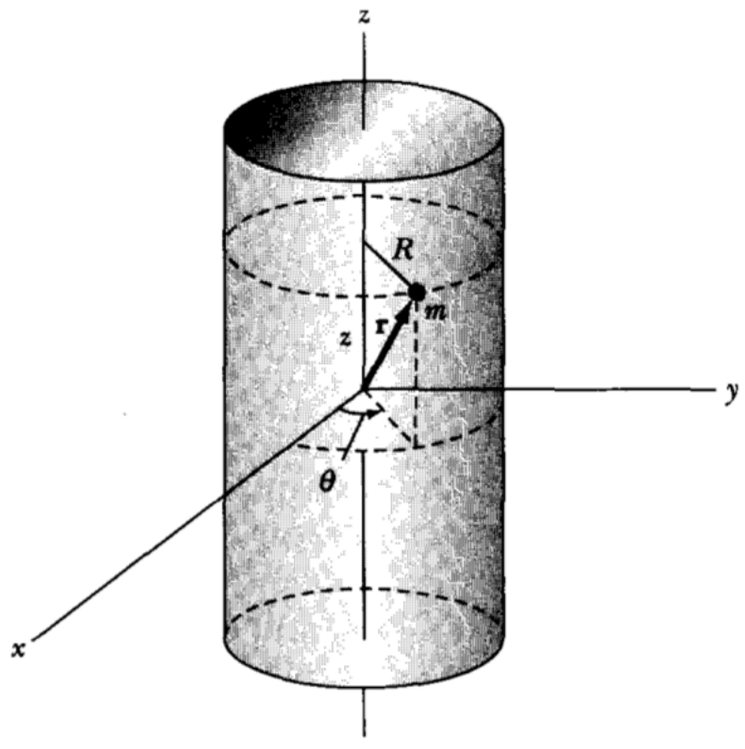


FIGURE 7-9 Example 7.11. A particle is constrained to move on the surface of a cylinder.

The potential corresponding to the force **F** is

$$U = \frac{1}{2} kr^2 = \frac{1}{2} k(x^2 + y^2 + z^2)$$
$$= \frac{1}{2} k(R^2 + z^2) \tag{7.164}$$

We can write the square of the velocity in cylindrical coordinates (see Equation 1.101) as

$$v^2 = \dot{R}^2 + \dot{R}^2\dot{\theta}^2 + \dot{z}^2 \tag{7.165}$$

But in this case, R is a constant, so the kinetic energy is

$$T = \frac{1}{2} m(R^2\dot{\theta}^2 + \dot{z}^2) \tag{7.166}$$

We may now write the Lagrangian as

$$L = T - U = \frac{1}{2} m(R^2\dot{\theta}^2 + \dot{z}^2) - \frac{1}{2} k(R^2 + z^2) \tag{7.167}$$

The generalized coordinates are θ and z, and the generalized momenta are

$$p_\theta = \frac{\partial L}{\partial \dot{\theta}} = mR^2\dot{\theta} \tag{7.168}$$

$$p_z = \frac{\partial L}{\partial \dot{z}} = m\dot{z} \tag{7.169}$$

Because the system is conservative and because the equations of transformation between rectangular and cylindrical coordinates do not explicitly involve the time, the Hamiltonian H is just the total energy expressed in terms of the variables θ , p_θ , z, and p_z. But θ does not occur explicitly, so

$$H(z, p_\theta, p_z) = T + U$$

$$= \frac{p_\theta^2}{2mR^2} + \frac{p_z^2}{2m} + \frac{1}{2}kz^2 \tag{7.170}$$

where the constant term $\frac{1}{2}kR^2$ has been suppressed. The equations of motion are therefore found from the canonical equations:

$$\dot{p}_\theta = -\frac{\partial H}{\partial \theta} = 0 \tag{7.171}$$

$$\dot{p}_z = -\frac{\partial H}{\partial z} = -kz \tag{7.172}$$

$$\dot{\theta} = \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{mR^2} \tag{7.173}$$

$$\dot{z} = \frac{\partial H}{\partial p_z} = \frac{p_z}{m} \tag{7.174}$$

Equations 7.173 and 1.174 just duplicate Equations 7.168 and 7.169. Equations 7.168 and 7.171 give

$$p_\theta = mR^2\dot{\theta} = \text{constant} \tag{7.175}$$

The angular momentum about the z-axis is thus a constant of the motion. This result is ensured, because the z-axis is the symmetry axis of the problem. Combining Equations 7.169 and 7.172, we find

$$\ddot{z} + \omega_0^2 z = 0 \tag{7.176}$$

where

$$\omega_0^2 \equiv k/m \tag{7.177}$$

The motion in the z direction is therefore simple harmonic.

The equations of motion for the preceding problem can also be found by the Lagrangian method using the function L defined by Equation 7.167. In this case, the Lagrange equations of motion are easier to obtain than are the canonical equations. In fact, it is quite often true that the Lagrangian method leads more readily to the equations of motion that does the Hamiltonian method. But because we have greater freedom in choosing the variable in the Hamiltonian formulation of a problem (the q_k and the p_k are independent, whereas the q_k, and the $\dot{q}_k$, are not), we often gain a certain practical advantage by using the Hamiltonian method. For example, in celestial mechanics - particularly in the event that the motions are subject to perturbations caused by the influence of other bodies - it proves convenient to formulate the problem in terms of Hamiltonian dynamics. Generally speaking, however, the great power of the Hamiltonian approach to dynamics does not manifest itself in simplifying the solutions to mechanics problems; rather, it provides a base we can extend to other fields.

The generalized coordinate q_k and the generalized momentum p_k are **canonically conjugate** quantities. According to Equations 7.160 and 7.161, if q_k does not appear in the Hamiltonian, then $\dot{p}_k = 0$, and the conjugate momentum p_k is a constant of the motion. Coordinates not appearing explicitly in the expressions for T and U are said to be *cyclic*. A coordinate cyclic in H is also cyclic in L. But, even if q_k does not appear in L, the generalized velocity $\dot{q}_k$ related to this coordinate is in general still present. Thus

$$L = L(q_1, \ldots, q_{k-1}, q_{k+1}, \ldots, q_s, \dot{q}_1 \ldots, \dot{q}_s, t)$$

and we accomplish no reduction in the number of degrees of freedom of the system, even though one coordinate is cyclic; there are still s second-order equations to be solved. However, in the canonical formulation, if q_k is cyclic, p_k is constant, $p_k = \alpha_k$, and

$$H = H(q_1, \ldots, q_{k-1}, q_{k+1}, \ldots, q_s, p_1, \ldots, p_{k-1}, \alpha_k, p_{k+1}, \ldots, p_s, t)$$

Thus, there are 2s - 2 first-order equations to be solved, and the problem has, in fact, been reduced in complexity; there are in effect only s - 1 degrees of freedom remaining.

The coordinate q_k is completely separated, and it is *ignorable* as far as the remainder of the problem is concerned. We calculate the constant α_k by applying the initial conditions, and the equation of motion for the cyclic coordinate is

$$\dot{q}_k = \frac{\partial H}{\partial \alpha_k} \equiv \omega_k \tag{7.178}$$

which can be immediately integrated to yield

$$q_k(t) = \int \omega_k \, dt \tag{7.179}$$

The solution for a cyclic coordinate is therefore trivial to reduce to quadrature. Consequently, the canonical formulation of Hamilton is particularly well suited for dealing with problems in which one or more of the coordinates are cyclic. The simplest possible solution to a problem would result if the problem could be formulated in such a way that all the coordinates were cyclic. Then, each coordinate would be described in a trivial manner as in Equation 7.179. It is, in fact, possible to find transformations that render all the coordinates cyclic(Hamilton-Jacobi theory), and these procedures lead naturally to a formulation of dynamics particularly useful in constructing modern theories of matter. The general discussion of these topics, however, is beyond the scope of this book.

EXAMPLE 7.12

Use the Hamiltonian method to find the equations of motion for a spherical pendulum of mass m and length b (see Figure 7-10).

Solution. The generalized coordinates are θ and ϕ . The kinetic energy is

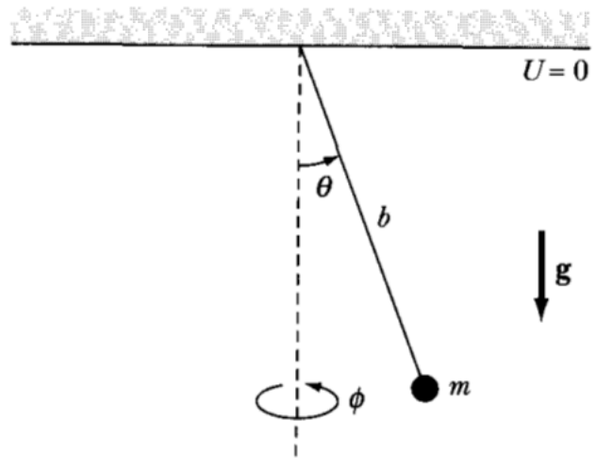


FIGURE 7-10 Example 7.12. A spherical pendulum with generalized coordinates θ and ϕ.

$$T = \frac{1}{2} mb^2\dot{\theta}^2 + \frac{1}{2} mb^2 \sin^2\theta\dot{\phi}^2$$

The only force acting on the pendulum (other than at the point of support) is gravity, and we define the potential zero to be at the pendulum's point of attachment.

$$U = -mgb\cos\theta$$

The generalized momenta are then

$$p_\theta = \frac{\partial L}{\partial\dot{\theta}} = mb^2\dot{\theta} \tag{7.180}$$

$$p_\phi = \frac{\partial L}{\partial\dot{\phi}} = mb^2 \sin^2\theta\dot{\phi} \tag{7.181}$$

We can solve Equations 7.180 and 7.181 for $\dot{\theta}$ and $\dot{\phi}$ in terms p_θ and p_ϕ .

We determine the Hamiltonian from Equation 7.155 or from H = T + U (because the conditions for Equation 7.130 apply)

$$\begin{aligned} H &= T + U \\ &= \frac{1}{2}mb^2\frac{p_\theta^2}{(mb^2)^2} + \frac{1}{2}\frac{mb^2\sin^2\theta p_\phi^2}{(mb^2\sin^2\theta)^2} - mgb\cos\theta \\ &= \frac{p_\theta^2}{2mb^2} + \frac{p_\phi^2}{2mb^2\sin^2\theta} - mgb\cos\theta \end{aligned}$$

The equations of motion are

$$\dot{\theta} = \frac{\partial H}{\partial p_\theta} = \frac{p_\theta}{mb^2}$$

$$\dot{\phi} = \frac{\partial H}{\partial p_\phi} = \frac{p_\phi}{mb^2 \sin^2 \theta}$$

$$\dot{p}_\theta = -\frac{\partial H}{\partial \theta} = \frac{p_\phi^2 \cos \theta}{mb^2 \sin^3 \theta} - mgb \sin \theta$$

$$\dot{p}_\phi = -\frac{\partial H}{\partial \phi} = 0$$

Because ϕ is cyclic, the momentum p_ϕ about the symmetry axis is constant.

7.11 Some Comments Regarding Dynamical Variables and Variational Calculations in Physics

We originally obtained Lagrange's equations of motion by stating Hamilton's Principle as a variational integral and then using the results of the preceding chapter on the calculus of variations. Because the method and the application were thereby separated, it is perhaps worthwhile to restate the argument in an orderly but abbreviated way.

Hamilton's Principle is expressed by

$$\delta \int_{t_1}^{t_2} L(q_j, \dot{q}_j, t)\, dt = 0 \tag{7.182}$$

Applying the variational procedure specified in Section 6.7, we have

$$\int_{t_1}^{t_2} \left(\frac{\partial L}{\partial q_j} \delta q_j + \frac{\partial L}{\partial \dot{q}_j} \delta \dot{q}_j \right) dt = 0$$

Next, we assert that the δq_j and the $\delta \dot{q}_j$ are not independent, so the variation operation and the time differentiation can be interchanged:

$$\delta \dot{q}_j = \delta\left(\frac{dq_j}{dt}\right) = \frac{d}{dt}\delta q_j \tag{7.183}$$

The varied integral becomes (after the integration by parts in which the δq_j are set equal to zero at the endpoints)

$$\int_{t_1}^{t_2}\left(\frac{\partial L}{\partial q_j} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_j}\right)\delta q_j\, dt = 0 \tag{7.184}$$

The requirement that the δq_j be independent variations leads immediately to Lagrange's equations.

In Hamilton's Principle, expressed by the variational integral in Equation 7.182, the Lagrangian is a function of the generalized coordinates and the generalized velocities. But only the q_j are considered as independent variables; the generalized velocities are simply the time derivatives of the q_j. When the integral is reduced to the form given by Equation 7.184, we state that the δq_j are independent variations; thus the integrand must vanish identically, and Lagrange's equations result. We may therefore pose this question: Because the dynamical motion of the system is completely determined by the initial conditions, what is the meaning of the variations δq_j? Perhaps a sufficient answer is that the variables are to be considered geometrically feasible within the limits of the given constraints - although they are not dynamically possible; that is, when using a variational procedure to obtain Lagrange's equations, it is convenient to ignore temporarily the fact that we are dealing with a physical system whose motion is completely determined and subject to no variation and to consider instead only a certain abstract mathematical problem. Indeed, this is the spirit in which any variational calculation relating to a physical process must be carried out.

In adopting such a viewpoint, we must not be overly concerned with the fact that the variational procedure may be contrary to certain known physical properties of the system. (For example, energy is generally not conserved in passing from the true path to the varied path.) A variational calculation simply tests various *possible* solutions to a problem and prescribes a method for selecting the *correct* solution.

The canonical equations of motion can also be obtained directly from a variational calculation based on the so-called **modified Hamilton's Principle**. The Lagrangian function can be expressed as (see Equation 7.153):

$$L = \sum_j p_j \dot{q}_j - H(q_j, p_j, t) \tag{7.185}$$

and the statement of Hamilton's Principle contained in Equation 7.182 can be modified to read

$$\boxed{\delta \int_{t_1}^{t_2} \left(\sum_j p_j \dot{q}_j - H \right) dt = 0} \tag{7.186}$$

Carrying out the variation in the standard manner, we obtain

$$\int_{t_1}^{t_2} \sum_j \left(p_j \delta \dot{q}_j + \dot{q}_j \delta p_j - \frac{\partial H}{\partial q_j} \delta q_j - \frac{\partial H}{\partial p_j} \delta p_j \right) dt = 0 \tag{7.187}$$

In the Hamiltonian formulation, the q_j and the p_j are considered to be independent. The $\dot{q}_j$ are again not independent of the q_j , so Equation 7.183 can be used to express the first term in Equation 7.187 as

$$\int_{t_1}^{t_2} \sum_j p_j \delta \dot{q}_j \, dt = \int_{t_1}^{t_2} \sum_j p_j \frac{d}{dt} \delta q_j \, dt$$

Integrating by parts, the integrated term vanishes, and we have

$$\int_{t_1}^{t_2} \sum_j p_j \delta \dot{q}_j \, dt = -\int_{t_1}^{t_2} \sum_j \dot{p}_j \delta q_j \, dt \tag{7.188}$$

Equation 7.187 then becomes

$$\int_{t_1}^{t_2} \sum_j \left\{ \left(\dot{q}_j - \frac{\partial H}{\partial p_j} \right) \delta p_j - \left(\dot{p}_j + \frac{\partial H}{\partial q_j} \right) \delta q_j \right\} dt = 0 \tag{7.189}$$

If δq_j and δp_j represent independent variations, the terms in the parentheses must separately vanish and Hamilton's canonical equations result

In the preceding section, we obtained the canonical equations by writing two different expressions for the total differential of the Hamiltonian (Equations 7.157 and 7.159) and then equating the coefficients of dq_j and dp_j. Such a procedure is valid if the q_j and the p_j are independent variables. Therefore, both in the previous derivation and in the preceding variational calculation, we obtained the canonical equations by exploring the independent nature of the generalized coordinates and the generalized momenta.

The coordinates and momenta are not actually "independent" in the ultimate sense of the word. For if the time dependence of each of the coordinates is known, $q_j = q_j(t)$, the problem is completely solved. The generalized velocities can be calculated from

$$\dot{q}_j(t) = \frac{d}{dt} q_j(t)$$

and the generalized momenta are

$$p_j = \frac{\partial}{\partial \dot{q}_j} L(q_j, \dot{q}_j, t)$$

The essential point is that, whereas the q_j and the $\dot{q}_j$ are related by a simple time derivative *independent of the manner in which the system behaves*, the connection between the qj and the pj are the *equations of motion themselves*. Finding the relations that connect the q_j and the p_j (and thereby eliminating the assumed independence of these quantities) is therefore tantamount to solving the problem.

7.12 Phase Space and Liouville's Theorem

We pointed out previously that the generalized coordinates q_j can be used to define an s-dimensional *configuration space* with every point representing a certain state of the system. Similarly, the generalized momenta p_j define an s-dimensional momentum space with every point representing a certain condition of motion of the system. A given point in configuration space specifies only the position of each of the particles in the system; nothing can be inferred regarding the motion of the particles. The reverse is true for momentum space. In Chapter 3, we found it profitable to represent geometrically the dynamics of simple oscillatory systems by phase diagrams. If we use this concept with more complicated dynamical systems, then a 2s-dimensional space consisting of the q_j and the p_j allows us to represent both the positions and the momenta of all particles. This generalization is called **Hamiltonian phase space** or, simply, **phase space**.

We previously plotted in the phase diagrams the position versus a quantity proportional to the velocity. In Hamiltonian phase space, this latter quantity becomes the generalized momentum.

EXAMPLE 7.13

Construct the phase diagram for the particle in Example 7.11.

Solution. The particle has two degrees of freedom (θ, z), so the phase space for this example is actually four dimensional: θ, p_θ, z, p_z. But p_θ is constant and therefore may be suppressed. In the z direction, the motion is simple harmonic, and so the projection onto the z-pz plane of the phase path for any total energy H is just an ellipse. Because $\dot{\theta}$ = constant, the phase path must represent motion increasing uniformly with θ. Thus, the phase path on any surface H = constant is a **uniform elliptic spiral** (Figure 7-11).

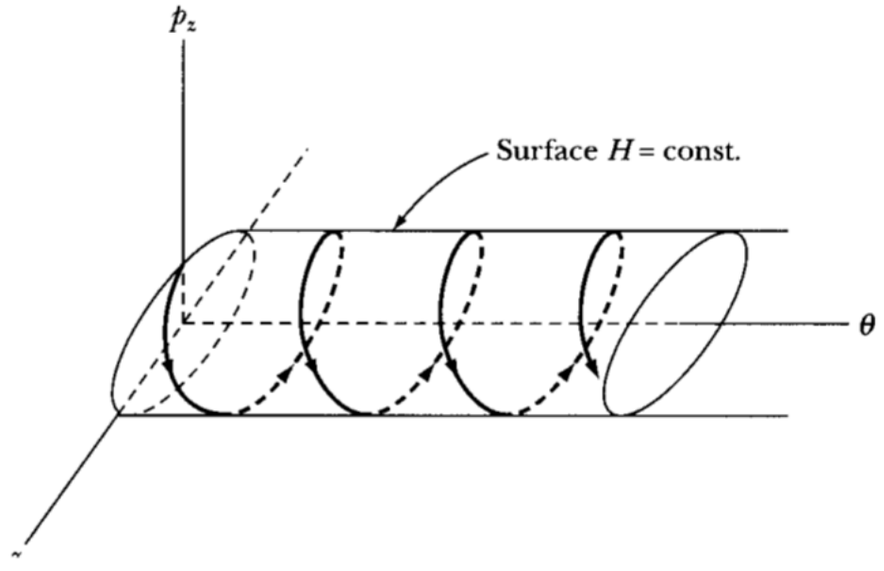


FIGURE 7-11 Example 7.13. The phase path for the particle in Example 7.11.

If, at a given time, the position and momenta of all the particles in a system are known, then with these quantities as initial conditions, the subsequent motion of the system is completely determined; that is, starting from a point $q_j(0)$, $p_j(0)$ in phase space, the representative point describing the system moves along a unique phase path. In principle, this procedure can always be followed and a solution obtained. But if the number of degrees of freedom of the system is large, the set of equations of motion may be too complicated to solve in a reasonable time. Moreover, for complex systems, such as a quantity of gas, it is a practical impossibility to determine the initial conditions for each constituent molecule. Because we cannot identify any particular point in phase space as representing the actual conditions at any given time, we must devise some alternative approach to study the dynamics of such systems. We therefore arrive at the point of departure of statistical mechanics. The Hamiltonian formulation of dynamics is ideal for the statistical study of complex systems. We demonstrate this in part by now proving a theorem that is fundamental for such investigations.

For a large collection of particles - say, gas molecules - we are unable to identify the particular point in phase space correctly representing the system. But we may fill the phase space with a collection of points, each representing a possible condition of the system; that is, we imagine a large number of systems

Because we are unable to discuss the details of the particles' motion in the actual system, we substitute a discussion of an ensemble of equivalent systems. Each representative point in phase space corresponds to a single system of the ensemble, and the motion of a particular point represents the independent motion of that system. Thus, no two of the phase paths may ever intersect.

We may consider the representative points to be sufficiently numerous that we can define a *density in phase space* ρ . The volume elements of the phase space defining the density must be sufficiently large to contain a large number of representative points, but they must also be sufficiently small so that the density varies continuously. The number N of systems whose representative points lie within a volume dv of phase space is

$$N = \rho\, dv \tag{7.190}$$

where

$$dv = dq_1\, dq_2 \cdots dq_s\, dp_1\, dp_2 \cdots dp_s \tag{7.191}$$

As before, s is the number of degrees of freedom of each system in the ensemble.
Consider an element of area in the q_k-p_k plane in phase space (Figure 7-12).

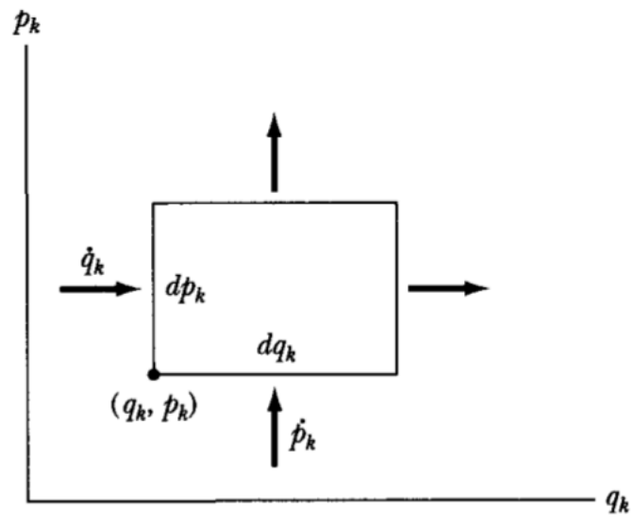


FIGURE 7-12 An element of area $dA = dq_k\, dp_k$ in the q_k–p_k plane in phase space.

The number of representative points moving across the left-hand edge into the area per unit time is

$$\rho \frac{dq_k}{dt} dp_k = \rho \dot{q}_k \, dp_k$$

and the number moving across the lower edge into the area per unit time is

$$\rho \frac{dp_k}{dt} dq_k = \rho \dot{p}_k \, dq_k$$

so that the total number of representative points moving into the area $dq_k dp_k$ per unit time is

$$\rho(\dot{q}_k \, dp_k + \dot{p}_k \, dq_k) \tag{7.192}$$

By a Taylor series expansion, the number of representative points moving out of the area per unit time is (approximately)

$$\left[\rho \dot{q}_k + \frac{\partial}{\partial q_k}(\rho \dot{q}_k) \, dq_k\right] dp_k + \left[\rho \dot{p}_k + \frac{\partial}{\partial p_k}(\rho \dot{p}_k) \, dp_k\right] dq_k \tag{7.193}$$

Thus, the total increase in density in $dq_k dp_k$ per unit time is the difference between Equations 7.192 and 7.193:

$$\frac{\partial \rho}{\partial t} dq_k \, dp_k = -\left[\frac{\partial}{\partial q_k}(\rho \dot{q}_k) + \frac{\partial}{\partial p_k}(\rho \dot{p}_k)\right] dq_k \, dp_k \tag{7.194}$$

After dividing by $dq_k dp_k$, and summing this expression over all possible values of k, we find

$$\frac{\partial \rho}{\partial t} + \sum_{k=1}^{s} \left(\frac{\partial \rho}{\partial q_k} \dot{q}_k + \rho \frac{\partial \dot{q}_k}{\partial q_k} + \frac{\partial \rho}{\partial p_k} \dot{p}_k + \rho \frac{\partial \dot{p}_k}{\partial p_k}\right) = 0 \tag{7.195}$$

From Hamilton's equations (Equations 7.160 and 7.161), we have (if the second partial derivatives of H are continuous)

$$\frac{\partial \dot{q}_k}{\partial q_k} + \frac{\partial \dot{p}_k}{\partial p_k} = 0 \tag{7.196}$$

so Equation 7.195 becomes

$$\frac{\partial \rho}{\partial t} + \sum_k \left(\frac{\partial \rho}{\partial q_k} \frac{dq_k}{dt} + \frac{\partial \rho}{\partial p_k} \frac{dp_k}{dt} \right) = 0 \tag{7.197}$$

But this is just the total time derivative of ρ, so we conclude that

$$\boxed{\frac{d\rho}{dt} = 0} \tag{7.198}$$

This important result, known as **Liouville's theorem**, states that the density of representative points in phase space corresponding to the motion of a system of particles remains constant during the motion. It must be emphasized that we have been able to establish the invariance of the density ρ only because the problem was formulated in phase space; an equivalent theorem for configuration space does not exist. Thus, we must use Hamiltonian dynamics (rather than Lagrangian dynamics) to discuss ensembles in statistical mechanics.

Liouville's theorem is important not only for aggregates of microscopic particles, as in the statistical mechanics of gaseous systems and the focusing properties of charged-particle accelerators, but also in certain macroscopic systems. For example, in stellar dynamics, the problem is inverted and by studying the distribution function ρ of stars in the galaxy, the potential U of the galactic gravitational field may be inferred.

7.13 Virial Theorem

Another important result of a statistical nature is worthy of mention. Consider a collection of particles whose position vectors $\mathbf{r}_\alpha$ and momenta $\mathbf{p}_\alpha$ are both bounded (i.e., remain finite for all values of the time). Define a quantity

$$S \equiv \sum_\alpha \mathbf{p}_\alpha \cdot \mathbf{r}_\alpha \tag{7.199}$$

The time derivative of S is

$$\frac{dS}{dt} = \sum_\alpha (\mathbf{p}_\alpha \cdot \dot{\mathbf{r}}_\alpha + \dot{\mathbf{p}}_\alpha \cdot \mathbf{r}_\alpha) \tag{7.200}$$

If we calculate the average value of dS/dt over a time interval t, we find

$$\left\langle \frac{dS}{dt} \right\rangle = \frac{1}{\tau} \int_0^\tau \frac{dS}{dt}\, dt = \frac{S(\tau) - S(0)}{\tau} \tag{7.201}$$

If the system's motion is periodic - and if τ is some integer multiple of the period - then S(t) = S(0), and $\langle \dot{S} \rangle$ vanishes. But even if the system does not exhibit any periodicity, then - because S is by hypothesis a bounded function - we can make $\langle \dot{S} \rangle$ as small as desired by allowing the time τ to become sufficiently long. Therefore, the time average of the right-hand side of Equation 7.201 can always be made to vanish (or at least to approach zero). Thus, in this limit, we have

$$\left\langle \sum_\alpha \mathbf{p}_\alpha \cdot \dot{\mathbf{r}}_\alpha \right\rangle = - \left\langle \sum_\alpha \dot{\mathbf{p}}_\alpha \cdot \mathbf{r}_\alpha \right\rangle \tag{7.202}$$

On the left-hand side of this equation, $\mathbf{p}_\alpha \cdot \dot{\mathbf{r}}_\alpha$ is twice the kinetic energy. On the right-hand side, $\dot{\mathbf{p}}_\alpha$ is just the force $\mathbf{F}_\alpha$ on the αth particle. Hence,

$$\left\langle 2\sum_{\alpha} T_{\alpha} \right\rangle = -\left\langle \sum_{\alpha} \mathbf{F}_{\alpha} \cdot \dot{\mathbf{r}}_{\alpha} \right\rangle \tag{7.203}$$

The sum over T_α is the total kinetic energy T of the system, so we have the general result

$$\langle T \rangle = -\frac{1}{2}\left\langle \sum_{\alpha} \mathbf{F}_{\alpha} \cdot \mathbf{r}_{\alpha} \right\rangle \tag{7.204}$$

The right-hand side of this equation was called by Clausius the **virial** of the system, and the **virial theorem** states that *the average kinetic energy of a system of particles is equal to its virial*.

EXAMPLE 7.14

Consider an ideal gas containing N atoms in a container of volume V, pressure P, and absolute temperature T_1 (not to be confused with the kinetic energy T). Use the virial theorem to derive the equation of state for a perfect gas.

Solution. According to the equipartition theorem, the average kinetic energy of each atom in the ideal gas is 3/2 kT_1, where k is the Boltzmann constant. The total average kinetic energy becomes

$$\langle T \rangle = \frac{3}{2} N k T_1 \tag{7.205}$$

The right-hand side of the virial theorem (Equation 7.204) contains the forces $\mathbf{F}_\alpha$. For an ideal perfect gas, no force of interaction occurs between atoms. The only force is represented by the force of constraint of the walls. The atoms bounce elastically off the walls, which are exerting a pressure on the atoms.

Because the pressure is force per unit area, we find the instantaneous differential force over a differential area to be

$$d\mathbf{F}_\alpha = -\mathbf{n}\,P\,dA \tag{7.206}$$

where **n** is a unit vector normal to the surface dA and pointing outward. The right-hand side of the virial theorem becomes

$$-\frac{1}{2}\left\langle \sum_\alpha \mathbf{F}_\alpha \cdot \mathbf{r}_\alpha \right\rangle = \frac{P}{2}\int \mathbf{n}\cdot\mathbf{r}\,dA \tag{7.207}$$

We use the divergence theorem to relate the surface integral to a volume integral.

$$\int \mathbf{n}\cdot\mathbf{r}\,dA = \int \nabla\cdot\mathbf{r}\,dV = 3\int dV = 3V \tag{7.208}$$

The virial theorem result is

$$\frac{3}{2}NkT = \frac{3PV}{2}$$

$$NkT = PV \tag{7.209}$$

which is the ideal gas law.

If the forces $\mathbf{F}_\alpha$ can be derived from potentials U_α, Equation 7.204 may be rewritten as

$$\langle T\rangle = \frac{1}{2}\left\langle \sum_\alpha \mathbf{r}_\alpha \cdot \nabla U_\alpha \right\rangle \tag{7.210}$$

Of particular interest is the case of two particles that interact according to a central power-law force: $F \propto r^n$. Then, the potential is of the form

$$U = kr^{n+1} \tag{7.211}$$

Therefore

$$\mathbf{r}\cdot\nabla U = \frac{dU}{dr} = k(n+1)r^{n+1} = (n+1)U \tag{7.212}$$

and the virial theorem becomes

$$\langle T \rangle = \frac{n+1}{2} \langle U \rangle \tag{7.213}$$

If the particles have a gravitational interaction, then n = -2, and

$$\langle T \rangle = -\frac{1}{2} \langle U \rangle, \quad n = -2$$

This relation is useful in calculating, for example, the energetics in planetary motion.

CHAPTER 8 - Central-Force Motion

8.1 Introduction

The motion of a system consisting of two bodies affected by a force directed along the line connecting the centers of the two bodies (i.e., a *central force*) is an extremely important physical problem - one we can solve completely. The importance of such a problem lies in large measure in two quite different realms of physics: the motion of celestial bodies - planets, moons, comets, double stars, and the like - and certain two-body nuclear interactions, such as the scattering of a particles by nuclei. In the pre-quantum mechanics days, physicists also described the hydrogen atom in terms of a classical two-body central force. Although such a description is still useful in a qualitative sense, the quantum-theoretical approach must be used for a detailed description. In addition to some general considerations regarding motion in central-force fields, we discuss in this and the following chapter several of the problems of two bodies encountered in celestial mechanics and in nuclear and particle physics.

8.2 Reduced Mass

Describing a system consisting of two particles requires the specification of six quantities; for example, the three components of each of the two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ for the particles. The orientation of the particles is assumed to be unimportant; that is, they are spherically symmetric (or are point particles). Alternatively, we may choose the three components of the center-of-mass vector $\mathbf{R}$ and the three components of $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ (see Figure 8-la).

Here, we restrict our attention to systems without frictional losses and for which the potential energy is a function only of $r = |\mathbf{r}_1 - \mathbf{r}_2|$. The Lagrangian for such a system may be written as

$$L = \frac{1}{2}m_1|\dot{\mathbf{r}}_1|^2 + \frac{1}{2}m_2|\dot{\mathbf{r}}_2|^2 - U(r) \tag{8.1}$$

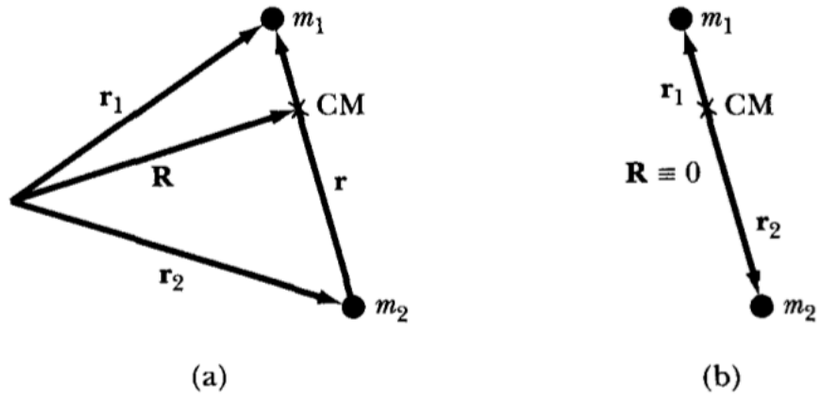


FIGURE 8-1 Two methods to describe the position of two particles. (a) From an arbitrary coordinate system origin, and (b) from the center of mass. The position vectors are $\mathbf{r}_1$ and $\mathbf{r}_2$, the center-of-mass vector is **R,** and the relative vector $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$.

Because translational motion of the system as a whole is uninteresting from the standpoint of the particle orbits with respect to one another, we may choose the origin for the coordinate system to be the particles' center of mass - that is, $\mathbf{R} = 0$ (see Figure 8-lb). Then (see Section 9.2)

$$m_1\mathbf{r}_1 + m_2\mathbf{r}_2 = 0 \tag{8.2}$$

This equation, combined with $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, yields

$$\left.\begin{aligned} \mathbf{r}_1 &= \frac{m_2}{m_1 + m_2}\mathbf{r} \\ \mathbf{r}_2 &= -\frac{m_1}{m_1 + m_2}\mathbf{r} \end{aligned}\right\} \tag{8.3}$$

Substituting Equation 8.3 into the expression for the Lagrangian gives

$$\boxed{L = \frac{1}{2}\mu|\dot{\mathbf{r}}|^2 - U(r)} \tag{8.4}$$

where μ , is the **reduced mass**,

$$\mu \equiv \frac{m_1 m_2}{m_1 + m_2} \tag{8.5}$$

We have therefore formally reduced the problem of the motion of two bodies to an equivalent one-body problem in which we must determine only the motion of a "particle" of mass μ , in the central field described by the potential function U(r).

Once we obtain the solution for **r**(t) by applying the Lagrange equations to Equation 8.4, we can find the individual motions of the particles, $\mathbf{r}_1(t)$ and $r_2(t)$, by using Equation 8.3. This latter step is not necessary if only the orbits relative to one another are required.

8.3 Conservation Theorems - First Integrals of the Motion

The system we wish to discuss consists of a particle of mass μ moving in a central-force field described by the potential function U(r). Because the potential energy depends only on the distance of the particle from the force center and not on the orientation, the system possesses **spherical symmetry**; that is, the system's rotation about any fixed axis through the center of force cannot affect the equations of motion. We have already shown (see Section 7.9) that under such conditions the angular momentum of the system is conserved:

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} = \text{constant} \tag{8.6}$$

From this relation, it should be clear that both the radius vector and the linear momentum vector of the particle lie always in a plane normal to the angular momentum vector L, which is fixed in space (see Figure 8-2). Therefore, we have only a two-dimensional problem, and the Lagrangian may then be conveniently expressed in plane polar coordinates:

$$L = \frac{1}{2}\mu(\dot{r}^2 + r^2\dot{\theta}^2) - U(r) \tag{8.7}$$

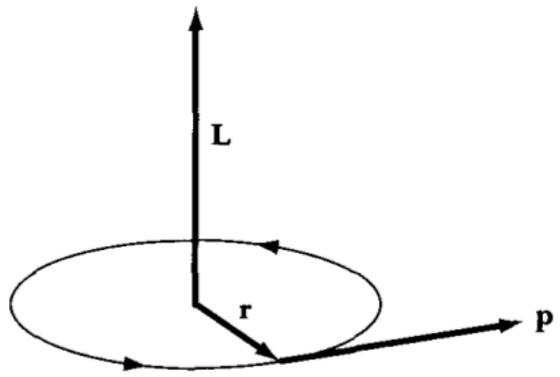


FIGURE 8-2 The motion of a particle of mass μ moving in a central-force field is described by the position vector **r**, linear momentum **p**, and constant angular momentum **L**.

Because the Lagrangian is cyclic in θ , the angular momentum conjugate to the coordinate θ is conserved:

$$\dot{p}_\theta = \frac{\partial L}{\partial \theta} = 0 = \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} \tag{8.8}$$

or

$$p_\theta \equiv \frac{\partial L}{\partial \dot{\theta}} = \mu r^2 \dot{\theta} = \text{constant} \tag{8.9}$$

The system's symmetry has therefore permitted us to integrate immediately one of the equations of motion. The quantity p_θ is a *first integral* of the motion, and we denote its constant value by the symbol l :

$$\boxed{l \equiv \mu r^2 \dot{\theta} = \text{constant}} \tag{8.10}$$

Note that l can be negative as well as positive. That l is constant has a simple geometric interpretation. Referring to Figure 8-3, we see that in describing the path **r**(t), the radius vector sweeps out an area $\frac{1}{2} r^2 d\theta$ in a time interval dt

$$dA = \frac{1}{2} r^2 d\theta \tag{8.11}$$

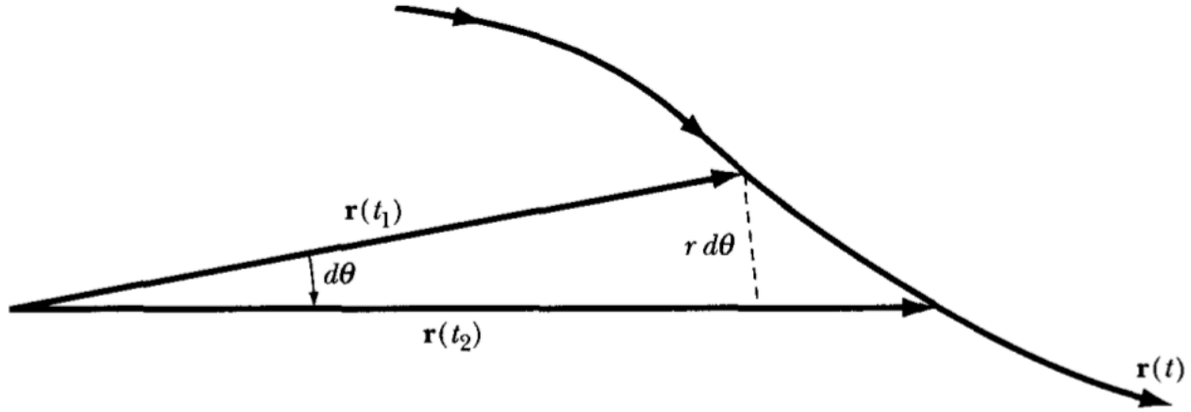


FIGURE 8-3 The path of a particle is described by $\mathbf{r}(t)$. The radius vector sweeps out an area $dA = \frac{1}{2}r^2 d\theta$ in a time interval dt.

On dividing by the time interval, the areal velocity is shown to be

$$\begin{aligned} \frac{dA}{dt} &= \frac{1}{2}r^2\frac{d\theta}{dt} = \frac{1}{2}r^2\dot{\theta} \\ &= \frac{l}{2\mu} = \text{constant} \end{aligned} \tag{8.12}$$

Thus, the areal velocity is constant in time. This result was obtained empirically by Kepler for planetary motion, and it is known as **Kepler's Second Law**. It is important to note that the conservation of the areal velocity is not limited to an inverse-square-law force (the case for planetary motion) but is a general result for central-force motion.

Because we have eliminated from consideration the uninteresting uniform motion of the system's center of mass, the conservation of linear momentum adds nothing new to the description of the motion. The conservation of energy is thus the only remaining first integral of the problem. The conservation of the total energy E is automatically ensured because we have limited the discussion to nondissipative systems. Thus,

$$T + U = E = \text{constant} \tag{8.13}$$

and

$$E = \frac{1}{2}\mu(\dot{r}^2 + r^2\dot{\theta}^2) + U(r)$$

or

$$\boxed{E = \frac{1}{2}\mu\dot{r}^2 + \frac{1}{2}\frac{l^2}{\mu r^2} + U(r)} \tag{8.14}$$

8.4 Equations of Motion

When U(r) is specified, Equation 8.14 completely describes the system, and the integration of this equation gives the general solution of the problem in terms of the parameters E and l . Solving Equation 8.14 for $\dot{r}$, we have

$$\dot{r} = \frac{dr}{dt} = \pm\sqrt{\frac{2}{\mu}(E - U) - \frac{l^2}{\mu^2 r^2}} \tag{8.15}$$

This equation can be solved for dt and integrated to yield the solution t = t{(r). An inversion of this result then gives the equation of motion in the standard form r = r(t). At present, however, we are interested in the equation of the path in terms of r and θ . We can write

$$d\theta = \frac{d\theta}{dt}\frac{dt}{dr}dr = \frac{\dot{\theta}}{\dot{r}}dr \tag{8.16}$$

Into this relation, we can substitute $\dot{\theta} = l/\mu r^2$ (Equation 8.10) and the expression for $\dot{r}$ from Equation 8.15. Integrating, we have

$$\theta(r) = \int \frac{\pm(l/r^2)\,dr}{\sqrt{2\mu\left(E - U - \dfrac{l^2}{2\mu r^2}\right)}} \tag{8.17}$$

Furthermore, because l is constant in time, $\dot{\theta}$ cannot change sign and therefore θ(t) must increase or decrease monotonically with time.

Although we have reduced the problem to the formal evaluation of an integral, the actual solution can be obtained only for certain specific forms of the force law. If the force is proportional to some power of the radial distance, $F(r) \propto r^n$, then the solution can be expressed in terms of elliptic integrals for certain integer and fractional values of n. Only for n = 1, -2, and -3 are the solutions expressible in terms of circular functions (sines and cosines). The case n = 1 is just that of the harmonic oscillator (see Chapter 3), and the case n = — 2 is the important inverse-square-law force treated in Sections 8.6 and 8.7. These two cases, n = 1, —2, are of prime importance in physical situations.

We have therefore solved the problem in a formal way by combining the equations that express the conservation of energy and angular momentum into a single result, which gives the equation of the orbit $\theta = \theta(r)$. We can also attack the problem using Lagrange's equation for the coordinate r:

$$\frac{\partial L}{\partial r} - \frac{d}{dt}\frac{\partial L}{\partial \dot{r}} = 0$$

Using Equation 8.7 for L, we find

$$\mu(\ddot{r} - r\dot{\theta}^2) = -\frac{\partial U}{\partial r} = F(r) \tag{8.18}$$

Equation 8.18 can be cast in a form more suitable for certain types of calculations by making a simple change of variable:

$$u \equiv \frac{1}{r}$$

First, we compute

$$\frac{du}{d\theta} = -\frac{1}{r^2}\frac{dr}{d\theta} = -\frac{1}{r^2}\frac{dr}{dt}\frac{dt}{d\theta} = -\frac{1}{r^2}\frac{\dot{r}}{\dot{\theta}}$$

But from Equation 8.10, $\dot{\theta} = l/\mu r^2$, so

$$\frac{du}{d\theta} = -\frac{\mu}{l}\dot{r}$$

Next, we write

$$\frac{d^2u}{d\theta^2} = \frac{d}{d\theta}\left(-\frac{\mu}{l}\dot{r}\right) = \frac{dt}{d\theta}\frac{d}{dt}\left(-\frac{\mu}{l}\dot{r}\right) = -\frac{\mu}{l\dot{\theta}}\ddot{r}$$

and with the same substitution for $\dot{\theta}$, we have

$$\frac{d^2u}{d\theta^2} = -\frac{\mu^2}{l^2}r^2\ddot{r}$$

Therefore, solving for $\ddot{r}$ and $r\dot{\theta}^2$ in terms of u, we find

$$\left.\begin{aligned} \ddot{r} &= -\frac{l^2}{\mu^2}u^2\frac{d^2u}{d\theta^2} \\ r\dot{\theta}^2 &= \frac{l^2}{\mu^2}u^3 \end{aligned}\right\} \tag{8.19}$$

Substituting Equation 8.19 into Equation 8.18, we obtain the transformed equation of motion:

$$\frac{d^2u}{d\theta^2} + u = -\frac{\mu}{l^2}\frac{1}{u^2}F(1/u) \tag{8.20}$$

which we may also write as

$$\boxed{\frac{d^2}{d\theta^2}\left(\frac{1}{r}\right) + \frac{1}{r} = -\frac{\mu r^2}{l^2}F(r)} \tag{8.21}$$

This form of the equation of motion is particularly useful if we wish to find the force law that gives a particular known orbit r = r(θ).

EXAMPLE 8.1

Find the force law for a central-force field that allows a particle to move in a logarithmic spiral orbit given by $r = ke^{\alpha\theta}$, where k and α are constants.

Solution. We use Equation 8.21 to determine the force law F(r). First, we determine

$$\frac{d}{d\theta}\left(\frac{1}{r}\right) = \frac{d}{d\theta}\left(\frac{e^{-\alpha\theta}}{k}\right) = \frac{-\alpha e^{-\alpha\theta}}{k}$$

$$\frac{d^2}{d\theta^2}\left(\frac{1}{r}\right) = \frac{\alpha^2 e^{-\alpha\theta}}{k} = \frac{\alpha^2}{r}$$

From Equation 8.21, we now determine F(r).

$$F(r) = \frac{-l^2}{\mu r^2}\left(\frac{\alpha^2}{r} + \frac{1}{r}\right) \tag{8.22}$$

$$F(r) = \frac{-l^2}{\mu r^3}(\alpha^2 + 1)$$

Thus, the force law is an attractive inverse cube.

EXAMPLE 8.2

Determine r(t) and θ(t) for the problem in Example 8.1.

Solution. From Equation 8.10, we find

$$\dot{\theta} = \frac{l}{\mu r^2} = \frac{l}{\mu k^2 e^{2\alpha\theta}} \tag{8.23}$$

Rearranging Equation 8.23 gives

$$e^{2\alpha\theta}\,d\theta = \frac{l}{\mu k^2}\,dt$$

and integrating gives

$$\frac{e^{2\alpha\theta}}{2\alpha} = \frac{lt}{\mu k^2} + C'$$

where C is an integration constant. Multiplying by 2α and letting C = 2αC' gives

$$e^{2\alpha\theta} = \frac{2\alpha lt}{\mu k^2} + C \tag{8.24}$$

We solve for θ(t) by taking the natural logarithm of Equation 8.24:

$$\theta(t) = \frac{1}{2\alpha}\ln\left(\frac{2\alpha lt}{\mu k^2} + C\right) \tag{8.25}$$

We can similarly solve for r(t) by examining Equations 8.23 and 8.24:

$$\frac{r^2}{k^2} = e^{2\alpha\theta} = \frac{2\alpha lt}{\mu k^2} + C$$

$$r(t) = \left[\frac{2\alpha l}{\mu}t + k^2 C\right]^{1/2} \tag{8.26}$$

The integration constant C and angular momentum l needed for Equations 8.25 and 8.26 are determined from the initial conditions.

EXAMPLE 8.3

What is the total energy of the orbit of the previous two examples?

Solution. The energy is found from Equation 8.14. In particular, we need $\dot{r}$ and U(r).

$$U(r) = -\int F\,dr = \frac{+l^2}{\mu}(\alpha^2 + 1)\int r^{-3}dr$$

$$U(r) = -\frac{l^2(\alpha^2 + 1)}{2\mu}\frac{1}{r^2} \tag{8.27}$$

where we have let $U(\infty) = 0$.

We rewrite Equation 8.10 to determine $\dot{r}$:

$$\dot{\theta} = \frac{d\theta}{dt} = \frac{d\theta}{dr}\frac{dr}{dt} = \frac{1}{\mu r^2} \tag{8.28}$$

$$\dot{r} = \frac{dr}{d\theta}\frac{l}{\mu r^2} = \alpha k e^{\alpha\theta}\frac{l}{\mu r^2} = \frac{\alpha l}{\mu r}$$

Substituting Equations 8.27 and 8.28 into Equation 8.14 gives

$$E = \frac{1}{2}\mu\left(\frac{\alpha l}{\mu r}\right)^2 + \frac{l^2}{2\mu r^2} - \frac{l^2(\alpha^2 + 1)}{2\mu r^2} \tag{8.29}$$

$$E = 0$$

The total energy of the orbit is zero if $U(r = \infty) = 0$.

8.5 Orbits in a Central Field

The radial velocity of a particle moving in a central field is given by Equation 8.15. This equation indicates that $\dot{r}$ vanishes at the roots of the radical, that is, at points for which

$$E - U(r) - \frac{l^2}{2\mu r^2} = 0 \tag{8.30}$$

The vanishing of $\dot{r}$ implies that a turning point in the motion has been reached (see Section 2.6). In general, Equation 8.30 possesses two roots: r_{max} and r_{min}. The motion of the particle is therefore confined to the annular region specified by $r_{max} \geq r \geq r_{min}$. Certain combinations of the potential function U(r) and the parameters E and l produce only a single root for Equation 8.30. In such a case, $\dot{r} = 0$ for all values of the time; hence, r = constant, and the orbit is circular.

If the motion of a particle in the potential U(r) is periodic, then the orbit is *closed*; that is, after a finite number of excursions between the radial limits r_{min} and r_{max} the motion exactly repeats itself. But if the orbit does not close on itself after a finite number of oscillations, the orbit is said to be *open* (Figure 8-4).

From Equation 8.17, we can compute the change in the angle θ that results from one complete transit of r from r_{min} to r_{max} and back to r_{min}. Because the motion is symmetric in time, this angular change is twice that which would result from the passage from r_{min} to r_{max}; thus

$$\Delta\theta = 2\int_{r_{min}}^{r_{max}} \frac{(l/r^2)\,dr}{\sqrt{2\mu\left(E - U - \dfrac{l^2}{2\mu r^2}\right)}} \tag{8.31}$$

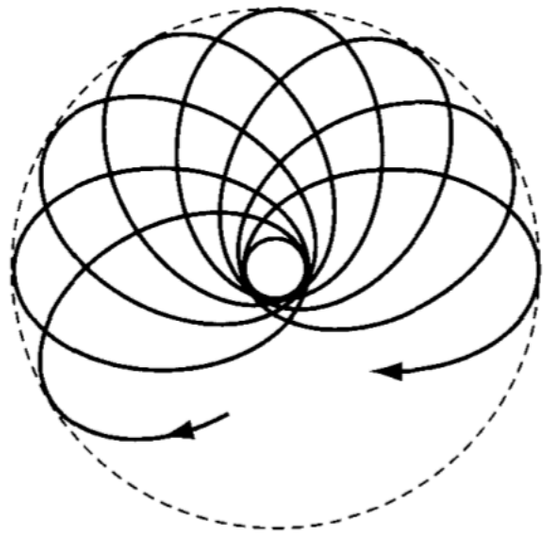


FIGURE 8-4 An orbit that does not close on itself after a finite number of oscillations is said to be *open*.

The path is closed only if $\Delta\theta$ is a rational fraction of 2π - that is, if , where a and b are integers. Under these conditions, after b periods the radius vector of the particle will have made a complete revolutions and will have returned to its original position. One can show that if the potential varies with some integer power of the radial distance, $\Delta\theta = 2\pi(a/b)$ then a closed noncircular path can result only if n = -2 or +1. The case n = -2 corresponds to an inverse-square-law force - for example, the gravitational or electrostatic force. The n = +1 case corresponds to the harmonic oscillator potential. For the two-dimensional case discussed in Section 3.4, we found that a closed path for the motion resulted if the ratio of the angular frequencies for the x and y motions were rational.

8.6 Centrifugal Energy and the Effective Potential

In the preceding expressions for $\dot{r}$, $\Delta\theta$, and so forth, a common term is the radical

$$\sqrt{E - U - \frac{l^2}{2\mu r^2}}$$

The last term in the radical has the dimensions of energy and, according to Equation 8.10, can also be written as

$$\frac{l^2}{2\mu r^2} = \frac{1}{2}\mu r^2\dot{\theta}^2$$

If we interpret this quantity as a "potential energy”,

$$U_c \equiv \frac{l^2}{2\mu r^2} \tag{8.32}$$

then the "force" that must be associated with U_c is

$$F_c = -\frac{\partial U_c}{\partial r} = \frac{l^2}{\mu r^3} = \mu r\dot{\theta}^2 \tag{8.33}$$

This quantity is traditionally called the **centrifugal force**, although it is not a force in the ordinary sense of the word(see later). We shall, however, continue to use this unfortunate terminology, because it is customary and convenient.

We see that the term $l^2/2\mu r^2$ can be interpreted as the *centrifugal potential energy* of the particle and, as such, can be included with U(r) in an *effective potential energy* defined by

$$\boxed{V(r) \equiv U(r) + \frac{l^2}{2\mu r^2}} \tag{8.34}$$

V(r) is therefore a *fictitious* potential that combines the real potential function U(r) with the energy term associated with the angular motion about the center of force. For the case of inverse-square-law central-force motion, the force is given by

$$F(r) = -\frac{k}{r^2} \tag{8.35}$$

from which

$$U(r) = -\int F(r)\,dr = -\frac{k}{r} \tag{8.36}$$

The effective potential function for gravitational attraction is therefore

$$V(r) = -\frac{k}{r} + \frac{l^2}{2\mu r^2} \tag{8.37}$$

This effective potential and its components are shown in Figure 8-5.

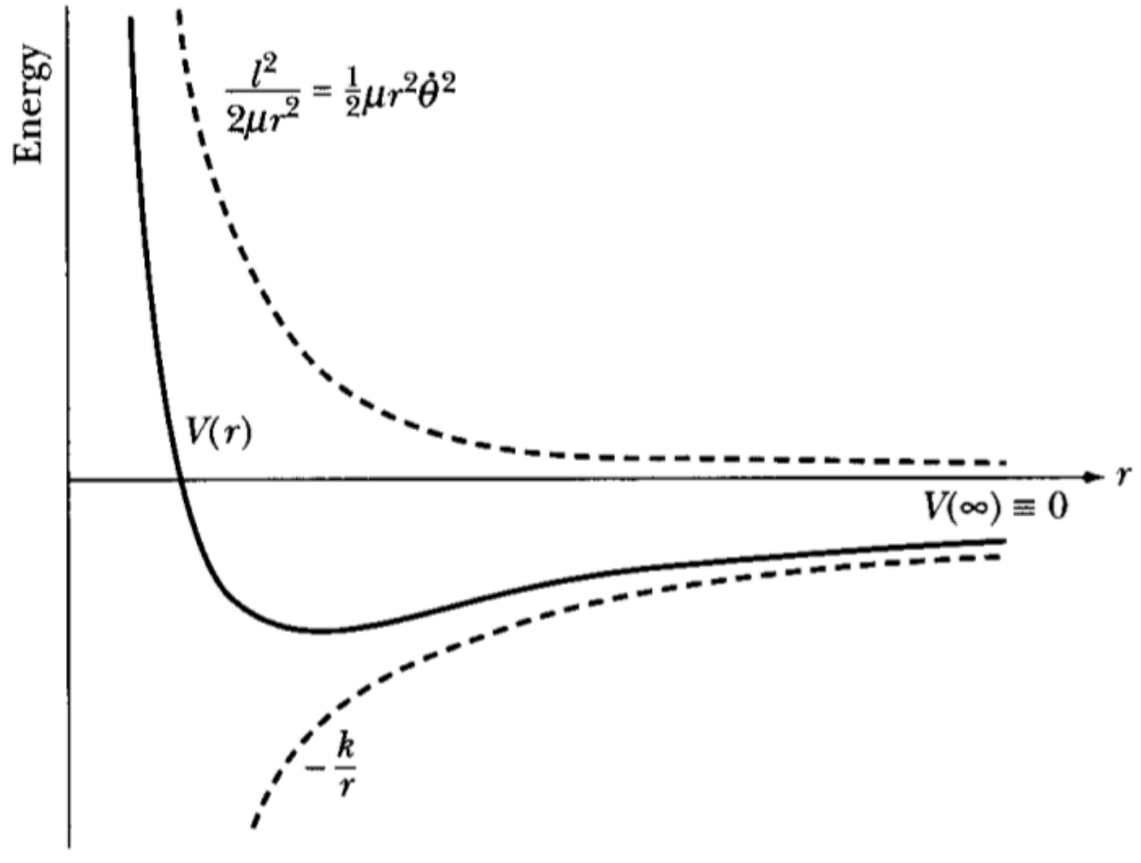


FIGURE 8-5 The effective potential for gravitational attraction $V(r)$ is composed of the real potential $-k/r$ term and the centrifugal potential energy $l^2/2\mu r^2$.

The value of the potential is arbitrarily taken to be zero at r = ∞ . (This is implicit in Equation 8.36, where we omitted the constant of integration.)

We may now draw conclusions similar to those in Section 2.6 on the motion of a particle in an arbitrary potential well. If we plot the total energy E of the particle on a diagram similar to Figure 8-5, we may identify three regions of interest (see Figure 8-6).

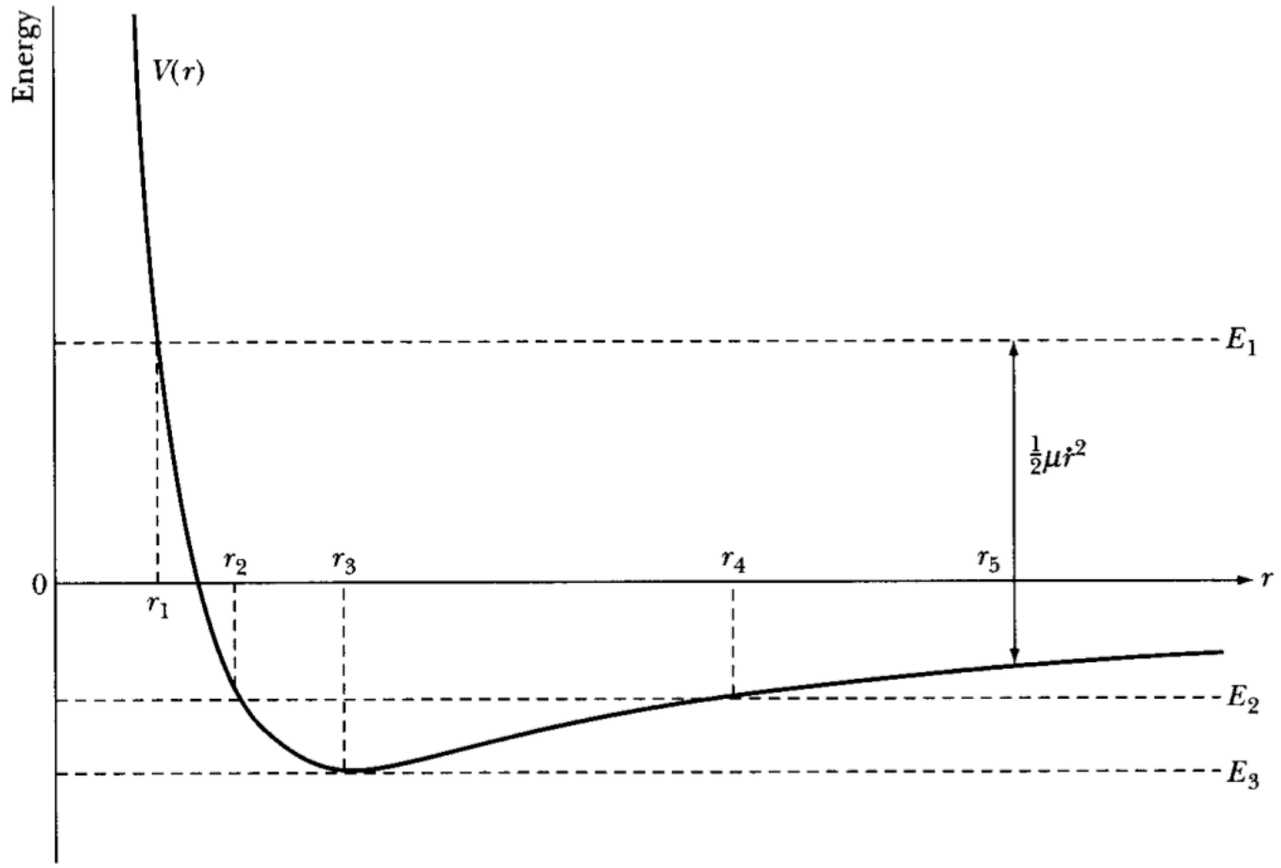


FIGURE 8-6 We can tell much about motion by looking at the total energy E on a potential energy plot. For example, for energy E_1 the particle's motion is unbounded. For energy E_2 the particle is bounded with $r_2 \leq r \leq r_4$. For energy E_3 the motion has $r = r_3$ and is circular.

If the total energy is positive or zero (e.g., $E_1 \geq 0$), then the motion is unbounded; the particle moves toward the force center (located at r = 0) from infinitely far away until it "strikes" the potential barrier at the turning point r = r_1 and is reflected back toward infinitely large r. Note that the height of the constant total energy line above V(r) at any r, such as r_5 in Figure 8-6, is equal to $\frac{1}{2}\mu\dot{r}^2$. Thus the radial velocity $\dot{r}$ vanishes and changes sign at the turning point (or points).

If the total energy is negative and lies between zero and the minimum value of V(r), as does E_2, then the motion is bounded, with $r_2 \leq r \leq r_4$.

The values r_2 and r_4 are the turning points, or the **apsidal distances**, of the orbit. If E equals the minimum value of the effective potential energy (see E_3 in Figure 8-6), then the radius of the particle's path is limited to the single value r_3, and then $\dot{r} = 0$ for all values of the time; hence the motion is circular.

Values of E less than $V_{min} = -(\mu k^2/2l^2)$ do not result in physically real motion; for such cases $\dot{r}^2 < 0$ and the velocity is imaginary.

The methods discussed in this section are often used in present-day research in general fields, especially atomic, molecular, and nuclear physics. For example, Figure 8-7 shows effective total nucleus-nucleus potentials for the scattering of ^{28}Si and ^{12}C.

The total potential includes the coulomb, nuclear, and the centrifugal contributions. The potential for $l = 0\hbar$ indicates the potential with no centrifugal term. For a relative angular momentum value of $l = 20\hbar$, a "pocket" exists where the two scattering nuclei may be bound together (even if only for a short time). For $l = 25\hbar$, the centrifugal "barrier" dominates, and the nuclei cannot form a bound state at all.

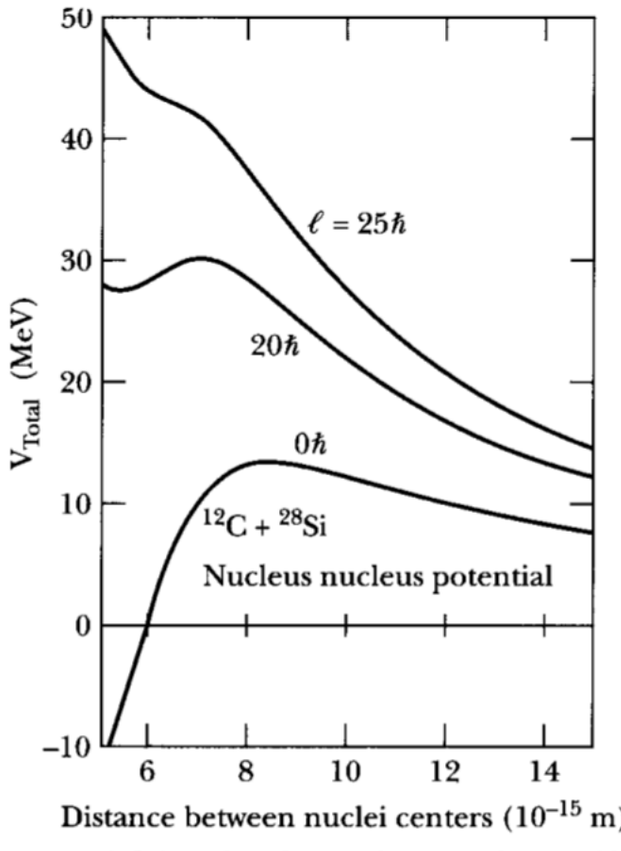


FIGURE 8-7 The total potential (coulomb, nuclear, and centrifugal) for scattering ^{28}Si nuclei from ^{12}C for various angular momentum l values as a function of distance between nuclei. For $l = 20\hbar$ a shallow pocket exists where the two nuclei may be bound together for a short time. For $l = 25\hbar$ the nuclei are not bound together.

8.7 Planetary Motion - Kepler's Problem

The equation for the path of a particle moving under the influence of a central force whose magnitude is inversely proportional to the square of the distance between the particle and the force center can be obtained (see Equation 8.17) from

$$\theta(r) = \int \frac{(l/r^2)\,dr}{\sqrt{2\mu\left(E + \frac{k}{r} - \frac{l^2}{2\mu r^2}\right)}} + \text{constant} \tag{8.38}$$

The integral can be evaluated if the variable is changed to $u = l/r$. If we define the origin of θ so that the minimum value of r is at $\theta = 0$, we find

$$\cos\theta = \frac{\frac{l^2}{\mu k}\cdot\frac{1}{r} - 1}{\sqrt{1 + \frac{2El^2}{\mu k^2}}} \tag{8.39}$$

Let us now define the following constants:

$$\left.\begin{aligned} \alpha &\equiv \frac{l^2}{\mu k} \\ \varepsilon &\equiv \sqrt{1 + \frac{2El^2}{\mu k^2}} \end{aligned}\right\} \tag{8.40}$$

Equation 8.39 can thus be written as

$$\boxed{\frac{\alpha}{r} = 1 + \varepsilon\cos\theta} \tag{8.41}$$

This is the equation of a conic section with one focus at the origin. The quantity ε is called the **eccentricity**, and 2α is termed the **latus rectum** of the orbit. Conic sections are formed by the intersection of a plane and a cone. A conic section is formed by the loci of points (formed in a plane), where the ratio of the distance from a fixed point (the focus) to a fixed line (called the directrix) is a constant. The directrix for the parabola is shown in Figure 8-8 by the vertical dashed line, drawn so that $r/r' = 1$.

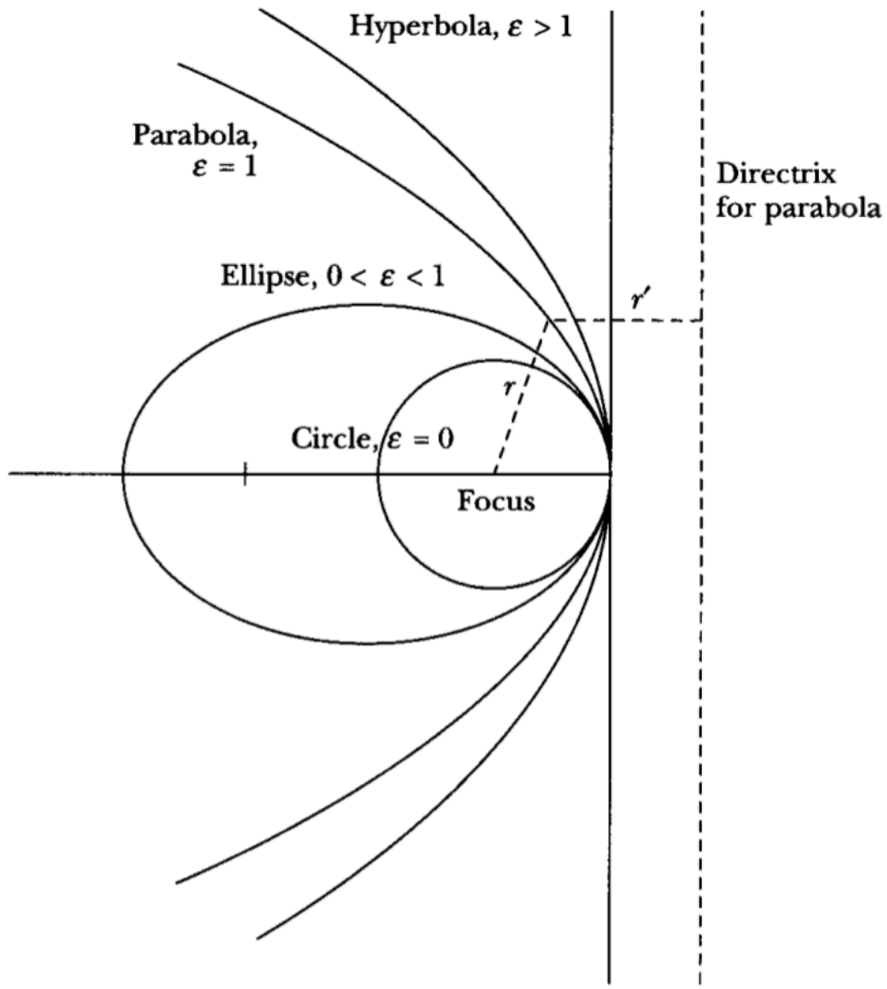


FIGURE 8-8 The orbits of the various conic sections are shown together with their eccentricities ε.

The minimum value for r in Equation 8.41 occurs when $\theta = 0$, or when $\cos\theta$ is a maximum. Thus the choice of the integration constant in Equation 8.38 corresponds to measuring θ from r_{min}, which position is called the **pericenter**; r_{max} corresponds to the **apocenter**. The general term for turning points is **apsides**. The corresponding terms for motion about the Sun are *perihelion* and *aphelion*, and for motion about Earth, *perigee* and *apogee*.

Various values of the eccentricity (and hence of the energy E) classify the orbits according to different conic sections (see Figure 8-8)

$\varepsilon > 1,$	$E > 0$	Hyperbola
$\varepsilon = 1,$	$E = 0$	Parabola
$0 < \varepsilon < 1,$	$V_{\min} < E < 0$	Ellipse
$\varepsilon = 0,$	$E = V_{\min}$	Circle

For planetary motion, the orbits are ellipses with major and minor axes (equal to 2a and 2b, respectively) given by

$$a = \frac{\alpha}{1 - \varepsilon^2} = \frac{k}{2|E|} \tag{8.42}$$

$$b = \frac{\alpha}{\sqrt{1 - \varepsilon^2}} = \frac{l}{\sqrt{2\mu|E|}} \tag{8.43}$$

Thus, the major axis depends only on the energy of the particle, whereas the minor axis is a function of both first integrals of the motion, E and L The geometry of elliptic orbits in terms of the parameters α, ε , a, and b is shown in Figure 8-9; P and P' are the foci.

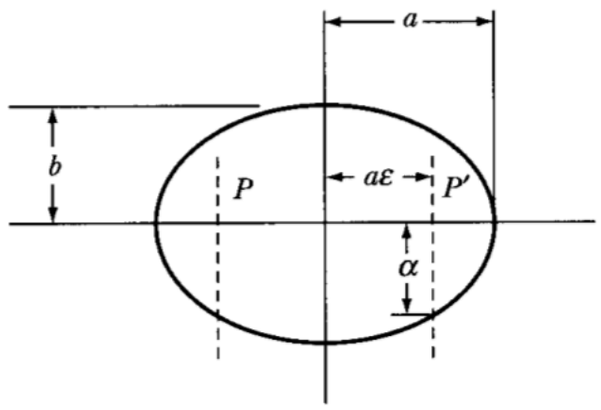


FIGURE 8-9 The geometry of elliptic orbits is shown in terms of parameters α, ε, a, and b. P and P' are the foci.

From this diagram, we see that the apsidal distances (r_{min} and r_{max} as measured from the foci to the orbit) are given by

$$\left.\begin{aligned} r_{\min} &= a(1-\varepsilon) = \frac{\alpha}{1+\varepsilon} \\ r_{\max} &= a(1+\varepsilon) = \frac{\alpha}{1-\varepsilon} \end{aligned}\right\} \tag{8.44}$$

To find the period for elliptic motion, we rewrite Equation 8.12 for the areal velocity as

$$dt = \frac{2\mu}{l}\,dA$$

Because the entire area A of the ellipse is swept out in one complete period τ ,

$$\int_0^{\tau} dt = \frac{2\mu}{l}\int_0^{A} dA \tag{8.45}$$

$$\tau = \frac{2\mu}{l}\,A$$

The area of an ellipse is given by $A = \pi ab$, and using a and b from Equations 8.42 and 8.43, we find

$$\tau = \frac{2\mu}{l}\cdot \pi ab = \frac{2\mu}{l}\cdot \pi \cdot \frac{k}{2|E|}\cdot \frac{l}{\sqrt{2\mu|E|}}$$

$$= \pi k\sqrt{\frac{\mu}{2}}\cdot |E|^{-3/2} \tag{8.46}$$

We also note from Equations 8.42 and 8.43 that the semiminor axis (the quantities a and b are called semimajor and semiminor axes, respectively.) can be written as

$$b = \sqrt{\alpha a} \tag{8.47}$$

Therefore, because $\alpha = l^2/\mu k$, the period τ can also be expressed as

$$\boxed{\tau^2 = \frac{4\pi^2\mu}{k} a^3} \tag{8.48}$$

This result, that the square of the period is proportional to the cube of the semimajor axis of the elliptic orbit, is known as **Kepler's Third Law**. Note that this result is concerned with the equivalent one-body problem, so account must be taken of the fact that it is the *reduced* mass μ that occurs in Equation 8.48. Kepler actually concluded that the squares of the periods of the planets were proportional to the cubes of the major axes of their orbits - with the same proportionality constant for all planets. In this sense, the statement is only approximately correct, because the reduced mass is different for each planet. In particular, because the gravitational force is given by

$$F(r) = -\frac{Gm_1m_2}{r^2} = -\frac{k}{r^2}$$

we identify $k = Gm_1m_2$. The expression for the square of the period therefore becomes

$$\tau^2 = \frac{4\pi^2a^3}{G(m_1 + m_2)} \cong \frac{4\pi^2a^3}{Gm_2}, \qquad m_1 \ll m_2 \tag{8.49}$$

and Kepler's statement is correct only if the mass m_1 of a planet can be neglected with respect to the mass m_2 of the Sun. (But note, for example, that the mass of Jupiter is about 1/1000 of the mass of the Sun, so the departure from the approximate law is not difficult to observe in this case.)

Kepler's laws can now be summarized:

I. *Planets move in elliptical orbits about the Sun with the Sun at one focus.*

II. *The area per unit time swept out by a radius vector from the Sun to a planet is constant.*

III. *The square of a planet's period is proportional to the cube of the major axis of the planet's orbit.*

See Table 8-1 for some properties of the principal objects in the solar system.

TABLE 8-1 Some Properties of the Principal Objects in the Solar System

Name	Semimajor axis of orbit (in astronomical units[a])	Period (yr)	Eccentricity	Mass (in units of Earth's mass[b])
Sun	—	—	—	332,830
Mercury	0.3871	0.2408	0.2056	0.0552
Venus	0.7233	0.6152	0.0068	0.814
Earth	1.0000	1.0000	0.0167	1.000
Eros (asteroid)	1.4583	1.7610	0.2229	2×10^{-9} (?)
Mars	1.5237	1.8809	0.0934	0.1074
Ceres (asteroid)	*c*	4.6035	0.0789	1/8000 (?)
Jupiter	5.2028	*c*	0.0483	317.89
Saturn	9.5388	29.456	0.0560	*c*
Uranus	19.191	84.07	0.0461	14.56
Neptune	30.061	164.81	0.0100	17.15
Pluto	39.529	248.53	0.2484	0.002
Halley (comet)	18	76	0.967	$\sim 10^{-10}$

[a] One astronomical unit (A.U.) is the length of the semimajor axis of Earth's orbit. One A.U. $\cong 1.495 \times 10^{11}$ m $\cong 93 \times 10^{6}$ miles.

[b] Earth's mass is approximately 5.976×10^{24} kg.

EXAMPLE 8.4

Halley's comet, which passed around the sun early in 1986, moves in a highly elliptical orbit with an eccentricity of 0.967 and a period of 76 years. Calculate its minimum and maximum distances from the Sun.

Solution. Equation 8.49 relates the period of motion with the semimajor axes. Because m (Halley's comet) $\ll m_{sun}$,

$$a = \left(\frac{Gm_{\text{Sun}}\tau^2}{4\pi^2}\right)^{1/3}$$

$$= \left[\frac{\left(6.67 \times 10^{-11}\,\frac{\text{Nm}^2}{\text{kg}^2}\right)(1.99 \times 10^{30}\,\text{kg})\left(76\,\text{yr}\,\frac{365\,\text{day}}{\text{yr}}\,\frac{24\,\text{hr}}{\text{day}}\,\frac{3600\,\text{s}}{\text{hr}}\right)^2}{4\pi^2}\right]^{1/3}$$

$$a = 2.68 \times 10^{12}\ \text{m}$$

Using Equation 8.44, we can determine r_{min} and r_{max}.

$$r_{\text{min}} = 2.68 \times 10^{12}\ \text{m}(1 - 0.967) = 8.8 \times 10^{10}\ \text{m}$$
$$r_{\text{max}} = 2.68 \times 10^{12}\ \text{m}(1 + 0.967) = 5.27 \times 10^{12}\ \text{m}$$

This orbit takes the comet inside the path of Venus, almost to Mercury's orbit, and out past even the orbit of Neptune and sometimes even to the moderately eccentric orbit of Pluto. Edmond Halley is generally given the credit for bringing Newton's work on gravitational and central forces to the attention of the world. After observing the comet personally in 1682, Halley became interested. Partly as a result of a bet between Christopher Wren and Robert Hooke, Halley asked Newton in 1684 what paths the planets must follow if the Sun pulled them with a force inversely proportional to the square of their distances. To the astonishment of Halley, Newton replied, "Why, in ellipses, of course." Newton had worked it out 20 years previously but had not published the result. With painstaking effort, Halley was able in 1705 to predict the next occurrence of the comet, now bearing his name, to be in 1758.

8.8 Orbital Dynamics

The use of central-force motion is nowhere more useful, important, and interesting than in space dynamics. Although space dynamics is actually quite complex because of the gravitational attraction of a spacecraft to various bodies and the orbital motion involved, we examine two rather simple aspects: a proposed trip to Mars and flybys past comets and planets.

Orbits are changed by single or multiple thrusts of the rocket engines. The simplest maneuver is a single thrust applied in the orbital plane that does not change the direction of the angular momentum but does change the eccentricity and energy simultaneously. The most economical method of interplanetary transfer consists of moving from one circular heliocentric (Sun-oriented motion) orbit to another in the same plane. Earth and Mars represent such a system reasonably well, and a Hohmann transfer (Figure 8-10) represents the path of minimum total energy expenditure.

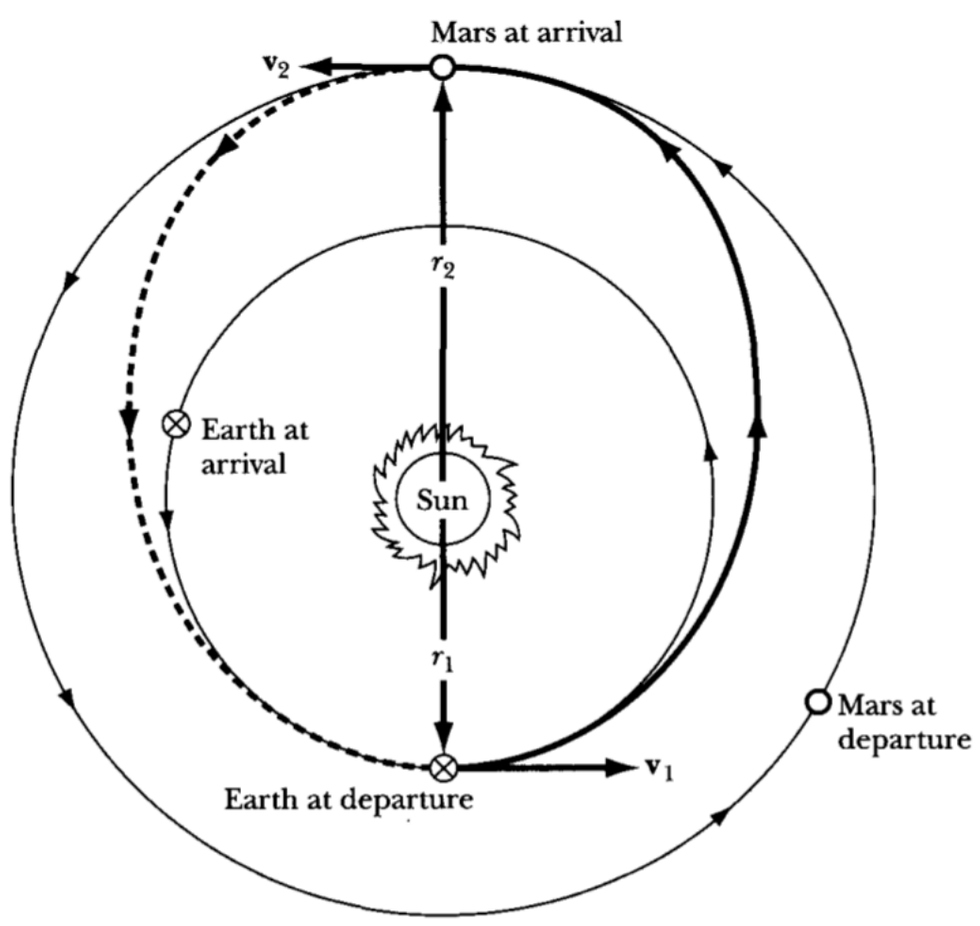


FIGURE 8-10 The Hohmann transfer for a round trip between Earth and Mars. It represents the minimum energy expenditure.

Two engine burns are required: A) the first burn injects the spacecraft from the circular Earth orbit to an elliptical transfer orbit that intersects Mars' orbit; B) the second burn transfers the spacecraft from the elliptical orbit into Mars' orbit.

We can calculate the velocity changes needed for a Hohmann transfer by calculating the velocity of a spacecraft moving in the orbit of Earth around the Sun (r_1 in Figure 8-10) and the velocity needed to "kick" it into an elliptical transfer orbit that can reach Mars' orbit. We are considering only the gravitational attraction of the Sun and not that of Earth and Mars.

For circles and ellipses we have, from Equation 8.42,

$$E = -\frac{k}{2a}$$

For a circular path around the Sun, this becomes

$$E = -\frac{k}{2r_1} = \frac{1}{2}mv_1^2 - \frac{k}{r_1} \tag{8.50}$$

where we have E = T+ U. We solve Equation 8.50 for v_1.

$$v_1 = \sqrt{\frac{k}{mr_1}} \tag{8.51}$$

We denote the semimajor axis of the transfer ellipse by a_t.

$$2a_t = r_1 + r_2$$

If we calculate the energy at the perihelion for the transfer ellipse, we have

$$E_t = \frac{-k}{r_1 + r_2} = \frac{1}{2}mv_{t1}^2 - \frac{k}{r_1} \tag{8.52}$$

where v_{t1} is the perihelion transfer speed. The direction of v_{t1} is along $\mathbf{v}_1$ in Figure 8-10. Solving Equation 8.52 for v_{t1} gives

$$v_{t1} = \sqrt{\frac{2k}{mr_1}\left(\frac{r_2}{r_1 + r_2}\right)} \tag{8.53}$$

The speed transfer Δv_1 needed is just

$$\Delta v_1 = v_{t1} - v_1 \tag{8.54}$$

Similarly, for the transfer from the ellipse to the circular orbit of radius r_2, we have

$$\Delta v_2 = v_2 - v_{t2} \tag{8.55}$$

where

$$v_2 = \sqrt{\frac{k}{mr_2}} \tag{8.56}$$

and

$$\left.\begin{aligned} v_{t2} &= \sqrt{\frac{2}{m}\left(E_t + \frac{k}{r_2}\right)} \\ v_{t2} &= \sqrt{\frac{2k}{mr_2}\left(\frac{r_1}{r_1 + r_2}\right)} \end{aligned}\right\} \tag{8.57}$$

The direction of v_{t2} is along $\mathbf{v}_2$ in Figure 8-10. The total speed increment can be determined by adding the speed changes, .

The total time required to make the transfer T_t is a half-period of the transfer orbit. From Equation 8.48, we have

$$\left.\begin{aligned} T_t &= \frac{\tau_t}{2} \\ T_t &= \pi\sqrt{\frac{m}{k}}\, a_t^{3/2} \end{aligned}\right\} \tag{8.58}$$

EXAMPLE 8.5

Calculate the time needed for a spacecraft to make a Hohmann transfer from Earth to Mars and the heliocentric transfer speed required assuming both planets are in coplanar orbits

Solution. We need to insert the appropriate constants in Equation 8.58.

$$\begin{aligned}\frac{m}{k} &= \frac{m}{GmM_{\text{Sun}}} = \frac{1}{GM_{\text{Sun}}} \\ &= \frac{1}{(6.67 \times 10^{-11}\,\text{m}^3/\text{s}^2 \cdot \text{kg})(1.99 \times 10^{30}\,\text{kg})} \\ &= 7.53 \times 10^{-21}\,\text{s}^2/\text{m}^3\end{aligned} \tag{8.59}$$

Because k/m occurs so often in solar system calculations, we write it as well.

$$\frac{k}{m} = 1.33 \times 10^{20}\ \text{m}^3/\text{s}^2$$

$$\begin{aligned}a_t &= \frac{1}{2}(r_{\text{Earth}-\text{Sun}} + r_{\text{Mars}-\text{Sun}}) \\ &= \frac{1}{2}(1.50 \times 10^{11}\,\text{m} + 2.28 \times 10^{11}\,\text{m}) \\ &= 1.89 \times 10^{11}\,\text{m} \\ T_t &= \pi(7.53 \times 10^{-21}\,\text{s}^2/\text{m}^3)^{1/2}(1.89 \times 10^{11}\,\text{m})^{3/2} \\ &= 2.24 \times 10^{7}\,\text{s} \\ &= 259\ \text{days}\end{aligned} \tag{8.60}$$

The heliocentric speed needed for the transfer is given in Equation 8.53.

$$v_{t1} = \left[\frac{2(1.33 \times 10^{20}\,\text{m}^3/\text{s}^2)(2.28 \times 10^{11}\,\text{m})}{(1.50 \times 10^{11}\,\text{m})(3.78 \times 10^{11}\,\text{m})}\right]^{1/2}$$

$$= 3.27 \times 10^{4}\ \text{m/s} = 32.7\ \text{km/s}$$

We can compare v_{t1} with the orbital speed of Earth (Equation 8.51).

$$v_1 = \left[\frac{1.33 \times 10^{20}\,\mathrm{m^3/s^2}}{1.50 \times 10^{11}\,\mathrm{m}}\right]^{1/2} = 29.8\ \mathrm{km/s}$$

For transfers to the outer planets, the spacecraft should be launched in the direction of Earth's orbit in order to gain Earth's orbital velocity. To transfer to the inner planets (e.g., to Venus), the spacecraft should be launched opposite Earth's motion. In each case, it is the relative velocity Δv_1 that is important to the spacecraft (i.e., relative to Earth).

Although the Hohmann transfer path represents the least energy expenditure, it does not represent the shortest time. For a round trip from Earth to Mars, the spacecraft would have to remain on Mars for 460 days until Earth and Mars were positioned correctly for the return trip (see Figure 8-11a).

The total trip B59 + 460 + 259 = 978 days = 2.7 yr) would probably be too long. Other schemes either use more fuel to gain speed (Figure 8-11b) or use the slingshot effect of flybys. Such a flyby mission past Venus (see Figure 8-11c) could be done in less than 2 years with only a few weeks near (or on) Mars.

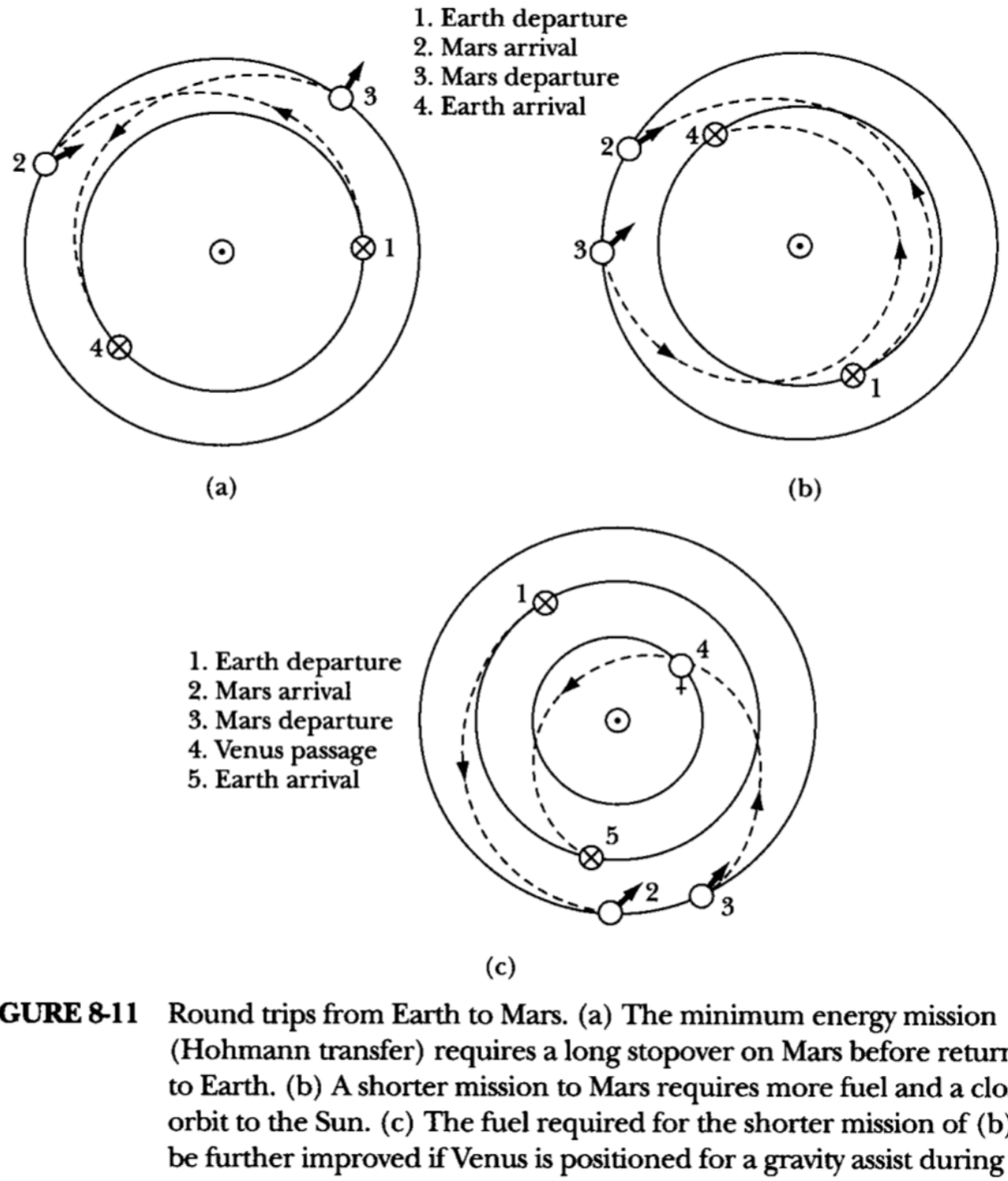


FIGURE 8-11 Round trips from Earth to Mars. (a) The minimum energy mission (Hohmann transfer) requires a long stopover on Mars before returning to Earth. (b) A shorter mission to Mars requires more fuel and a closer orbit to the Sun. (c) The fuel required for the shorter mission of (b) can be further improved if Venus is positioned for a gravity assist during flyby.

Several spacecraft in recent years have escaped Earth's gravitational attraction to explore our solar system. Such interplanetary transfer can be divided into three segments: A) the escape from Earth, B) a heliocentric transfer to the area of interest, and C) an encounter with another body - so far, either a planet or a comet. The spacecraft fuel required for such missions can be enormous, but a clever trick has been designed to "steal" energy from other solar system bodies. Because the mass of a spacecraft is so much smaller than the planets (or their moons), the energy loss of the heavenly body is negligible.

We examine a simple version of this flyby or slingshot effect that utilizes gravity assist. A spacecraft coming from infinity approaches a body (labeled B), interacts with B, and recedes. The path is a hyperbola (Figure 8-12). The initial and final velocities, *with respect to* B, are denoted by v_i' and v_f', respectively. The net effect on the spacecraft is a deflection angle of δ with respect to B.

If we examine the system in some inertial frame in which the motion of B occurs, the velocities of the spacecraft can be quite different *because of the motion of* B. The initial velocity v_i is shown in Figure 8-13a, and both v_i and v_f are shown in Figure 8-13b. Notice that the spacecraft has increased its speed as well as changed its direction. An increase in velocity occurs when the spacecraft passes behind B's direction of motion. Similarly, a decrease in velocity occurs when the spacecraft passes in *front* of B's motion.

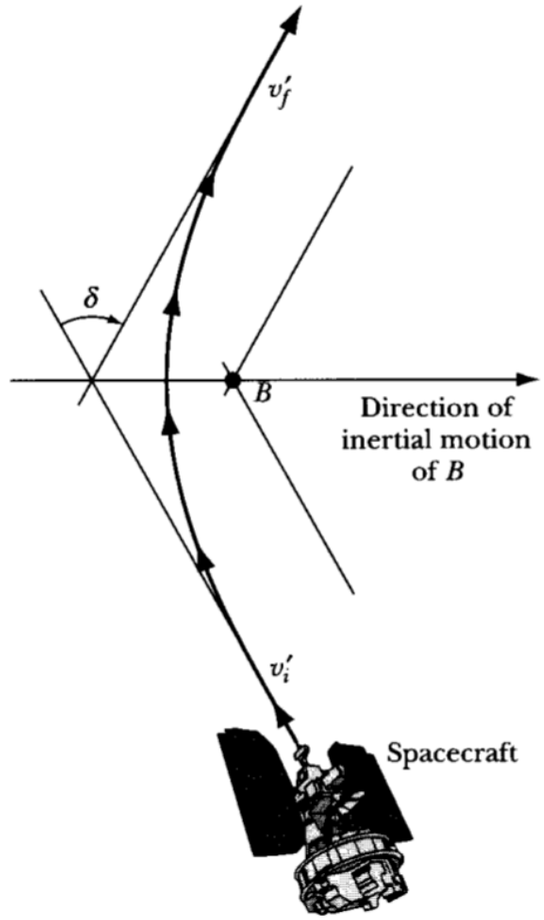


FIGURE 8-12 A spacecraft flies by a large body *B* (like a planet) and gains speed when it flies behind *B*'s direction of motion. Similarly, the spacecraft loses speed when it passes in front of *B*'s direction of motion. The direction of the spacecraft also changes.

During the 1970s, scientists at the Jet Propulsion Laboratory of the National Aeronautics and Space Administration (NASA) realized that the four largest planets of our solar system would be in a fortuitous position to allow a spacecraft to fly past them and many of their 32 known moons in a single, relatively short "Grand Tour" mission using the gravity-assist method just discussed. This opportunity of the planets' alignment would not occur again for 175 years.

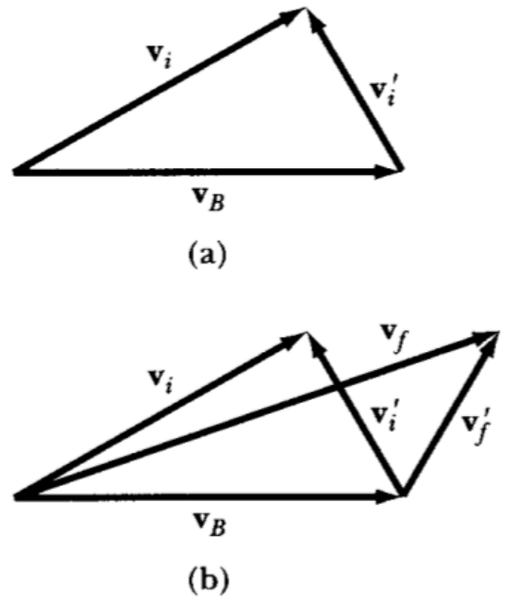


FIGURE 8-13 The vectors $\mathbf{v}_i'$ and $\mathbf{v}_f'$ are the initial and final velocities of the spacecraft with respect to B. The vectors $\mathbf{v}_i$ and $\mathbf{v}_f$ are the velocities in an inertial frame. (a) $\mathbf{v}_i = \mathbf{v}_B + \mathbf{v}_i'$. (b) $\mathbf{v}_f = \mathbf{v}_B + \mathbf{v}_f'$.

Because of budget constraints, there was not time to develop the new technology needed, and a mission to last only 4 years to visit just Jupiter and Saturn was approved and planned. No special equipment was put on board the twin *Voyager* spacecrafts for an encounter with Uranus and Neptune. *Voyagers 1 and 2* were launched in 1977 for visits to Jupiter in 1979 and Saturn in 1980 (*Voyager 1*) and 1981 (*Voyager 2*). Because of the success of these visits to Jupiter and Saturn, funding was later approved to extend *Voyager 2*'s mission to include Uranus and Neptune. The *Voyagers* are now on their way out of our solar system.

The path of *Voyager 2* is shown in Figure 8-14. The slingshot effect of gravity allowed the path of *Voyager 2* to be redirected, for example, toward Uranus as it passed Saturn by the method shown in Figure 8-12. The gravitational attraction from Saturn was used to pull the spacecraft off its straight path and redirect it at a different angle. The effect of the orbital motion of Saturn allows an increase in the spacecraft's speed. It was only by using this gravity-assist technique that the spectacular mission of *Voyager 2* was made possible in only a brief 12-year period.

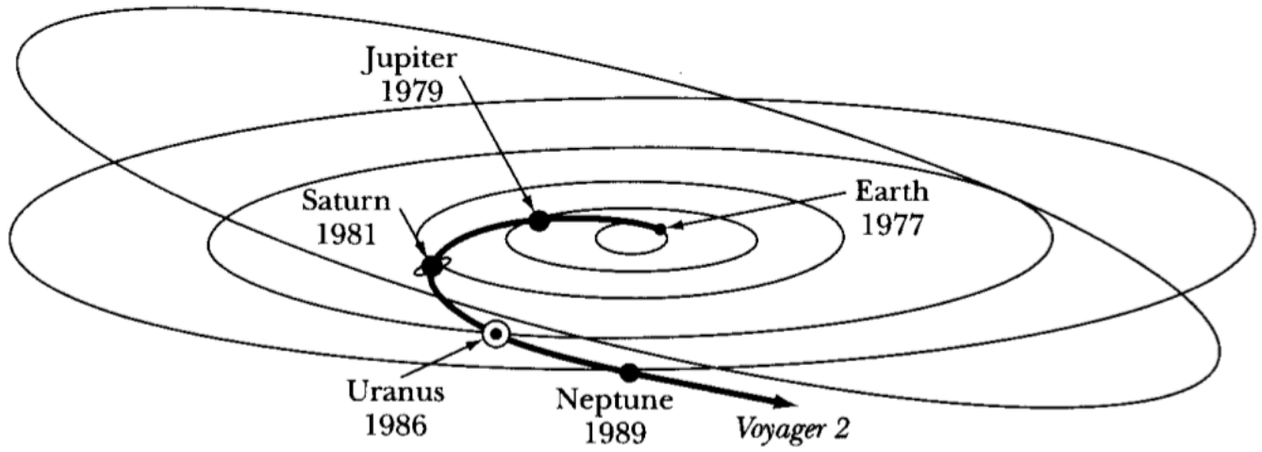


FIGURE 8-14 *Voyager 2* was launched in 1977 and passed by Jupiter, Saturn, Uranus, and Neptune. Gravitational assists were used in the mission.

Voyager 2 passed Uranus in 1986 and Neptune in 1989 before proceeding into interstellar space in one of the most successful space missions ever undertaken. Most planetary missions now take advantage of gravitational assists; for example, the Galileo satellite, which photographed the spectacular collisions of the Shoemaker-Levy comet with Jupiter in 1994 and reached Jupiter in 1995, was launched in 1989 but went by Earth twice (1990 and 1992) as well as Venus (1990) to gain speed and redirection.

A spectacular display of flybys occurred in the years 1982-1985 by a spacecraft initially called the International Sun-Earth Explorer 3 (ISEE-J). Launched in 1978, its mission was to monitor the solar wind between the Sun and Earth. For 4 years, the spacecraft circled in the elliptical plane about 2 million miles from Earth. In 1982 - because the United States had decided not to participate in a joint European, Japanese, and Soviet spacecraft investigation of Halley's comet in 1986 - NASA decided to reprogram the ISEE-3, renamed it the *International Cometary Explorer* (ICE), and sent it through the Giacobini-Zinner comet in September 1985, some 6 months before the flybys of other spacecraft with Halley's comet. The subsequent three-year journey of ICE was spectacular (Figure 8-15). The path of ICE included two close trips to Earth and five flybys of the moon along its billion-mile trip to the comet. During one flyby, the satellite came within 75 miles of the lunar surface. The entire path could be planned precisely because the force law is very well known. The eventual interaction with the comet, some 44 million miles from Earth, included a 20-minute trip through the comet - about 5,000 miles behind the comet's nucleus.

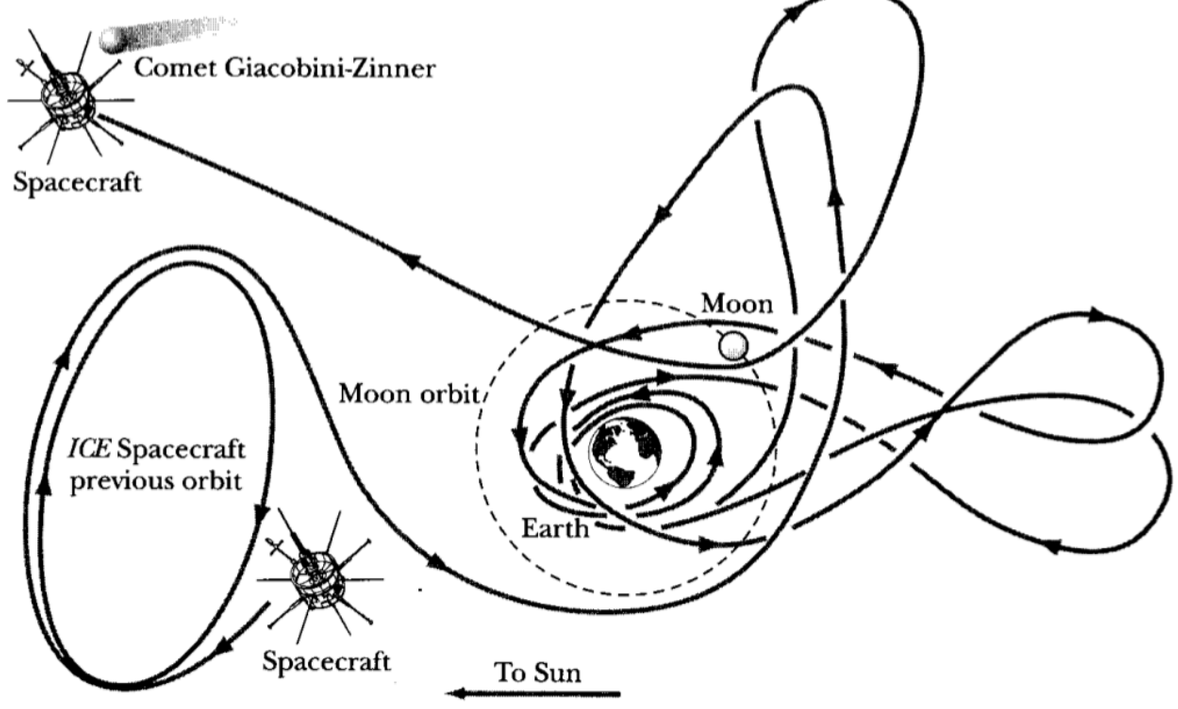


FIGURE 8-15 The NASA spacecraft initially called *ISEE-3* was reprogrammed to be the International Cometary Explorer and was sent on a spectacular three-year journey utilizing gravity assists on its way by the Comet Giacobini-Zinner.

8.9 Apsidal Angles and Precession

If a particle executes bounded, noncircular motion in a central-force field, then the radial distance from the force center to the particle must always be in the range $r_{max} \geq r \geq r_{min}$; that is, r must be bounded by the apsidal distances. Figure 8-6 indicates that only two apsidal distances exist for bounded, noncircular motion. But in executing one complete revolution in θ, the particle may not return to its original position (see Figure 8-4). The angular separation between two successive values of $r = r_{max}$ depends on the exact nature of the force. The angle between any two consecutive apsides is called the **apsidal angle**, and because a closed orbit must be symmetric about any apsis, it follows that all apsidal angles for such motion must be equal. The apsidal angle for elliptical motion, for example, is just π . If the orbit is not closed, the particle reaches the apsidal distances at different points in each revolution; the apsidal angle is not then a rational fraction of 2π, as is required for a closed orbit. If the orbit is *almost* closed, the apsides precess, or rotate slowly in the plane of the motion. This effect is exactly analogous to the slow rotation of the elliptical motion of a two-dimensional harmonic oscillator whose natural frequencies for the x and y motions are almost equal (see Section 3.3).

Because an inverse-square-law force requires that all elliptical orbits be exactly closed, the apsides must stay fixed in space for all time. If the apsides move with time, however slowly, this indicates that the force law under which the body moves does not vary exactly as the inverse square of the distance. This important fact was realized by Newton, who pointed out that any advance or regression of a planet's perihelion would require the radial dependence of the force law to be slightly different from $1/r^2$. Thus, Newton argued, the observation of the time dependence of the perihelia of the planets would be a sensitive test of the validity of the form of the universal gravitation law.

In point of fact, for planetary motion within the solar system, one expects that, because of the perturbations introduced by the existence of all the other planets, the force experienced by any planet does not vary exactly as $1/r^2$, if r is measured from the Sun. This effect is small, however, and only slight variations of planetary perihelia have been observed. The perihelion of Mercury, for example, which shows the largest effect, advances only about 574" of arc per century. This precession is in addition to the general precession of the equinox with respect to the "fixed" stars, which amounts to 5025.645" ± 0.050" per century Detailed calculations of the influence of the other planets on the motion of Mercury predict that the rate of advance of the perihelion should be approximately 531" per century. The uncertainties in this calculation are considerably less than the difference of 43" between observation and calculation, and for a considerable time, this discrepancy was the outstanding unresolved difficulty in the Newtonian theory. We now know that the modification introduced into the equation of motion of a planet by the general theory of relativity almost exactly accounts for the difference of 43". This result is one of the major triumphs of relativity theory.

We next indicate the way the advance of the perihelion can be calculated from the modified equation of motion. To perform this calculation, it is convenient to use the equation of motion in the form of Equation 8.20. If we use the universal gravitational law for F(r), we can write

$$\frac{d^2u}{d\theta^2} + u = -\frac{m}{l^2}\frac{1}{u^2}F(1/u) = \frac{Gm^2M}{l^2} \tag{8.61}$$

where we consider the motion of a body of mass m in the gravitational field of a body of mass M. The quantity u is therefore the reciprocal of the distance between m and M.

The modification of the gravitational force law required by the general theory of relativity introduces into the force a small component that varies as $l/r^4 (= u^4)$. Thus, we have

$$\frac{d^2u}{d\theta^2} + u = \frac{Gm^2M}{l^2} + \frac{3GM}{c^2}u^2 \tag{8.62}$$

where c is the velocity of propagation of the gravitational interaction and is identified with the velocity of light. To simplify the notation, we define

$$\left.\begin{aligned} \frac{1}{\alpha} &\equiv \frac{Gm^2M}{l^2} \\ \delta &\equiv \frac{3GM}{c^2} \end{aligned}\right\} \tag{8.63}$$

and we can write Equation 8.62 as

$$\frac{d^2u}{d\theta^2} + u = \frac{1}{\alpha} + \delta u^2 \tag{8.64}$$

This is a nonlinear equation, and we use a successive approximation procedure to obtain a solution. We choose the first solution to be the solution of Equation 8.64 in the case that the term δu^2 is neglected (we eliminate the necessity of introducing an arbitrary phase into the argument of the cosine term by choosing to measure θ from the position of perihelion; i.e., u_1 is a maximum (and hence r_1 is a minimum) at $\theta = 0$:

$$u_1 = \frac{1}{\alpha}(1 + \varepsilon \cos\theta) \tag{8.65}$$

This is the familiar result for the pure inverse-square-law force (see Equation 8.41). Note that α is here the same as that defined in Equation 8.40 except that μ, has been replaced by m. If we substitute this expression into the right-hand side of Equation 8.64, we find

$$\begin{aligned} \frac{d^2u}{d\theta^2} + u &= \frac{1}{\alpha} + \frac{\delta}{\alpha^2}[1 + 2\varepsilon\cos\theta + \varepsilon^2\cos^2\theta] \\ &= \frac{1}{\alpha} + \frac{\delta}{\alpha^2}\left[1 + 2\varepsilon\cos\theta + \frac{\varepsilon^2}{2}(1 + \cos 2\theta)\right] \end{aligned} \tag{8.66}$$

where $\cos^2\theta$ has been expanded in terms of $\cos 2\theta$. The first trial function u_1, when substituted into the left-hand side of Equation 8.64, reproduces only the first term on the right-hand side: $1/\alpha$. We can therefore construct a second trial function by adding to u_1 a term that reproduces the remainder of the right-hand side (in Equation 8.66). We can verify that such a particular integral is

$$u_p = \frac{\delta}{\alpha^2}\left[\left(1 + \frac{\varepsilon^2}{2}\right) + \varepsilon\theta\sin\theta - \frac{\varepsilon^2}{6}\cos 2\theta\right] \tag{8.67}$$

The second trial function is therefore $u_2 = u_1 + u_p$

If we stop the approximation procedure at this point, we have

$$\begin{aligned} u \cong u_2 &= u_1 + u_p \\ &= \left[\frac{1}{\alpha}(1 + \varepsilon\cos\theta) + \frac{\delta\varepsilon}{\alpha^2}\theta\sin\theta\right] \\ &\quad + \left[\frac{\delta}{\alpha^2}\left(1 + \frac{\varepsilon^2}{2}\right) - \frac{\delta\varepsilon^2}{6\alpha^2}\cos 2\theta\right] \end{aligned} \tag{8.68}$$

where we have regrouped the terms in u_1 and u_p.

Consider the terms in the second set of brackets in Equation 8.68: the first of these is just a constant, and the second is only a small and periodic disturbance of the normal Keplerian motion. Therefore, on a long time scale neither of these terms contributes, on the average, to any change in the positions of the apsides. But in the first set of brackets, the term proportional to θ produces secular and therefore observable effects. Let us consider the first set of brackets:

$$u_{\text{secular}} = \frac{1}{\alpha}\left[1 + \varepsilon \cos\theta + \frac{\delta\varepsilon}{\alpha}\theta \sin\theta\right] \tag{8.69}$$

Next, we can expand the quantity

$$1 + \varepsilon \cos\left(\theta - \frac{\delta}{\alpha}\theta\right) = 1 + \varepsilon\left(\cos\theta \cos\frac{\delta}{\alpha}\theta + \sin\theta \sin\frac{\delta}{\alpha}\theta\right) \tag{8.70}$$

$$\cong 1 + \varepsilon \cos\theta + \frac{\delta\varepsilon}{\alpha}\theta \sin\theta$$

where we have used the fact that δ is small to approximate

$$\cos\frac{\delta}{\alpha}\theta \cong 1, \qquad \sin\frac{\delta}{\alpha}\theta \cong \frac{\delta}{\alpha}\theta$$

Hence, we can write u_{secular} as

$$u_{\text{secular}} \cong \frac{1}{\alpha}\left[1 + \varepsilon \cos\left(\theta - \frac{\delta}{\alpha}\theta\right)\right] \tag{8.71}$$

We have chosen to measure θ from the position of perihelion at t = 0. Successive appearances at perihelion result when the argument of the cosine
term in u_{secular} increases to 2π , 4π ,... , and so forth. But an increase of the argument by 2π requires that

$$\theta - \frac{\delta}{\alpha}\theta = 2\pi$$

or

$$\theta = \frac{2\pi}{1 - (\delta/\alpha)} \cong 2\pi\left(1 + \frac{\delta}{\alpha}\right)$$

Therefore, the effect of the relativistic term in the force law is to displace the perihelion in each revolution by an amount

$$\Delta \cong \frac{2\pi\delta}{\alpha} \tag{8.72a}$$

that is, the apsides rotate slowly in space. If we refer to the definitions of α and δ (Equations 8.63), we find

$$\Delta \cong 6\pi\left(\frac{GmM}{cl}\right)^2 \tag{8.72b}$$

TABLE 8-2 Precessional Rates for the Perihelia of Some Planets

	Precessional rate (seconds of arc/century)	
Planet	**Calculated**	**Observed**
Mercury	43.03 ± 0.03	43.11 ± 0.45
Venus	8.63	8.4 ± 4.8
Earth	3.84	5.0 ± 1.2
Mars	1.35	—
Jupiter	0.06	—

From Equations 8.40 and 8.42, we can write $l^2 = \mu k a(1 - \varepsilon^2)$: then, because k = GmM and $\mu \simeq m$, we have

$$\boxed{\Delta \cong \frac{6\pi GM}{ac^2(1 - \varepsilon^2)}}$$

We see therefore that the effect is enhanced if the semimajor axis a is small and if the eccentricity is large. Mercury, which is the planet nearest the sun and which has the most eccentric orbit of any planet (except Pluto), provides the most sensitive test of the theory. Alternatively, we can say that the relativistic advance of the perihelion is a maximum for Mercury because the orbital velocity is greatest for Mercury and the relativistic parameter v/c largest. The calculated value of the precessional rate for Mercury is 43.03" ± 0.03" of arc per century. The observed value (corrected for the influence of the other planets) is 43.11" ± 0.45", so the prediction of relativity theory is confirmed in striking fashion. The precessional rates for some of the planets are given in Table 8-2.

8.10 Stability of Circular Orbits

In Section 8.6, we pointed out that the orbit is circular if the total energy equals the minimum value of the effective potential energy, $E = V_{min}$. More generally, however, a circular orbit is allowed for any attractive potential, because the attractive force can always be made to just balance the centrifugal force by the proper choice of radial velocity. Although circular orbits are therefore always possible in a central, attractive force field, such orbits are not necessarily stable. A circular orbit at r = ρ exists if $\dot{r}|_{r=\rho} = 0$ for all t; this is possible if $(\partial V/\partial r)|_{r=\rho} = 0$. But only if the effective potential has a true minimum does stability result. All other equilibrium circular orbits are unstable.

Let us consider an attractive central force with the form

$$F(r) = -\frac{k}{r^n} \tag{8.73}$$

The potential for such a force is

$$U(r) = -\frac{k}{n-1} \cdot \frac{1}{r^{(n-1)}} \tag{8.74}$$

and the effective potential function is

$$V(r) = -\frac{k}{n-1} \cdot \frac{1}{r^{(n-1)}} + \frac{l^2}{2\mu r^2} \tag{8.75}$$

The conditions for a minimum of V(r) and hence for a stable circular orbit with a radius ρ are

$$\boxed{\left.\frac{\partial V}{\partial r}\right|_{r=\rho} = 0 \quad \text{and} \quad \left.\frac{\partial^2 V}{\partial r^2}\right|_{r=\rho} > 0} \tag{8.76}$$

Applying these criteria to the effective potential of Equation 8.75, we have

$$\left.\frac{\partial V}{\partial r}\right|_{r=\rho} = \frac{k}{\rho^n} - \frac{l^2}{\mu\rho^3} = 0$$

or

$$\rho^{(n-3)} = \frac{\mu k}{l^2} \tag{8.77}$$

and

$$\left.\frac{\partial^2 V}{\partial r^2}\right|_{r=\rho} = -\frac{nk}{\rho^{(n+1)}} + \frac{3l^2}{\mu\rho^4} > 0$$

so

$$-\frac{nk}{\rho^{(n-3)}} + \frac{3l^2}{\mu} > 0 \tag{8.78}$$

Substituting $\rho^{(n-3)}$ from Equation 8.77 into Equation 8.78, we have

$$(3 - n)\frac{l^2}{\mu} > 0 \tag{8.79}$$

The condition that a stable circular orbit exists is thus n < 3.

Next, we apply a more general procedure and inquire about the frequency of oscillation about a circular orbit in a general force field. We write the force as

$$F(r) = -\mu g(r) = -\frac{\partial U}{\partial r} \tag{8.80}$$

Equation 8.18 can now be written as

$$\ddot{r} - r\dot{\theta}^2 = -g(r) \tag{8.81}$$

Substituting for $\dot{\theta}$ from Equation 8.10,

$$\ddot{r} - \frac{l^2}{\mu^2 r^3} = -g(r) \tag{8.82}$$

We now consider the particle to be initially in a circular orbit with radius ρ and apply a perturbation of the form $r \to \rho + x$, where x is small. Because ρ = constant, we also have $\ddot{r} \to \ddot{\rho}$. Thus

$$\ddot{x} - \frac{l^2}{\mu^2 \rho^3 [1 + (x/\rho)]^3} = -g(\rho + x) \tag{8.83}$$

But by hypothesis $(x/\rho) \ll 1$, so we can expand the quantity:

$$[1 + (x/\rho)]^{-3} = 1 - 3(x/\rho) + \cdots \tag{8.84}$$

We also assume that g(r) = g(ρ + x) can be expanded in a Taylor series about the point r = ρ :

$$g(\rho + x) = g(\rho) + xg'(\rho) + \cdots \tag{8.85}$$

where

$$g'(\rho) \equiv \left.\frac{dg}{dr}\right|_{r=\rho}$$

If we neglect all terms in x^2 and higher powers, then the substitution of Equations 8.84 and 8.85 into Equation 8.83 yields

$$\ddot{x} - \frac{l^2}{\mu^2 \rho^3}[1 - 3(x/\rho)] \cong -[g(\rho) + xg'(\rho)] \tag{8.86}$$

Recall that we assumed the particle to be initially in a circular orbit with $r = \rho$. Under such a condition, no radial motion occurs - that is, $\dot{r}|_{r=\rho} = 0$. Then, also, $\ddot{r}|_{r=\rho} = 0$. Therefore, evaluating Equation 8.82 at $r = \rho$, we have

$$g(\rho) = \frac{l^2}{\mu^2 \rho^3} \tag{8.87}$$

Substituting this relation into Equation 8.86, we have, approximately,

$$\ddot{x} - g(\rho)[1 - 3(x/\rho)] \cong -[g(\rho) + xg'(\rho)]$$

or

$$\ddot{x} + \left[\frac{3g(\rho)}{\rho} + g'(\rho)\right]x \cong 0 \tag{8.88}$$

If we define

$$\omega_0^2 \equiv \frac{3g(\rho)}{\rho} + g'(\rho) \tag{8.89}$$

then Equation 8.88 becomes the familiar equation for the undamped harmonic oscillator:

$$\ddot{x} + \omega_0^2 x = 0 \tag{8.90}$$

The solution to this equation is

$$x(t) = Ae^{+i\omega_0 t} + Be^{-i\omega_0 t} \tag{8.91}$$

If $\omega_0^2 < 0$, so that ω_0 is imaginary, then the second term becomes $B\exp(|\omega_0|t)$, which clearly increases without limit as time increases. The condition for oscillation is therefore $\omega_0^2 > 0$, or

$$\frac{3g(\rho)}{\rho} + g'(\rho) > 0 \tag{8.92a}$$

Because $g(\rho) > 0$ (see Equation 8.87), we can divide through by $g(\rho)$ and write this inequality as

$$\frac{g'(\rho)}{g(\rho)} + \frac{3}{\rho} > 0 \tag{8.92b}$$

or, because g(r) and F(r) are related by a constant multiplicative factor, stability results if

$$\boxed{\frac{F'(\rho)}{F(\rho)} + \frac{3}{\rho} > 0} \tag{8.93}$$

We now compare the condition on the force law imposed by Equation 8.93 with that previously obtained for a power-law force:

$$F(r) = -\frac{k}{r^n} \tag{8.94}$$

Equation 8.93 becomes

$$\frac{nk\rho^{-(n+1)}}{-k\rho^{-n}} + \frac{3}{\rho} > 0$$

or

$$(3 - n) \cdot \frac{1}{\rho} > 0 \tag{8.95}$$

and we are led to the same condition as before - that is, n < 3. (We must note, however, that the case n = 3 needs further examination.)

EXAMPLE 8.6

Investigate the stability of circular orbits in a force field described by the potential function

$$U(r) = \frac{-k}{r} e^{-(r/a)} \tag{8.96}$$

where k > 0 and a > 0.

Solution. This potential is called the **screened Coulomb potential** (when $k = Ze^2/4\pi\varepsilon_0$, where Z is the atomic number and e is the electron charge) because it falls off with distance more rapidly than 1/r and hence approximates the electrostatic potential of the atomic nucleus in the vicinity of the nucleus by taking into account the partial "cancellation" or "screening" of the nuclear by the atomic electrons. The force is found from

and

$$F(r) = -\frac{\partial U}{\partial r} = -k\left(\frac{1}{ar} + \frac{1}{r^2}\right)e^{-(r/a)}$$

$$\frac{\partial F}{\partial r} = k\left(\frac{1}{a^2 r} + \frac{2}{ar^2} + \frac{2}{r^3}\right)e^{-(r/a)}$$

The condition for stability (see Equation 8.93) is

$$3 + \rho\frac{F'(\rho)}{F(\rho)} > 0$$

Therefore

$$3 + \frac{\rho k\left(\frac{1}{a^2\rho} + \frac{2}{a\rho^2} + \frac{2}{\rho^3}\right)}{-k\left(\frac{1}{a\rho} + \frac{1}{\rho^2}\right)} > 0$$

which simplifies to

$$a^2 + a\rho - \rho^2 > 0$$

We may write this as

$$\frac{a^2}{\rho^2} + \frac{a}{\rho} - 1 > 0$$

Stability thus results for all $q \equiv a/\rho$ that exceed the value satisfying the equation

$$q^2 + q - 1 = 0$$

The positive (and therefore the only physically meaningful) solution is

$$q = \frac{1}{2}(\sqrt{5} - 1) \cong 0.62$$

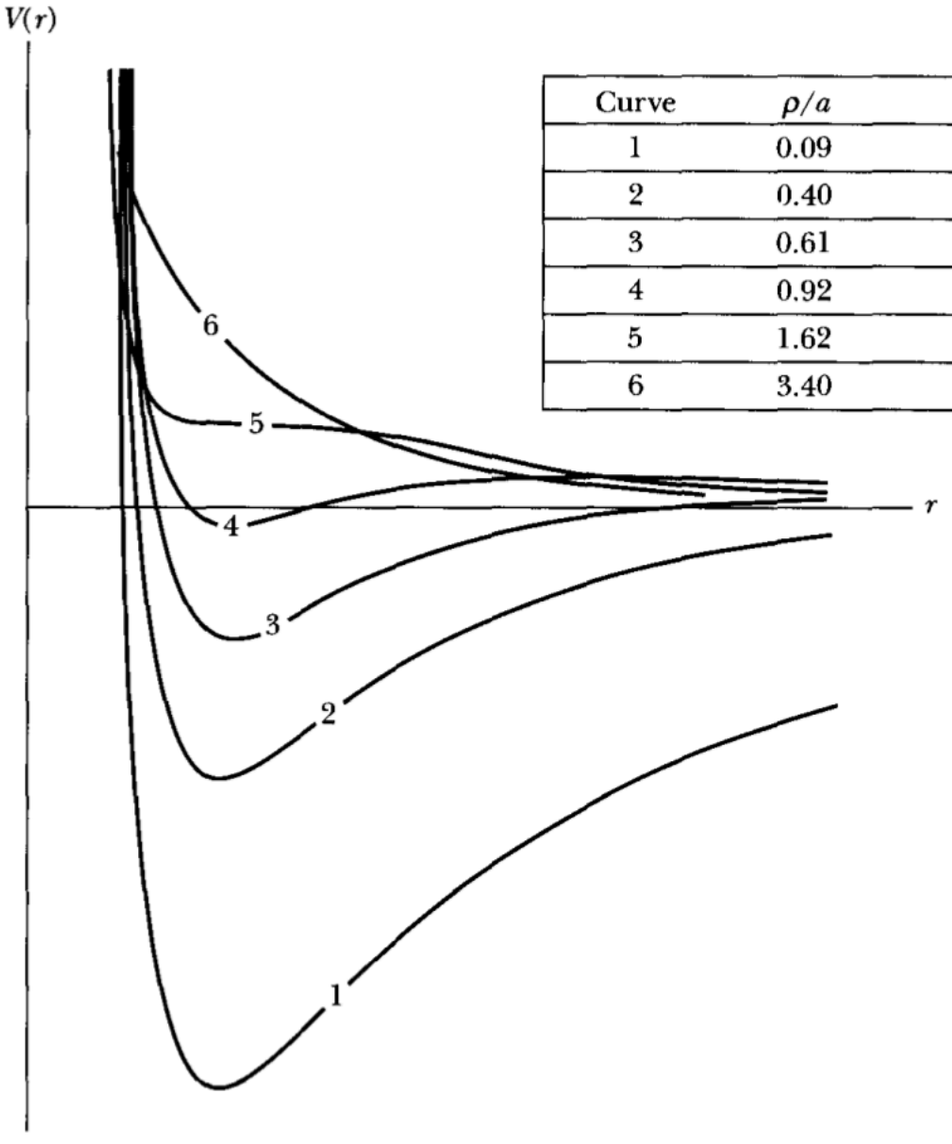
If, then, the angular momentum and energy allow a circular orbit at $r = \rho$, the motion is stable if

$$\frac{a}{\rho} \gtrsim 0.62$$

or

$$\rho \lesssim 1.62a \tag{8.97}$$

The stability condition for orbits in a screened potential is illustrated graphically in Figure 8-16, which shows the potential V(r) for various values of ρ/a . The force constant k is the same for all the curves, but $l^2/2\mu$ has been adjusted to maintain the minimum of the potential at the same value of the radius as a is changed.



Curve	ρ/a	$l^2/2\mu$
1	0.09	$1.93V_0$
2	0.40	$1.82V_0$
3	0.61	$1.75V_0$
4	0.92	$1.50V_0$
5	1.62	$\equiv V_0$
6	3.40	$0.59V_0$

FIGURE 8-16 Example 8.7. Potentials 1–4 produce a stable, circular orbit for values of $\rho/a \lesssim 1.62$.

For $\rho/a < 1.62$, a true minimum exists for the potential, indicating that the circular orbit is stable with respect to small oscillations. For $\rho/a > 1.62$, there is no minimum, so circular orbits cannot exist. For $\rho/a = 1.62$, the potential has zero slope at the position that a circular orbit would occupy. The orbit is unstable at this position, because (ω_0^2 is zero in Equation 8.90 and the displacement x increases linearly with time.

An interesting feature of this potential function is that under certain conditions there can exist bound orbits for which the total energy is positive (see, for example, curve 4 in Figure 8-16).

EXAMPLE 8.7

Determine whether a particle moving on the inside surface of a cone under the influence of gravity (see Example 7.4) can have a stable circular orbit.

Solution. In Example 7.4, we found that the angular momentum about the z-axis was a constant of the motion:

$$l = mr^2\dot{\theta} = \text{constant}$$

We also found the equation of motion for the coordinate r:

$$\ddot{r} - r\dot{\theta}^2 \sin^2\alpha + g \sin\alpha \cos\alpha = 0 \tag{8.98}$$

If the initial conditions are appropriately selected, the particle can move in a circular orbit about the vertical axis with the plane of the orbit at a constant height z_0 above the horizontal plane passing through the apex of the cone. Although this problem does not involve a central force, certain aspects of the motion are the same as for the central-force case. Thus we may discuss, for example, the stability of circular orbits for the particle. To do this, we perform a perturbation calculation.

First, we assume that a circular orbit exists for $r = \rho$. Then, we apply the perturbation $r \to \rho + x$. The quantity $r\dot{\theta}^2$ in Equation 8.98 can be expressed as

$$r\dot{\theta}^2 = r \cdot \frac{l^2}{m^2 r^4} = \frac{l^2}{m^2 r^3}$$

$$= \frac{l^2}{m^2}(\rho + x)^{-3} = \frac{l^2}{m^2\rho^3}\left(1 + \frac{x}{\rho}\right)^{-3}$$

$$\cong \frac{l^2}{m^2\rho^3}\left(1 - 3\frac{x}{\rho}\right)$$

where we have retained only the first term in the expansion, because x/ρ is by hypothesis a small quantity.

Then, because $\ddot{\rho} = 0$, Equation 8.98 becomes, approximately,

$$\ddot{x} - \frac{l^2 \sin^2 \alpha}{m^2\rho^3}\left(1 - 3\frac{x}{\rho}\right) + g \sin \alpha \cos \alpha = 0$$

or

$$\ddot{x} + \left(\frac{3l^2 \sin^2 \alpha}{m^2\rho^4}\right)x - \frac{l^2 \sin \alpha}{m^2\rho^3} + g \sin \alpha \cos \alpha = 0 \qquad (8.99)$$

If we evaluate Equation 8.98 at $r = \rho$, then $\ddot{r} = 0$, and we have

$$g \sin \alpha \cos \alpha = \rho\dot{\theta}^2 \sin^2 \alpha$$

$$= \frac{l^2}{m^2\rho^3} \sin^2 \alpha$$

In view of this result, the last two terms in Equation 8.99 cancel, and there remains

$$\ddot{x} + \left(\frac{3l^2 \sin^2 \alpha}{m^2\rho^4}\right)x = 0$$

The solution to this equation is just a harmonic oscillation with a frequency ω where

$$\omega = \frac{\sqrt{3}l}{m\rho^2} \sin \alpha$$

Thus, the circular orbit is stable.

CHAPTER 9 - Dynamics of a System of Particles

9.1 Introduction

Thus far, we have treated our dynamical problems primarily in terms of single particles. Even though we have considered extended objects such as projectiles and planets, we have been able to treat them as single particles. Generally, we have not had to deal with the internal interactions between the many particles that make up the extended body.

Later, when we treat the dynamics of rigid bodies, we must describe rotational as well as translational motion. We need to prepare the techniques that will allow us to do this.

We first extend our discussion to describe the system of n particles. These particles may form a loose aggregate - such as a pile of rocks or a volume of gas molecules - or form a rigid body in which the constituent particles are restrained from moving relative to one another. We devote the latter part of the chapter to a study of the interaction of two particles (n= 2). For the three-body problem (n = 3), the solutions become formidable. Perturbation techniques often are used, although great progress has been made through the use of numerical methods with high-speed computers. Finally, we shall examine rocket motion.

Newton's Third Law plays a prominent role in the dynamics of a system of particles because of the internal forces between the particles in the system. We need to make two assumptions concerning the internal forces:

The forces exerted by two particles α and β on each other are equal in magnitude and opposite in direction. Let $\mathbf{f}_{\alpha\beta}$ represent the force on the α^{th} particle due to the β^{th} particle. The so-called "weak" form of Newton's Third Law is

$$\mathbf{f}_{\alpha\beta} = -\mathbf{f}_{\beta\alpha} \tag{9.1}$$

The forces exerted by two particles α and β on each other, in addition to being equal and opposite, must lie on the straight line joining the two particles. This more restrictive form of Newton's Third Law, often called the "strong" form, is displayed in Figure 9-1.

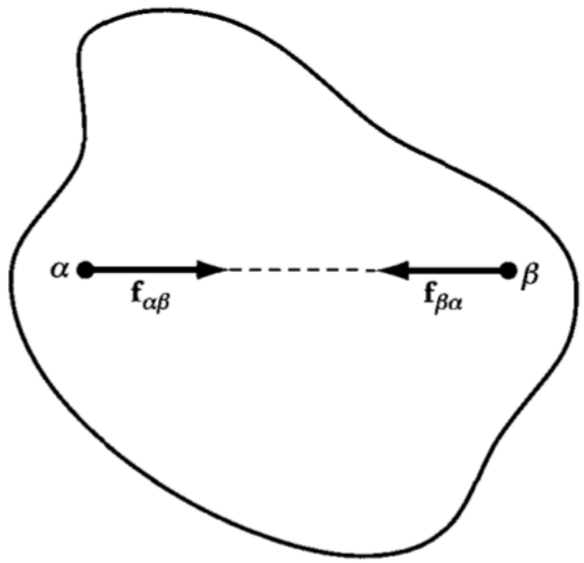


FIGURE 9-1 Example of the strong form of Newton's Third Law, where the equal and opposite forces between two particles must lie along a straight line joining the two particles. The force is attractive, as in the molecular attraction in a solid.

We must be careful to remember when each form of Newton's Third Law applies. We recall from Section 2.2 that the Third Law is not always valid for moving charged particles; electromagnetic forces are *velocity dependent*. For example, magnetic forces, those forces exerted on a moving charge q in a magnetic field **B** ($\mathbf{F} = q\mathbf{v} \times \mathbf{B}$), obey the weak form, but not the strong form, of the Third Law.

9.2 Center of Mass

Consider a system composed of n particles, with each particle's mass described by m_α, where α is an index from $\alpha = 1$ to $\alpha = n$. The total mass of the system is denoted by M,

$$M = \sum_\alpha m_\alpha \tag{9.2}$$

where the summation over α (as in all summations with Greek indices) runs from $\alpha = 1$ to $\alpha = n$.

Such a system is displayed in Figure 9-2.

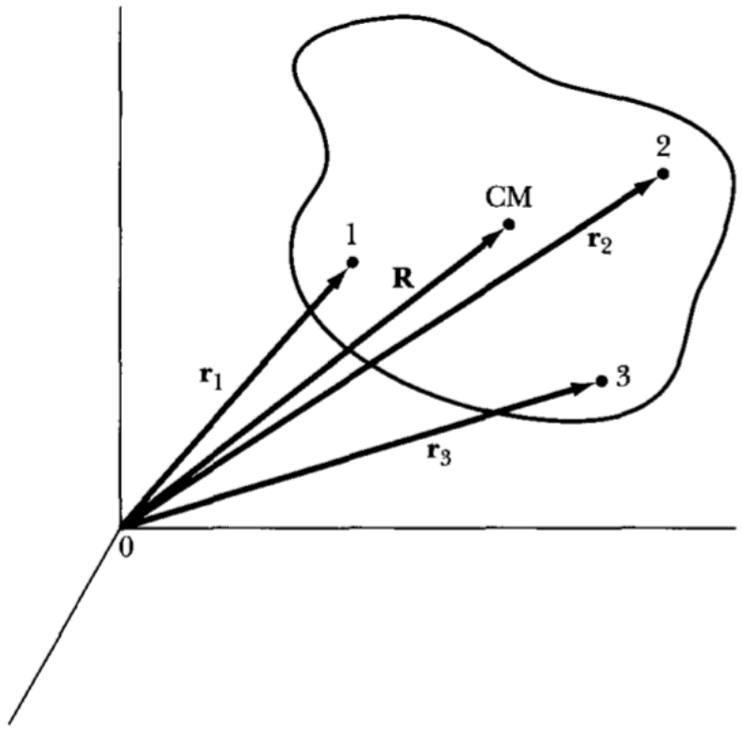


FIGURE 9-2 The position vectors to particles 1, 2, and 3 in the body are indicated, along with the center of mass position vector **R.**

If the vector connecting the origin with the α^{th} particle is $\mathbf{r}_\alpha$, then the vector defining the position of the system's center of mass is

$$\mathbf{R} = \frac{1}{M}\sum_{\alpha} m_\alpha \mathbf{r}_\alpha \tag{9.3}$$

For a continuous distribution of mass, the summation is replaced by an integral.

$$\mathbf{R} = \frac{1}{M}\int \mathbf{r}\, dm \tag{9.4}$$

The location of the center of mass of a body is uniquely defined, but the position vector **R** depends on the coordinate system chosen. If the origin in Figure 9-2
were chosen elsewhere, the vector **R** would be different.

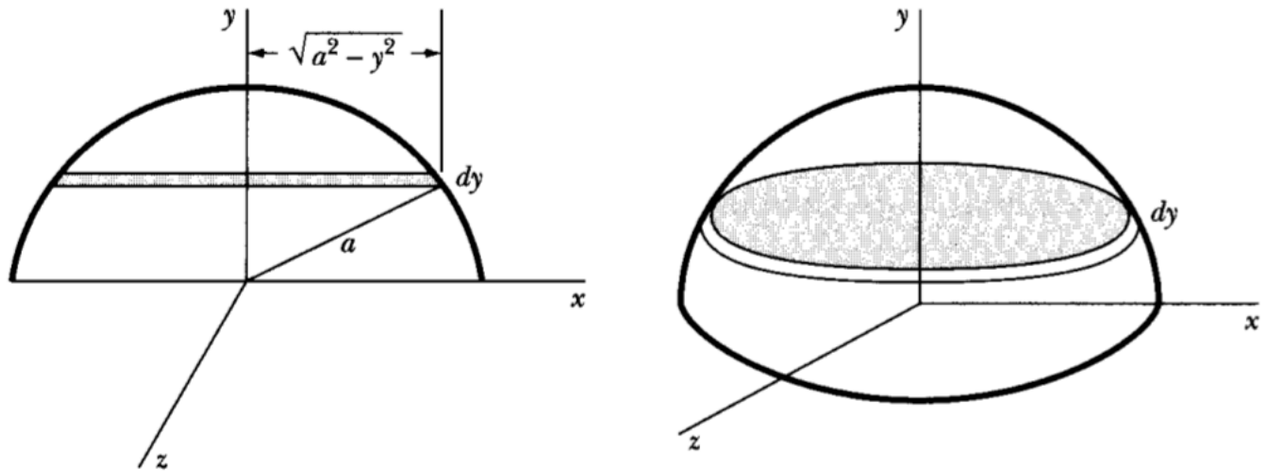
EXAMPLE 9.1

Find the center of mass of a solid hemisphere of constant density.

Solution. Let the density be ρ , the hemispherical mass be M, and the radius be a.

$$\rho = \frac{M}{\frac{2}{3}\pi a^3}$$

We want to choose the origin of our coordinate system carefully (Figure 9-3) to make the problem as simple as possible.



(a) (b)

FIGURE 9-3 Example 9.1. (a) We choose a thin slice *dy* of a solid hemisphere of constant density to find the center of mass position value *Y*. (b) The area of the slice *dy* is circular.

The position coordinates of R are (X, Y, Z). From symmetry, X = 0, Z = 0. This should be obvious from Equation 9.4,

$$X = \frac{1}{M}\int_{-a}^{a} x\, dm$$

$$Z = \frac{1}{M}\int_{-a}^{a} z\, dm$$

because we are integrating over an odd power of a variable with symmetric limits. For Y, however, the limits are asymmetric.

$$Y = \frac{1}{M}\int_0^a y\, dm$$

Construct dm so it is placed at a constant value of y. A circular slice perpendicular to the y-axis suffices (see Figure 9-3).

$$dm = \rho dV = \rho\pi(a^2 - y^2)\, dy$$

$$Y = \frac{1}{M}\int_0^a \rho\pi y(a^2 - y^2)\, dy$$

$$Y = \frac{\pi\rho a^4}{4M} = \frac{3a}{8}$$

The position of the center of mass is (0,3a/8, 0).

9.3 Linear Momentum of the System

If a certain group of particles constitutes a *system*, then the resultant force acting on a particle within the system (say, the α^{th} particle) is in general composed of two parts. One part is the resultant of all forces whose origin lies outside of the system; this is called the **external force**, $\mathbf{F}_\alpha^{(e)}$. The other part is the resultant of the forces arising from the interaction of all of the other n - 1 particles with the α^{th} particle; this is called the **internal force**, $\mathbf{f}_\alpha$. Force $\mathbf{f}_\alpha$ is given by the vector sum of all the individual forces $\mathbf{f}_{\alpha\beta}$,

$$\mathbf{f}_\alpha = \sum_\beta \mathbf{f}_{\alpha\beta} \tag{9.5}$$

where $\mathbf{f}_{\alpha\beta}$ represents the force on the α^{th} particle due to the β^{th} particle. The total force acting on the α^{th} particle is therefore

$$\mathbf{F}_\alpha = \mathbf{F}_\alpha^{(e)} + \mathbf{f}_\alpha \tag{9.6}$$

Also, according to the weak statement of Newton's Third Law, we have

$$\mathbf{f}_{\alpha\beta} = -\mathbf{f}_{\beta\alpha} \tag{9.7}$$

Newton's Second Law for the α^{th} particle can be written as

$$\dot{\mathbf{p}}_\alpha = m_\alpha \ddot{\mathbf{r}}_\alpha = \mathbf{F}_\alpha^{(e)} + \mathbf{f}_\alpha \tag{9.1}$$

or

$$\frac{d^2}{dt^2}(m_\alpha \mathbf{r}_\alpha) = \mathbf{F}_\alpha^{(e)} + \sum_\beta \mathbf{f}_{\alpha\beta} \tag{9.8}$$

Summing this expression over α , we have

$$\frac{d^2}{dt^2}\sum_\alpha m_\alpha \mathbf{r}_\alpha = \sum_\alpha \mathbf{F}_\alpha^{(e)} + \sum_{\substack{\alpha \\ \alpha\neq\beta}}\sum_\beta \mathbf{f}_{\alpha\beta} \tag{9.9}$$

where the terms $\alpha = \beta$ do not enter in the second sum on the right-hand side, because $\mathbf{f}_{\alpha\alpha} = 0$. The summation on the left-hand side just yields M**R** (see Equation 9.3), and the second time derivative is $M\ddot{\mathbf{R}}$. The first term on the right-hand side is the sum of all the external forces and can be written as

$$\sum_\alpha \mathbf{F}_\alpha^{(e)} \equiv \mathbf{F} \tag{9.10}$$

The second term on the right-hand side in Equation 9.9 can be expressed as

$$\sum_{\substack{\alpha \\ \alpha\neq\beta}}\sum_\beta \mathbf{f}_{\alpha\beta} \equiv \sum_{\alpha,\beta\neq\alpha} \mathbf{f}_{\alpha\beta} = \sum_{\alpha<\beta} (\mathbf{f}_{\alpha\beta} + \mathbf{f}_{\beta\alpha})$$

which vanishes according to Equation 9.1. Thus, we have the first important result

$$M\ddot{\mathbf{R}} = \mathbf{F} \tag{9.11}$$

which we can express as follows:

I. *The center of mass of a system moves as if it were a single particle of mass equal to the total mass of the system, acted on by the total external force, and independent of the nature of the internal forces (as long as they follow* $\mathbf{f}_{\alpha\beta} = -\mathbf{f}_{\beta\alpha}$*, the weak form of Newton's Third Law).*

The total linear momentum of the system is

$$\mathbf{P} = \sum_{\alpha} m_{\alpha} \dot{\mathbf{r}}_{\alpha} = \frac{d}{dt} \sum_{\alpha} m_{\alpha} \mathbf{r}_{\alpha} = \frac{d}{dt}(M\mathbf{R}) = M\dot{\mathbf{R}} \tag{9.12}$$

and

$$\dot{\mathbf{P}} = M\ddot{\mathbf{R}} = \mathbf{F} \tag{9.13}$$

Thus, the total linear momentum of the system is conserved if there is no external force. From Equations 9.12 and 9.13, we note our second and third important results:

II. *The linear momentum of the system is the same as if a single particle of mass* M *were located at the position of the center of mass and moving in the manner the center of mass moves.*

III. *The total linear momentum for a system free of external forces is constant and equal to the linear momentum of the center of mass* (the law of conservation of linear momentum for a system).

All measurements must be made in an inertial reference system. An example of the linear momentum of a system is given by the explosion of an artillery shell above ground. Because the explosion is an internal effect, the only external force affecting the center of mass velocity is due to gravity. The center of mass of the artillery shell fragments immediately after the explosion must continue with the velocity of the shell just before the explosion.

EXAMPLE 9.2

A chain of uniform linear mass density ρ , length b, and mass M (ρ = M/b) hangs as shown in Figure 9-4.

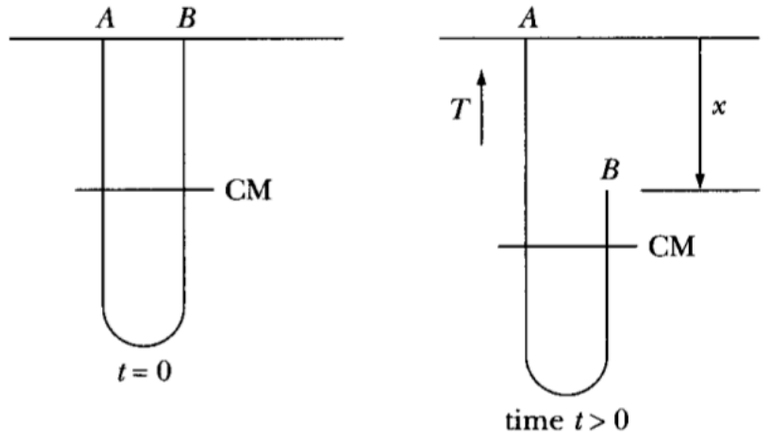


FIGURE 9-4 Example 9.2. (a) A chain of uniform linear mass density hangs at points *A* and *B* before *B* is released at time $t = 0$. (b) At time t the end *B* has fallen a distance x.

At time t = 0, the ends A and B are adjacent, but end B is released. Find the tension in the chain at point A after end B has fallen a distance x by (a) assuming free fall and (b) by using energy conservation.

Solution. (a) In the case of free fall, let's assume the only forces acting on the system at time t are the tension T acting vertically upward at point A and the gravitational force Mg pulling the chain down. The center of mass momentum reacts to these forces such that

$$\dot{P} = Mg - T \tag{9.14}$$

The right side of the chain, with mass ρ(b - x)/2, is moving at the speed $\dot{x}$, and the left side of the chain is not moving. The total momentum of the system is therefore

$$P = \rho\left(\frac{b - x}{2}\right)\dot{x}$$

and

$$\dot{P} = \frac{\rho}{2}[-\dot{x}^2 + \ddot{x}(b - x)] \tag{9.15}$$

For free fall, we have $x = gt^2/2$, so that

$$\dot{x} = gt = \sqrt{2gx}$$

$$\ddot{x} = g$$

and

$$\dot{P} = \frac{\rho}{2}(gb - 3gx) = Mg - T$$

and finally.

$$T = \frac{Mg}{2}\left(\frac{3x}{b} + 1\right) \tag{9.16}$$

(b) Calkin and March (Am. J. Phys. 57, 154 [1989]) have found that chains act much like a perfectly flexible, inextensible rope that conserves energy when it falls, with no dissipative mechanisms. We treat the chain as one-dimensional motion, ignoring the small horizontal motion. Let the potential energy U be measured relative to the fixed end of the chain, so that the initial potential energy $U(t = 0) = U_0 = -\rho g b^2/4$. A careful geometric construction shows that the potential energy after the chain has dropped a distance x is

$$U = -\frac{1}{4}\rho g(b^2 + 2bx - x^2)$$

The kinetic energy (where we use K instead of T to avoid confusion with tension) is determined from the speed x of the right side of the chain, so that

$$K = \frac{\rho}{4}(b - x)\dot{x}^2$$

Because energy is conserved, we must have $K + U = U_0$.

$$\frac{\rho}{4}(b - x)\dot{x}^2 - \frac{1}{4}\rho g(b^2 + 2bx - x^2) = -\frac{1}{4}\rho g b^2$$

We solve for $\dot{x}^2$ to obtain

$$\dot{x}^2 = \frac{g(2bx - x^2)}{b - x} \tag{9.17}$$

To find the tension from Equations 9.14 and 9.15, we need to determine $\ddot{x}$. We take the derivative of Equation 9.17 and find

$$\ddot{x} = g + \frac{g(2bx - x^2)}{2(b - x)^2}$$

We now insert $\dot{x}^2$ and $\ddot{x}$ from the two previous equations into Equation 9.15 to determine P and insert this value of Pinto Equation 9.14. After collection of terms and solving for T, we obtain

$$T = \frac{Mg}{4b}\frac{1}{(b - x)}(2b^2 + 2bx - 3x^2) \tag{9.18}$$

Note the difference between the two results, Equations 9.16 and 9.18, for the free fall and energy conserving methods. It should be rather easy by experimentation to determine which is correct, because the latter result has the tension rising dramatically (T—> ∞) at the end when x—> b. Experiments by Calkin and March confirm that the tension does increase rapidly at the end to a maximum of about 25 times the chain's weight, and the observations as a function of x agree well with the calculations. Real chains cannot have an infinite tension.

For the free fall case, the tension in the chain is discontinuous on either side of the bottom bend; the tension is $T_1 = \rho\dot{x}^2/2$ on the fixed side and $T_2 = 0$ on the free side. For the energy conserving case, the tension T_2 on the free side is not zero, and this tension helps gravity pull the chain down. The result is that the chain falls about 15% faster than calculated for the free fall case. For energy-conserving chains, the tension is continuous: $T_1 = T_2 = \rho\dot{x}^2/4$. We examine further properties of the falling chain in the problems.

9.4 Angular Momentum of the System

It is often more convenient to describe a system by a position vector with respect to the center of mass. The position vector $\mathbf{r}_\alpha$ in the inertial reference system (see Figure 9-5) becomes

$$\mathbf{r}_\alpha = \mathbf{R} + \mathbf{r}'_\alpha \tag{9.19}$$

where $\mathbf{r}'_\alpha$ is the position vector of the particle a with respect to the center of mass.

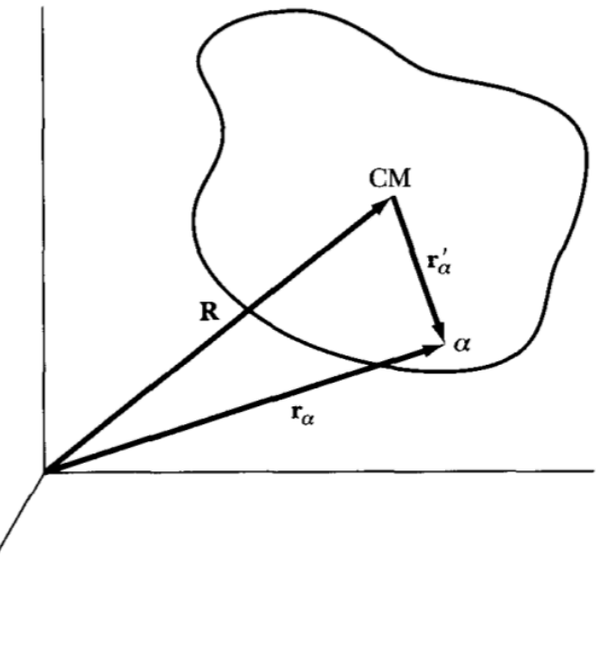


FIGURE 9-5 We can also describe a system by position vectors $\mathbf{r}'_\alpha$ with respect to the center of mass.

The angular momentum of the α^{th} particle about the origin is given by Equation 2.81:

$$\mathbf{L}_\alpha = \mathbf{r}_\alpha \times \mathbf{p}_\alpha \tag{9.20}$$

Summing this expression over α, and using Equation 9.19, we have

$$\begin{aligned}
\mathbf{L} = \sum_\alpha \mathbf{L}_\alpha &= \sum_\alpha (\mathbf{r}_\alpha \times \mathbf{p}_\alpha) = \sum_\alpha (\mathbf{r}_\alpha \times m_\alpha \dot{\mathbf{r}}_\alpha) \\
&= \sum_\alpha (\mathbf{r}'_\alpha + \mathbf{R}) \times m_\alpha (\dot{\mathbf{r}}'_\alpha + \dot{\mathbf{R}}) \\
&= \sum_\alpha m_\alpha \left[(\mathbf{r}'_\alpha \times \dot{\mathbf{r}}'_\alpha) + (\mathbf{r}'_\alpha \times \dot{\mathbf{R}}) + (\mathbf{R} \times \dot{\mathbf{r}}'_\alpha) + (\mathbf{R} \times \dot{\mathbf{R}}) \right]
\end{aligned} \tag{9.21}$$

The middle two terms can be written as

$$\left(\sum_\alpha m_\alpha \mathbf{r}'_\alpha\right) \times \dot{\mathbf{R}} + \mathbf{R} \times \frac{d}{dt}\left(\sum_\alpha m_\alpha \mathbf{r}'_\alpha\right)$$

which vanishes because

$$\begin{aligned} \sum_\alpha m_\alpha \mathbf{r}'_\alpha &= \sum_\alpha m_\alpha (\mathbf{r}_\alpha - \mathbf{R}) = \sum_\alpha m_\alpha \mathbf{r}_\alpha - \mathbf{R} \sum_\alpha m_\alpha \\ \sum_\alpha m_\alpha \mathbf{r}'_\alpha &= M\mathbf{R} - M\mathbf{R} \equiv 0 \end{aligned} \tag{9.22}$$

This indicates that $\sum_\alpha m_\alpha \mathbf{r}'_\alpha$ specifies the position of the center of mass in the center-of-mass coordinate system and is therefore a null vector. Thus, Equation 9.21 becomes

$$\mathbf{L} = M\mathbf{R} \times \dot{\mathbf{R}} + \sum_\alpha \mathbf{r}'_\alpha \times \mathbf{p}'_\alpha = \mathbf{R} \times \mathbf{P} + \sum_\alpha \mathbf{r}'_\alpha \times \mathbf{p}'_\alpha \tag{9.23}$$

Our fourth important result is

IV. *The total angular momentum about an origin is the sum of the angular momentum of the center of mass about that origin and the angular momentum of the system about the position of the center of mass.*

The time derivative of the angular momentum of the α^{th} particle is, from Equation 2.83,

$$\dot{\mathbf{L}}_\alpha = \mathbf{r}_\alpha \times \dot{\mathbf{p}}_\alpha \tag{9.24}$$

and, using Equations 9.7 and 9.8, we have

$$\dot{\mathbf{L}}_\alpha = \mathbf{r}_\alpha \times \left(\mathbf{F}_\alpha^{(e)} + \sum_\beta \mathbf{f}_{\alpha\beta}\right) \tag{9.25}$$

Summing this expression over a, we have

$$\dot{\mathbf{L}} = \sum_\alpha \dot{\mathbf{L}}_\alpha = \sum_\alpha (\mathbf{r}_\alpha \times \mathbf{F}_\alpha^{(e)}) + \sum_{\alpha, \beta \neq \alpha} (\mathbf{r}_\alpha \times \mathbf{f}_{\alpha\beta}) \tag{9.26}$$

The last term may be written as

$$\sum_{\alpha,\beta\neq\alpha}(\mathbf{r}_\alpha \times \mathbf{f}_{\alpha\beta}) = \sum_{\alpha<\beta}[(\mathbf{r}_\alpha \times \mathbf{f}_{\alpha\beta}) + (\mathbf{r}_\beta \times \mathbf{f}_{\beta\alpha})]$$

The vector connecting the α^{th} and β^{th} particles (see Figure 9-6) is defined to be

$$\mathbf{r}_{\alpha\beta} \equiv \mathbf{r}_\alpha - \mathbf{r}_\beta \tag{9.27}$$

and then, because $\mathbf{f}_{\alpha\beta} = -\mathbf{f}_{\beta\alpha}$. we have

$$\sum_{\alpha,\beta\neq\alpha}(\mathbf{r}_\alpha \times \mathbf{f}_{\alpha\beta}) = \sum_{\alpha<\beta}(\mathbf{r}_\alpha - \mathbf{r}_\beta) \times \mathbf{f}_{\alpha\beta}$$

$$= \sum_{\alpha<\beta}(\mathbf{r}_{\alpha\beta} \times \mathbf{f}_{\alpha\beta}) \tag{9.28}$$

Now we want to limit the discussion to central internal forces and apply the "strong" version of Newton's Third Law. Hence, $\mathbf{f}_{\alpha\beta}$ is along the same direction as $\pm\mathbf{r}_{\alpha\beta}$ and

$$\mathbf{r}_{\alpha\beta} \times \mathbf{f}_{\alpha\beta} \equiv 0 \tag{9.29}$$

and

$$\dot{\mathbf{L}} = \sum_\alpha[\mathbf{r}_\alpha \times \mathbf{F}_\alpha^{(e)}] \tag{9.30}$$

The right-hand side of this expression is just the sum of all the external torques:

$$\dot{\mathbf{L}} = \sum_\alpha \mathbf{N}_\alpha^{(e)} = \mathbf{N}^{(e)} \tag{9.31}$$

This leads to our next important result:

V. *If the net resultant external torques about a given axis vanish, then the total angular momentum of the system about that axis remains constant in time.*

Note also that the term

$$\sum_{\beta} \mathbf{r}_\alpha \times \mathbf{f}_{\alpha\beta} \tag{9.32}$$

is the torque on the α^{th} particle due to all the internal forces - that is, it is the *internal torque*. Because the sum of this quantity over all the particles α vanishes (see Equation 9.28),

$$\sum_{\alpha,\beta\neq\alpha} (\mathbf{r}_\alpha \times \mathbf{f}_{\alpha\beta}) = \sum_{\alpha<\beta} (\mathbf{r}_{\alpha\beta} \times \mathbf{f}_{\alpha\beta}) = 0 \tag{9.33}$$

the total internal torque must vanish, which we can state as

VI. *The total internal torque must vanish if the internal forces are central—that is, if* $\mathbf{f}_{\alpha\beta} = -\mathbf{f}_{\beta\alpha}$, *and the angular momentum of an isolated system cannot be altered without the application of external forces.*

EXAMPLE 9.3

A light string of length a has bobs of mass m_1 and m_2 ($m_2 > m_1$) on its ends. The end with m_1 is held and m_2 is whirled vigorously by hand above the head in a counterclockwise direction (looking down from above) and then released. Describe the subsequent motion, and find the tension in the string after release.

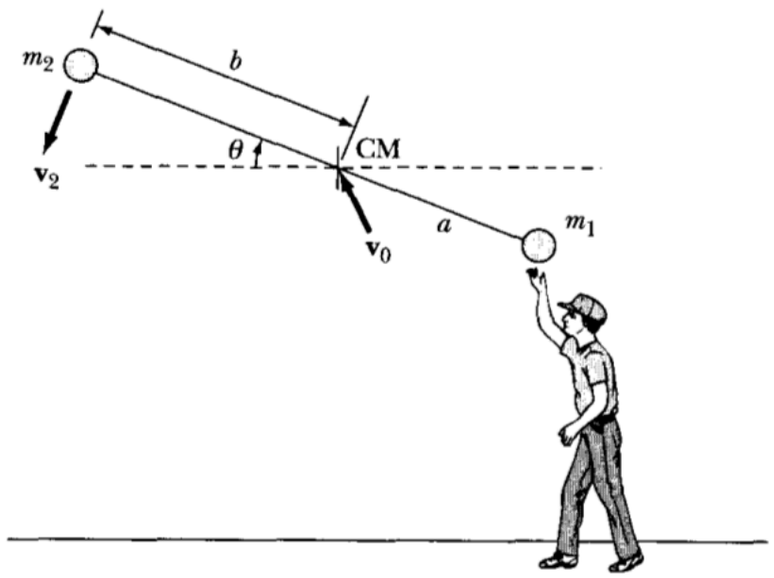


FIGURE 9-7 Example 9.3. A light string with masses m_1 and m_2 at its ends is whirled around by hand above the head and released.

Solution. The system is shown in Figure 9-7. The center of mass is a distance $b = [m_1/(m_1 + m_2)]a$ from mass m_2. After being released, the only forces on the system are the gravitational forces on m_1 and m_2. Assume that $\mathbf{v}_0$ is the initial velocity of the center of mass CM. The CM will continue in a parabolic path under the influence of gravity as if all the mass $(m_1 + m_2)$ were concentrated at the CM. But when released, mass m_2 is rotating around m_1 rapidly. Because no external torque exists, the system will continue to rotate. But now both m_1 and m_2 rotate about the CM, and the angular momentum is conserved. If mass m_2 is traveling with the linear velocity $\mathbf{v}_2$ when released, then we must have $v_2 = b\dot{\theta}$ [similarly, $v_1 = (a - b)\dot{\theta}$]. The tension in the string is, however, due to the centrifugal reaction of the masses rotating, which is, in this case,

$$\text{Centrifugal force} = \frac{m_2(b\dot{\theta})^2}{b} = \text{Tension}$$

$$\text{Tension} = m_2 b\dot{\theta}^2 = m_2\left(\frac{m_1 a}{m_1 + m_2}\right)\dot{\theta}^2 = \frac{m_1 m_2 a\dot{\theta}^2}{m_1 + m_2}$$

9.5 Energy of the System

The final conservation theorem, that of energy, may be derived for a system of particles as follows. Consider the work done on the system in moving it from a Configuration 1, in which all the coordinates $\mathbf{r}_\alpha$ are specified, to a Configuration 2, in which the coordinates $\mathbf{r}_\alpha$ have some different specification. (Note that the individual particles may just be rearranged in such a process, and that, for example, the position of the center of mass could remain stationary.) In analogy with Equation 2.84, we write

$$W_{12} = \sum_\alpha \int_1^2 \mathbf{F}_\alpha \cdot d\mathbf{r}_\alpha \tag{9.34}$$

where $\mathbf{F}_\alpha$ is the net resultant force acting on particle α . Using a procedure similar to that used to obtain Equation 2.86, we have

$$W_{12} = \sum_\alpha \int_1^2 d\left(\frac{1}{2} m_\alpha v_\alpha^2\right) = T_2 - T_1 \tag{9.35}$$

where

$$T = \sum_\alpha T_\alpha = \sum_\alpha \frac{1}{2} m_\alpha v_\alpha^2 \tag{9.36}$$

Using the relation (see Equation 9.19)

$$\dot{\mathbf{r}}_\alpha = \dot{\mathbf{r}}'_\alpha + \dot{\mathbf{R}} \tag{9.37}$$

we have

$$\begin{aligned} \dot{\mathbf{r}}_\alpha \cdot \dot{\mathbf{r}}_\alpha = v_\alpha^2 &= (\dot{\mathbf{r}}'_\alpha + \dot{\mathbf{R}}) \cdot (\dot{\mathbf{r}}'_\alpha + \dot{\mathbf{R}}) \\ &= (\dot{\mathbf{r}}'_\alpha \cdot \dot{\mathbf{r}}'_\alpha) + 2(\dot{\mathbf{r}}'_\alpha \cdot \dot{\mathbf{R}}) + (\dot{\mathbf{R}} \cdot \dot{\mathbf{R}}) \\ &= v_\alpha'^2 + 2(\dot{\mathbf{r}}'_\alpha \cdot \dot{\mathbf{R}}) + V^2 \end{aligned}$$

where $\mathbf{v}' \equiv \dot{\mathbf{r}}'$ and where V is the velocity of the center of mass. Then

$$\begin{aligned} T &= \sum_\alpha \frac{1}{2} m_\alpha v_\alpha^2 \\ &= \sum_\alpha \frac{1}{2} m_\alpha v_\alpha'^2 + \sum_\alpha \frac{1}{2} m_\alpha V^2 + \dot{\mathbf{R}} \cdot \frac{d}{dt} \sum_\alpha m_\alpha \mathbf{r}'_\alpha \end{aligned} \tag{9.38}$$

But, by a previous argument $\sum_\alpha m_\alpha \mathbf{r}'_\alpha = 0$, and the last term vanishes. Thus,

$$\boxed{T = \sum \frac{1}{2} m_\alpha v_\alpha'^2 + \frac{1}{2} M V^2} \tag{9.39}$$

which can be stated:

VII. ***The total kinetic energy of the system is equal to the sum of the kinetic energy of a particle of mass M moving with the velocity of the center of mass and the kinetic energy of motion of the individual particles relative to the center of mass.***

Next, the total force in Equation 9.34 can be separated as in Equation 9.6:

$$W_{12} = \sum_{\alpha} \int_1^2 \mathbf{F}_{\alpha}^{(e)} \cdot d\mathbf{r}_{\alpha} + \sum_{\alpha, \beta \neq \alpha} \int_1^2 \mathbf{f}_{\alpha\beta} \cdot d\mathbf{r}_{\alpha} \tag{9.40}$$

If the forces $\mathbf{F}_{\alpha}^{(e)}$ and $\mathbf{f}_{\alpha\beta}$ are conservative, then they are derivable from potential functions, and we can write

$$\left.\begin{aligned} \mathbf{F}_{\alpha}^{(e)} &= -\nabla_{\alpha} U_{\alpha} \\ \mathbf{f}_{\alpha\beta} &= -\nabla_{\alpha} \bar{U}_{\alpha\beta} \end{aligned}\right\} \tag{9.41}$$

where U_{α} and $\bar{U}_{\alpha\beta}$ are the potential functions but which do not necessarily have the same form. The notation ∇_{α} means that the gradient operation is performed with respect to the coordinates of the α^{th} particle

The first term in Equation 9.40 becomes

$$\begin{aligned} \sum_{\alpha} \int_1^2 \mathbf{F}_{\alpha}^{(e)} \cdot d\mathbf{r}_{\alpha} &= -\sum_{\alpha} \int_1^2 (\nabla_{\alpha} U_{\alpha}) \cdot d\mathbf{r}_{\alpha} \\ &= -\sum_{\alpha} U_{\alpha} \Big|_1^2 \end{aligned} \tag{9.42}$$

The second term* in Equation 9.40 is

$$\begin{aligned} \sum_{\alpha, \beta \neq \alpha} \int_1^2 \mathbf{f}_{\alpha\beta} \cdot d\mathbf{r}_{\alpha} &= \sum_{\alpha < \beta} \int_1^2 (\mathbf{f}_{\alpha\beta} \cdot d\mathbf{r}_{\alpha} + \mathbf{f}_{\beta\alpha} \cdot d\mathbf{r}_{\beta}) \\ &= \sum_{\alpha < \beta} \int_1^2 \mathbf{f}_{\alpha\beta} \cdot (d\mathbf{r}_{\alpha} - d\mathbf{r}_{\beta}) = \sum_{\alpha < \beta} \int_1^2 \mathbf{f}_{\alpha\beta} \cdot d\mathbf{r}_{\alpha\beta} \end{aligned} \tag{9.43}$$

where, following the definition in Equation 9.27, $d\mathbf{r}_{\alpha\beta} = d\mathbf{r}_{\alpha} - \mathbf{r}_{\beta}$.

Because $\bar{U}_{\alpha\beta}$ is a function only of the distance between m_{α} and m_{β} , it therefore depends on six quantities - that is, the three coordinates of m_{α} (the $x_{\alpha i}$) and the three coordinates of m_{β}(the $x_{\beta i}$).

The total derivative of $\bar{U}_{\alpha\beta}$ is therefore the sum of six partial derivatives and is given by

$$d\bar{U}_{\alpha\beta} = \sum_i \left(\frac{\partial \bar{U}_{\alpha\beta}}{\partial x_{\alpha,i}} dx_{\alpha,i} + \frac{\partial \bar{U}_{\alpha\beta}}{\partial x_{\beta,i}} dx_{\beta,i} \right) \tag{9.44}$$

where the $x_{\beta i}$ are held constant in the first term and the $x_{\alpha i}$ are held constant in the second. Thus,

$$d\bar{U}_{\alpha\beta} = (\nabla_\alpha \bar{U}_{\alpha\beta}) \cdot d\mathbf{r}_\alpha + (\nabla_\beta \bar{U}_{\alpha\beta}) \cdot d\mathbf{r}_\beta \tag{9.45}$$

Now

$$\nabla_\alpha \bar{U}_{\alpha\beta} = -\mathbf{f}_{\alpha\beta} \tag{9.46}$$

but $\bar{U}_{\alpha\beta} = \bar{U}_{\beta\alpha}$, so

$$\nabla_\beta \bar{U}_{\alpha\beta} = \nabla_\beta \bar{U}_{\beta\alpha} = -\mathbf{f}_{\beta\alpha} = \mathbf{f}_{\alpha\beta} \tag{9.47}$$

Therefore,

$$\begin{aligned} d\bar{U}_{\alpha\beta} &= -\mathbf{f}_{\alpha\beta} \cdot (d\mathbf{r}_\alpha - d\mathbf{r}_\beta) \\ &= -\mathbf{f}_{\alpha\beta} \cdot d\mathbf{r}_{\alpha\beta} \end{aligned} \tag{9.48}$$

Using this result in Equation 9.43, we have

$$\sum_{\alpha,\beta \neq \alpha} \int_1^2 \mathbf{f}_{\alpha\beta} \cdot d\mathbf{r}_\alpha = -\sum_{\alpha<\beta} \int_1^2 d\bar{U}_{\alpha\beta} = -\sum_{\alpha<\beta} \bar{U}_{\alpha\beta} \Big|_1^2 \tag{9.49}$$

Combining Equations 9.42 and 9.49 to evaluate W12 in Equation 9.40, we find

$$W_{12} = -\sum_\alpha U_\alpha \Big|_1^2 - \sum_{\alpha<\beta} \bar{U}_{\alpha\beta} \Big|_1^2 \tag{9.50}$$

We obtained this equation assuming that both the external and internal forces were derivable from potentials. In such a case, the *total potential energy* (both internal and external) for the system can be written as

$$U = \sum_{\alpha} U_{\alpha} + \sum_{\alpha<\beta} \overline{U}_{\alpha\beta} \tag{9.51}$$

Then,

$$W_{12} = -U\Big|_1^2 = U_1 - U_2 \tag{9.52}$$

Combining this result with Equation 9.35, we have

$$T_2 - T_1 = U_1 - U_2$$

or

$$T_1 + U_1 = T_2 + U_2$$

so that

$$\boxed{E_1 = E_2} \tag{9.53}$$

which expresses the conservation of energy for the system. This result is valid for a system in which all the forces are derivable from potentials that do not depend explicitly on the time; we say that such a system is *conservative*.

VIII. ***The total energy for a conservative system is constant.***

In Equation 9.51, the term

$$\sum_{\alpha<\beta} \overline{U}_{\alpha\beta}$$

represents the internal potential energy of the system. If the system is a rigid body with the constituent particles restrained to maintain their relative positions, then, in any process involving the body, the internal potential energy remains constant. In such a case, the internal potential energy can be ignored when computing the total potential energy of the system. This amounts simply to redefining the position of zero potential energy, but this position is arbitrarily chosen anyway; that is, it is only the difference in potential energy that is physically significant. The absolute value of the potential energy is an arbitrary quantity.

EXAMPLE 9.4

A projectile of mass M explodes while in flight into three fragments (Figure 9-8).

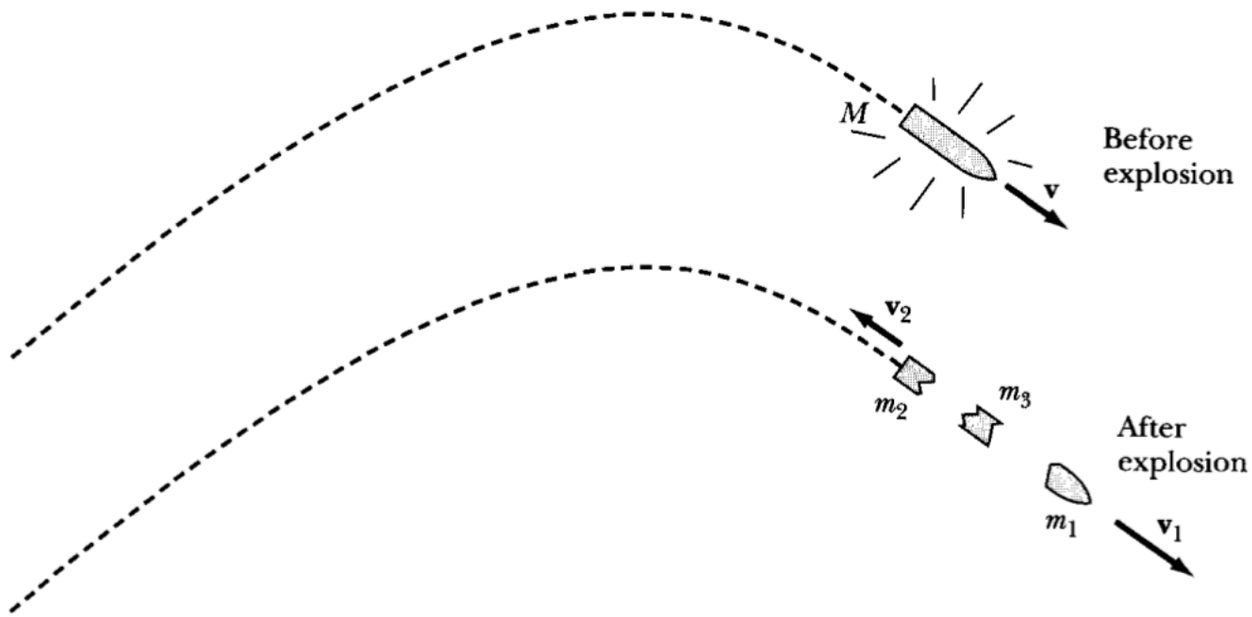


FIGURE 9-8 Example 9.4. A projectile of mass M explodes in flight into three fragments of masses m_1, m_2, and m_3.

One mass ($m_1 = M/2$) travels in the original direction of the projectile, mass m_2 ($= M/6$) travels in the opposite direction, and mass m_3 ($= M/3$) comes to rest. The energy E released in the explosion is equal to five times the projectile's kinetic energy at explosion. What are the velocities?

Solution. Let the velocity of the projectile of mass M be **v**. The three fragments have the following masses and velocities:

$$m_1 = \frac{M}{2}, \quad \mathbf{v}_1 = k_1\mathbf{v} \qquad \text{Forward direction, } k_1 > 0$$

$$m_2 = \frac{M}{6}, \quad \mathbf{v}_2 = -k_2\mathbf{v} \qquad \text{Opposite direction, } k_2 > 0$$

$$m_3 = \frac{M}{3}, \quad \mathbf{v}_3 = 0 \qquad \text{At rest}$$

The conservation of linear momentum and energy give

$$Mv = \frac{M}{2}k_1 v - \frac{M}{6}k_2 v \tag{9.54}$$

$$E + \frac{1}{2}Mv^2 = \frac{1}{2}\frac{M}{2}(k_1 v)^2 + \frac{1}{2}\frac{M}{6}(k_2 v)^2 \tag{9.55}$$

From Equation 9.54, $k_2 = 3k_1 - 6$, which we can insert into Equation 9.55:

$$5\left(\frac{1}{2}Mv^2\right) + \frac{1}{2}Mv^2 = \frac{Mv^2}{4}k_1^2 + \frac{Mv^2}{12}(3k_1 - 6)^2$$

which reduces to $k_1^2 - 3k_1 = 0$, giving the results $k_1 = 0$ and $k_1 = 3$. For $k_1 = 0$, the value of $k_2 = -6$, which is inconsistent with $k_2 > 0$. For $k_1 = 3$, the value of $k_2 = 3$. The velocities become

$$\mathbf{v}_1 = 3\mathbf{v}$$

$$\mathbf{v}_2 = -3\mathbf{v}$$

$$\mathbf{v}_3 = 0$$

EXAMPLE 9.5

A rope of uniform linear density ρ and mass m is wrapped one complete turn around a hollow cylinder of mass M and radius R. The cylinder rotates freely about its axis as the rope unwraps (Figure 9-9). The rope ends are at x = 0 (one fixed, one loose) when point Pis at $\theta = 0$, and the system is slightly displaced from equilibrium at rest. Find the angular velocity as a function of angular displacement θ of the cylinder.

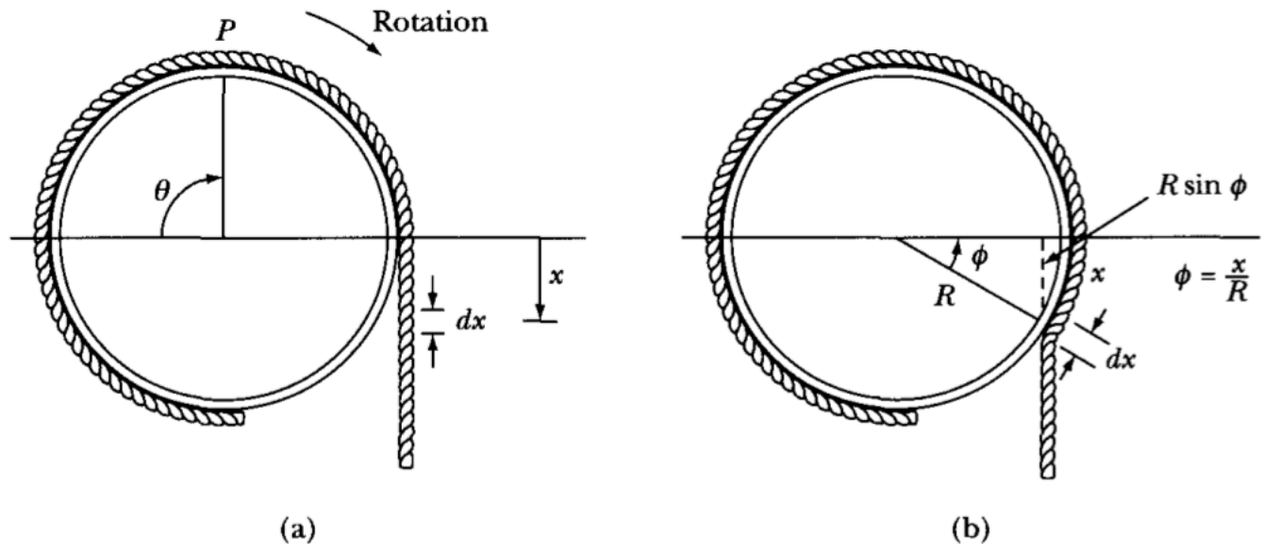


FIGURE 9-9 Example 9.5. (a) A rope is wound around a cylinder. Both ends are at $x = 0$ when $\theta = 0$. (b) Work is done to place section dx back up next to the cylinder.

Solution. Gravity has done work on the system to unwind the rope. Consider a section dx of the rope located a distance x from where it unwinds. The mass of this section is ρdx. If we were to perform work by reaching up and wrapping this loose end of the rope against the cylinder, how far up would the section dx actually travel? The distance x would be on the circumference of the cylinder (see Figure 9-9), and dx would be R sin(x/R) below x= 0. The total distance the section dx would move up is

$$\text{Distance } dx \text{ moves} = x - R\sin\left(\frac{x}{R}\right)$$

$$\text{Work done} = (\rho dx)g\left[x - R\sin\left(\frac{x}{R}\right)\right] \tag{9.56}$$

The total work done by gravity in unwrapping the rope through an angle θ is, therefore,

$$W = \int_0^{R\theta} \rho g\left[x - R\sin\left(\frac{x}{R}\right)\right] dx$$

$$W = \rho g R^2\left(\frac{\theta^2}{2} + \cos\theta - 1\right) \tag{9.57}$$

The work done by gravity must equal the kinetic energy gained by the rope and the cylinder.

$$T = \frac{1}{2}m(R\dot{\theta})^2 + \frac{1}{2}M(R\dot{\theta})^2 \tag{9.58}$$

Because W= T and $\rho = m/(2\pi R)$,

$$\frac{mgR}{2\pi}\left[\frac{\theta^2}{2} + \cos\theta - 1\right] = \frac{1}{2}(m + M)R^2\dot{\theta}^2$$

and

$$\dot{\theta}^2 = \frac{mg(\theta^2 + 2\cos\theta - 2)}{2\pi R(m + M)} \tag{9.59}$$

9.6 Elastic Collisions of Two Particles

For the next few sections, we apply the conservation laws to the interaction of two particles. When two particles interact, the motion of one particle relative to the other is governed by the force law that describes the interaction. This interaction may result from actual contact, as in the collision of two billiard balls, or the interaction may take place through the intermediary of a force field. For example, a *free* object (i.e., one not bound in a solar orbit) may "scatter" from the sun by a gravitational interaction, or an α-particle may be scattered by the electric field of an atomic nucleus. We demonstrated in the previous chapter that once the force law is known, the two-body problem can be completely solved. But even if the force of interaction between two particles is not known, a great deal can still be learned about the relative motion by using only the results of the conservation of momentum and energy. Thus, if the initial state of the system is known (i.e., if the velocity vector of each of the particles is specified), the conservation laws allow us to obtain information regarding the velocity vectors in the final state.

The "initial state" of the system is the condition of the particles when they are not yet sufficiently close to interact appreciably; the "final state' is the condition after the interaction has taken place. For a contact interaction, these conditions are obvious. But if the interaction takes place by a force field, then the rate of decrease of the force with distance must be taken into account in specifying the initial and final states.

On the basis of the conservation theorems alone, it is not possible to predict, for example, the angle between the initial and final velocity vectors of one of the particles; knowledge of the force law is required for such details. In this section and the next, we derive those relationships that require only the conservation of momentum and energy. Then, we examine the features of the collision process, which demand that the force law be specified. We limit our discussion primarily to elastic collisions, because the essential features of two-particle kinematics are adequately demonstrated by elastic collisions. The results obtained under the assumption only of momentum and energy conservation are valid (in the non-relativistic velocity region) even for quantum mechanical systems, because these conservation theorems are applicable to quantum as well as to classical systems.

We demonstrated on several occasions that the description of many physical processes is considerably simplified if one chooses coordinate systems at rest with respect to the system's center of mass. In the problem we now discuss - the elastic collision of two particles - the usual situation (and the one to which we confine our attention) is one in which the collision is between a moving particle and a particle at rest. A collision is elastic if no change in the internal energy of the particles results; thus, the conservation of energy may be applied without regard to the internal energy. Notice that heat may be generated when two mechanical bodies collide inelastically. Heat is just a manifestation of the agitation of a body's constituent particles and may therefore be considered a part of the internal energy. Although it is indeed simpler to describe the effects of the collision in a coordinate system in which the *center of mass is at rest*, the actual measurements are made in the **laboratory coordinate system** in which the observer is at rest.

In this system, one of the particles is normally moving, and the struck particle is normally at rest. We here refer to these two coordinate systems simply as the **CM** and the **LAB** systems.

We use the following notation:

$$\left.\begin{matrix} m_1 = \\ m_2 = \end{matrix}\right\} \text{Mass of the } \left\{\begin{matrix}\text{moving}\\ \text{struck}\end{matrix}\right\} \text{particle}$$

In general, primed quantities refer to the CM system:

$$\left.\begin{matrix} \mathbf{u}_1 = \text{Initial} \\ \mathbf{v}_1 = \text{Final} \end{matrix}\right\} \text{velocity of } m_1 \text{ in the LAB system}$$

$$\left.\begin{matrix} \mathbf{u}_1' = \text{Initial} \\ \mathbf{v}_1' = \text{Final} \end{matrix}\right\} \text{velocity of } m_1 \text{ in the CM system}$$

and similarly for $\mathbf{u}_2$, $\mathbf{u}_2$, $\mathbf{u}_2'$, and $\mathbf{v}_2'$ (but $\mathbf{u}_2 = 0$):

$$\left.\begin{matrix} T_0 = \\ T_0' = \end{matrix}\right\} \text{Total initial kinetic energy in } \left\{\begin{matrix}\text{LAB}\\ \text{CM}\end{matrix}\right\} \text{system}$$

$$\left.\begin{matrix} T_1 = \\ T_1' = \end{matrix}\right\} \text{Final kinetic energy of } m_1 \text{ in } \left\{\begin{matrix}\text{LAB}\\ \text{CM}\end{matrix}\right\} \text{system}$$

and similarly for T_2 and T_2',

$\mathbf{V}$ = velocity of the center of mass in the LAB system
ψ = angle through which m_1 is deflected in the LAB system
ζ = angle through which m_2 is deflected in the LAB system
θ = angle through which m_1 and m_2 are deflected in the CM system

Figure 9-10 illustrates the geometry of an elastic collision (we assume throughout that the scattering is axially symmetric so that no azimuthal angle need be introduced. However, axial symmetry is not always found in scattering problems; this is particularly true in certain quantum mechanical systems. in both the LAB and CM systems. The final state in the LAB and CM systems for the scattered particle m_1 may be conveniently summarized by the diagrams in Figure 9-11.

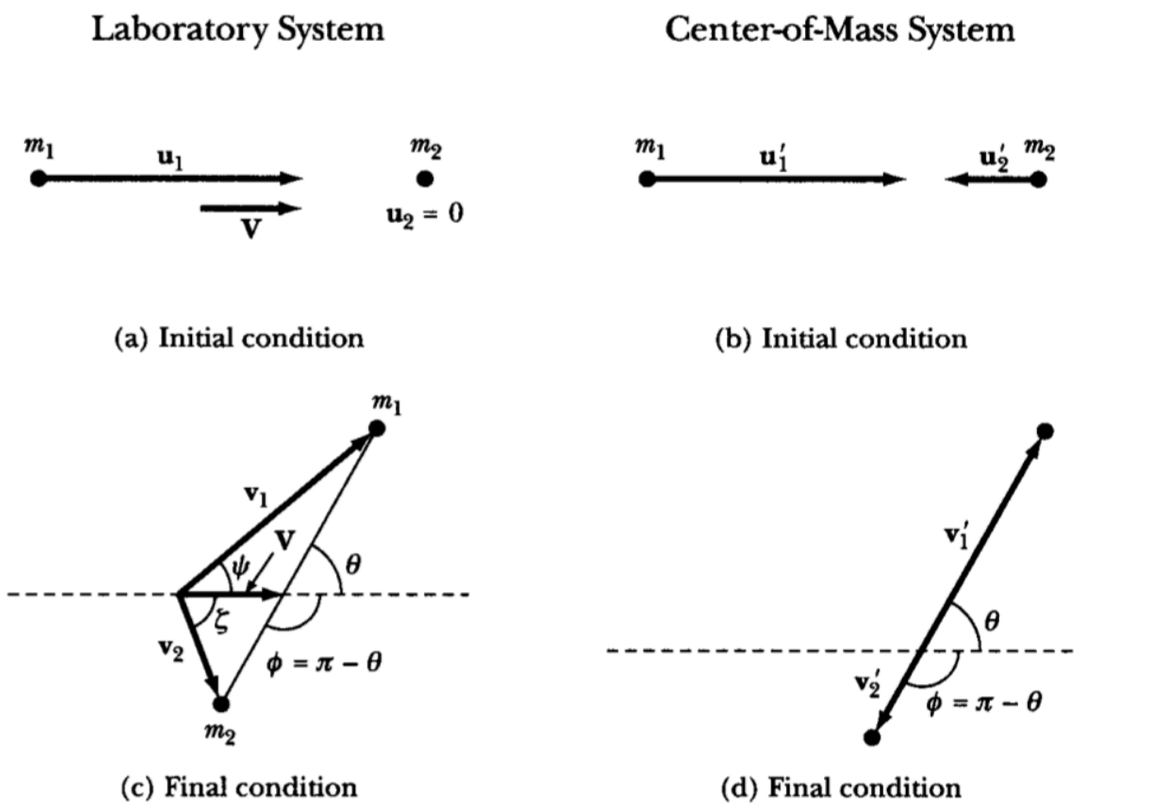


FIGURE 9-10 Geometry and notations of an elastic collision in the LAB and CM system: (a) Initial condition with $\mathbf{u}_2 = 0$ in the LAB system, (b) initial condition in the CM system, (c) final condition in the LAB system, and (d) final condition in the CM system. Note carefully the scattering angles.

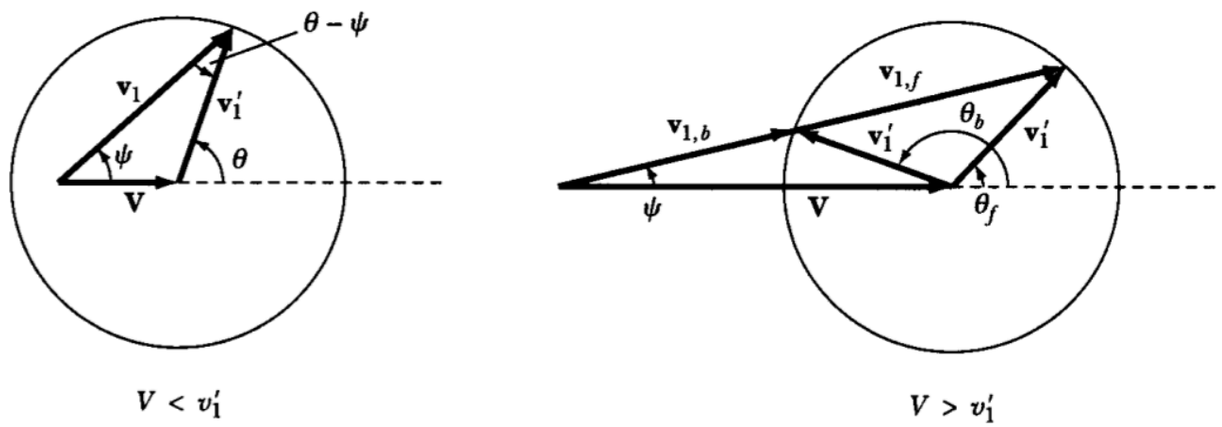


FIGURE 9-11 The final state of mass m_1 for the elastic collision of two particles for the case (a) $V < v_1'$ for which there is one trajectory and (b) $V > v_1'$ for which there are two possible trajectories (b stands for backward and f for forward).

We can interpret these diagrams in the following manner. To the velocity $\mathbf{V}$ of the CM, we can add the final CM velocity $\mathbf{v}_1'$ of the scattered particle. Depending on the angle θ at which the scattering takes place, the possible vectors $\mathbf{v}_1'$ lie on the circle of radius v_1' whose center is at the terminus of the vector $\mathbf{V}$. The LAB velocity $\mathbf{v}_1$ and LAB scattering angle ψ are then obtained by connecting the point of origin of $\mathbf{V}$ with the terminus of $\mathbf{v}_1'$.

If $V < v'_1$, only one possible relationship exists between $\mathbf{V}$, $\mathbf{v}_1$, $\mathbf{v}'_1$ and θ (see Figure 9-1 la). But if $V > v'_1$, then for every set $\mathbf{V}$, $\mathbf{v}'_1$ there exists two possible scattering angles and laboratory velocities: $\mathbf{v}_{1b}$, θ , and $\mathbf{v}_{1f}$, θ (see Figure 9-1 lb), where the designations b and f stand for *backward* and *forward*. This situation results from the fact that if the final CM velocity $\mathbf{v}'_1$ is insufficient to overcome the velocity $\mathbf{V}$ of the center of mass, then, even if m_1 is scattered into the backward direction in the CM system ($\theta > \pi/2$), the particle will appear at a forward angle in the LAB system ($\psi < \pi/2$). Thus, for $V > v'_1$, the velocity $\mathbf{v}_1$ in the LAB system is a double-valued function of $\mathbf{v}'_1$. In an experiment, we usually measure ψ, not the velocity *vector* $\mathbf{v}_1$, so that a single value of ψ can correspond to two different values of θ. Note, however, that a specification of the vectors $\mathbf{V}$ and $\mathbf{v}'_1$ always leads to a unique combination $\mathbf{v}_1$, θ ; but a specification of $\mathbf{V}$ and only the *direction* of $\mathbf{v}_1$ (i.e., ψ) allows the possibility of two final vectors, $\mathbf{v}_{1b}$ and $\mathbf{v}_{1f}$, if $V > v'_1$.

Having given a qualitative description of the scattering process, we now obtain some of the equations relating the various quantities.

According to the definition of the center of mass (Equation 9.3), we have

$$m_1\mathbf{r}_1 + m_2\mathbf{r}_2 = M\mathbf{R} \tag{9.60}$$

Differentiating with respect to the time, we find

$$m_1\mathbf{u}_1 + m_2\mathbf{u}_2 = M\mathbf{V} \tag{9.61}$$

But $\mathbf{u}_2 = 0$ and $M = m_1 + m_2$, the center of mass must therefore be moving (in the LAB system) toward m_2 with a velocity

$$\mathbf{V} = \frac{m_1\mathbf{u}_1}{m_1 + m_2} \tag{9.62}$$

By the same reasoning, because m_2 is initially at rest, the initial CM speed of m_2 must just equal V:

$$u_2' = V = \frac{m_1 u_1}{m_1 + m_2} \tag{9.63}$$

Note, however, that the motion and the velocities are opposite in direction and that vectorially $\mathbf{u}'_2 = -\mathbf{V}$.

The great advantage of using the CM coordinate system is because the total linear momentum in such a system is zero, so that before the collision the particles move directly toward each other and after the collision they move in exactly opposite directions. If the collision is elastic, as we have specified, then the masses do not change, and the conservation of linear momentum and kinetic energy is sufficient to provide that the CM speeds before and after collision are equal:

$$u_1' = v_1', \quad u_2' = v_2' \tag{9.64}$$

Term u_1 is the *relative* speed of the two particles in either the CM or the LAB system, $u_1 = u'_1 + u'_2$. We therefore have, for the final CM speeds.

$$v_2' = \frac{m_1 u_1}{m_1 + m_2} \tag{9.65a}$$

$$v_1' = u_1 - u_2' = \frac{m_2 u_1}{m_1 + m_2} \tag{9.65b}$$

We have (see Figure 9-1 la)

$$v_1' \sin\theta = v_1 \sin\psi \tag{9.66a}$$

and

$$v_1' \cos\theta + V = v_1 \cos\psi \tag{9.66b}$$

Dividing Equation 9.66a by Equation 9.66b,

$$\tan\psi = \frac{v_1' \sin\theta}{v_1' \cos\theta + V} = \frac{\sin\theta}{\cos\theta + (V/v_1')} \tag{9.67}$$

According to Equations 9.62 and 9.65b, V/v'$_1$ is given by

$$\frac{V}{v_1'} = \frac{m_1 u_1/(m_1 + m_2)}{m_2 u_1/(m_1 + m_2)} = \frac{m_1}{m_2} \tag{9.68}$$

Thus, the ratio m_1/m_2 governs whether Figure 9-11a or Figure 9-11b describes the scattering process:

Figure 9-11a: $V < v_1', \quad m_1 < m_2$

Figure 9-11b: $V > v_1', \quad m_1 > m_2$

If we combine Equations 9.67 and 9.68 and write

$$\boxed{\tan\psi = \frac{\sin\theta}{\cos\theta + (m_1/m_2)}} \tag{9.69}$$

we see that if $m_1 \ll m_2$, the LAB and CM scattering angles are approximately equal; that is, the particle m_2 is but little affected by the collision with m_1 and acts essentially as a fixed scattering center

Thus

$$\boxed{\psi \cong \theta, \qquad m_1 \ll m_2} \tag{9.70}$$

However, if $m_1 = m_2$, then

$$\tan\psi = \frac{\sin\theta}{\cos\theta + 1} = \tan\frac{\theta}{2}$$

so that

$$\boxed{\psi = \frac{\theta}{2}, \qquad m_1 = m_2} \tag{9.71}$$

and the LAB scattering angle is one half the CM scattering angle. Because the maximum value of θ is 180°, Equation 9.71 indicates that for $m_1 = m_2$, there can be no scattering in the LAB system at angles greater than 90°

Let us now refer to Figure 9-10c and construct a diagram for the recoil particle m_2 similar to Figure 9-1la. The situation is illustrated in Figure 9-12, from which we find

$$v_2 \sin \zeta = v_2' \sin \theta \tag{9.72a}$$

$$v_2 \cos \zeta = V - v_2' \cos \theta \tag{9.72b}$$

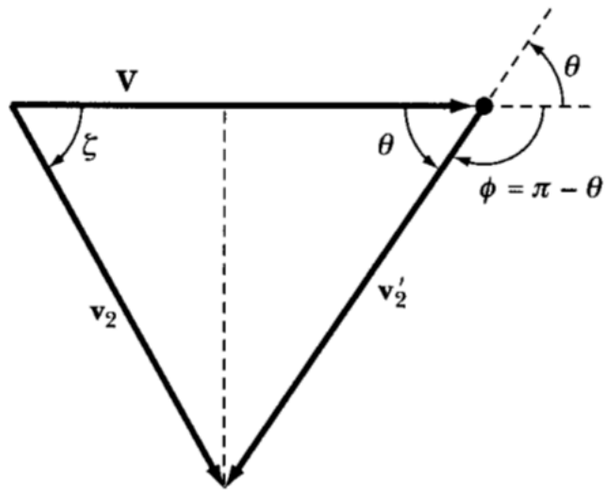


FIGURE 9-12 The final state of recoil mass m_2 in the elastic collision of two particles.

Dividing Equation 9.72a by Equation 9.72b, we have

$$\tan \zeta = \frac{v_2' \sin \theta}{V - v_2' \cos \theta} = \frac{\sin \theta}{(V/v_2') - \cos \theta}$$

But, according to Equations 9.63 and 9.65a, V and v'_1 are equal. Therefore,

$$\tan \zeta = \frac{\sin \theta}{1 - \cos \theta} = \cot \frac{\theta}{2} \tag{9.73}$$

which we may write as

$$\tan \zeta = \tan \left(\frac{\pi}{2} - \frac{\theta}{2} \right)$$

Thus,

$$2\zeta = \pi - \theta = \phi \tag{9.74}$$

For particles with equal mass, $m_1 = m_2$, we have $\theta = 2\psi$. Combining this result in Equation 9.74, we have

$$\boxed{\zeta + \psi = \frac{\pi}{2}, \qquad m_1 = m_2} \tag{9.75}$$

Hence, the scattering of particles of equal mass always produces a final state in which the velocity vectors of the particles are at right angles if one of the particles is initially at rest (see Figure 9-13). This result is valid only in the nonrelativistic limit; see Equation 14.131 for the relativistic expression governing this case.

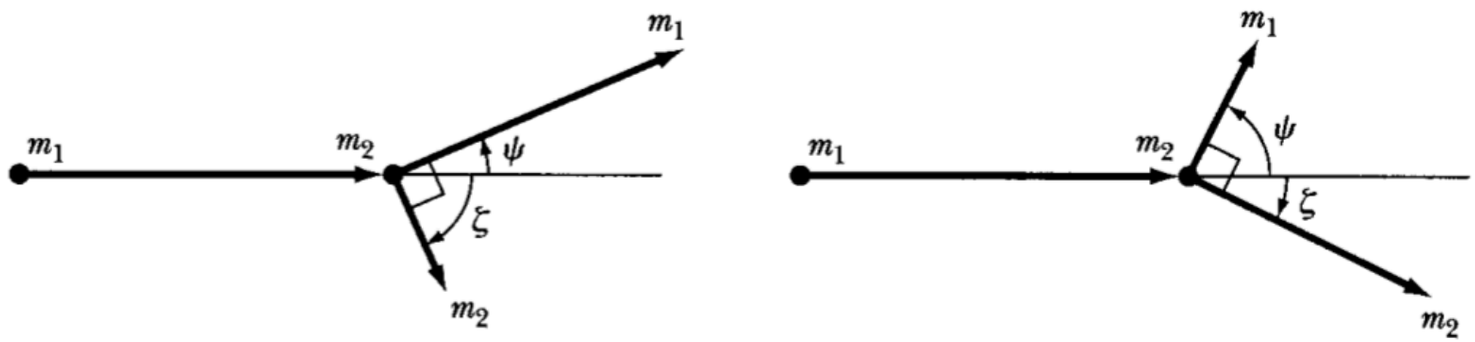


FIGURE 9-13 For the elastic scattering of two particles of equal mass ($m_1 = m_2$) with one of them initially at rest in the LAB system, the final velocities (trajectories) of the two masses are at right angles to each other. Two such possibilities are shown.

EXAMPLE 9.6

What is the maximum angle that ψ can attain for the case $V > v'_1$? What is ψ_{max} for m1 > m2 and m1 = m2?

Solution. For the case of ψ_{max} , Figure 9-llb becomes as shown in Figure 9-14.

The angle between $\mathbf{v}'_1$ and $\mathbf{v}_1$ is 90° for ψ to be a. maximum.

$$\sin \psi_{\max} = \frac{v_1'}{V} \tag{9.76}$$

According to Equation 9.68, this is just

$$\sin \psi_{\max} = \frac{m_2}{m_1}$$

from which

$$\psi_{\max} = \sin^{-1}\left(\frac{m_2}{m_1}\right) \tag{9.77}$$

For $m_1 \gg m_2$,$\psi_{max} = 0$ (no scattering), and for $m_1 = m_2$, $\psi_{max} = 90°$. Generally, for $m_1 > m_2$, no scattering of m_1 backward of 90° can occur.

9.7 Kinematics of Elastic Collisions

Relationships involving the energies of the particles may be obtained as follows. First, we have simply

$$T_0 = \frac{1}{2} m_1 u_1^2 \tag{9.78}$$

and, in the CM system.

$$T_0' = \frac{1}{2}\left(m_1 u_1'^2 + m_2 u_2'^2\right)$$

which, on using Equations 9.65a and 9.65b, becomes

$$T_0' = \frac{1}{2} \frac{m_1 m_2}{m_1 + m_2} u_1^2 = \frac{m_2}{m_1 + m_2} T_0 \tag{9.79}$$

This result shows that the initial kinetic energy in the CM system T_0' is always a fraction $m_2/(m_1 + m_2) < 1$ of the initial LAB energy. For the final CM energies, we find

$$T_1' = \frac{1}{2}m_1 v_1'^2 = \frac{1}{2}m_1\left(\frac{m_2}{m_1 + m_2}\right)^2 u_1^2 = \left(\frac{m_2}{m_1 + m_2}\right)^2 T_0 \tag{9.80}$$

and

$$T_2' = \frac{1}{2}m_2 v_2'^2 = \frac{1}{2}m_2\left(\frac{m_1}{m_1 + m_2}\right)^2 u_1^2 = \frac{m_1 m_2}{(m_1 + m_2)^2}T_0 \tag{9.81}$$

To obtain T_1 in terms of T_0, we write

$$\frac{T_1}{T_0} = \frac{\frac{1}{2}m_1 v_1^2}{\frac{1}{2}m_1 u_1^2} = \frac{v_1^2}{u_1^2} \tag{9.82}$$

Referring to Figure 9-1 la and using the cosine law, we can write

$$v_1'^2 = v_1^2 + V^2 - 2v_1 V\cos\psi$$

or

$$\frac{T_1}{T_0} = \frac{v_1^2}{u_1^2} = \frac{v_1'^2}{u_1^2} - \frac{V^2}{u_1^2} + 2\frac{v_1 V}{u_1^2}\cos\psi \tag{9.83}$$

From the previous definitions, we have

$$\frac{v_1'}{u_1} = \frac{m_2}{m_1 + m_2} \quad \text{and} \quad \frac{V}{u_1} = \frac{m_1}{m_1 + m_2} \tag{9.84}$$

The squares of these quantities give the desired expressions for the first two terms on the right-hand side of Equation 9.83. To evaluate the third term, we write, using Equation 9.66a.

$$2\frac{v_1 V}{u_1^2}\cos\psi = 2\left(v_1'\frac{\sin\theta}{\sin\psi}\right)\cdot\frac{V}{u_1^2}\cos\psi \tag{9.85}$$

The quantity of $v_1'V/u_1^2$ can be obtained from the product of the equations in Equation 9.84, and using Equation 9.69, we have

$$\frac{\sin\theta\cos\psi}{\sin\psi} = \frac{\sin\theta}{\tan\psi} = \cos\theta + \frac{m_1}{m_2}$$

so that

$$2\frac{v_1 V}{u_1^2}\cos\psi = \frac{2m_1m_2}{(m_1+m_2)^2}\left(\cos\theta + \frac{m_1}{m_2}\right) \tag{9.86}$$

Substituting Equations 9.84 and 9.86 into Equation 9.83, we obtain

$$\frac{T_1}{T_0} = \left(\frac{m_2}{m_1+m_2}\right)^2 - \left(\frac{m_1}{m_1+m_2}\right)^2 + \frac{2m_1m_2}{(m_1+m_2)^2}\left(\cos\theta + \frac{m_1}{m_2}\right)$$

which simplifies to

$$\frac{T_1}{T_0} = 1 - \frac{2m_1m_2}{(m_1+m_2)^2}(1-\cos\theta) \tag{9.87a}$$

Similarly, we can also obtain the ratio T_1/T_0 in terms of the LAB scattering angle ψ :

$$\frac{T_1}{T_0} = \frac{m_1^2}{(m_1+m_2)^2}\left[\cos\psi \pm \sqrt{\left(\frac{m_2}{m_1}\right)^2 - \sin^2\psi}\right]^2 \tag{9.87b}$$

where the plus (+) sign for the radical is to be taken unless $m_1 > m_2$ - in which case the result is double-valued, and Equation 9.77 specifies the maximum value allowed for ψ .

The LAB energy of the recoil particle m_2 can be calculated from

$$\frac{T_2}{T_0} = 1 - \frac{T_1}{T_0} = \frac{4m_1m_2}{(m_1+m_2)^2}\cos^2\zeta, \qquad \zeta \le \pi/2 \tag{9.88}$$

If $m_1 = m_2$, we have the simple relation

$$\boxed{\frac{T_1}{T_0} = \cos^2 \psi, \qquad m_1 = m_2} \tag{9.89a}$$

with the restriction noted in the discussion following Equation 9.71 that $\psi < 90°$. Also,

$$\boxed{\frac{T_2}{T_0} = \sin^2 \psi, \qquad m_1 = m_2} \tag{9.89b}$$

Several further relationships are

$$\sin \zeta = \sqrt{\frac{m_1 T_1}{m_2 T_2}} \sin \psi \tag{9.90}$$

$$\tan \psi = \frac{\sin 2\zeta}{(m_1/m_2) - \cos 2\zeta} \tag{9.91}$$

$$\tan \psi = \frac{\sin \phi}{(m_1/m_2) - \cos \phi} \tag{9.92}$$

As an example of applying the kinematic relations we have derived, consider the following situation. Suppose that we have a beam of projectiles, all with mass m_1 and energy T_0. We direct this beam toward a target consisting of a group of particles whose masses m_2 may not all be the same. Some of the incident particles interact with the target particles and are scattered. The incident particles all move in the same direction in a beam of small cross-sectional area, and we assume that the target particles are localized in space so that the scattered particles emerge from a small region. If we position a detector at, say, 90° to the incident beam and with this detector measure the energies of the scattered particles, we can display the results as in the lower portion of Figure 9-15.

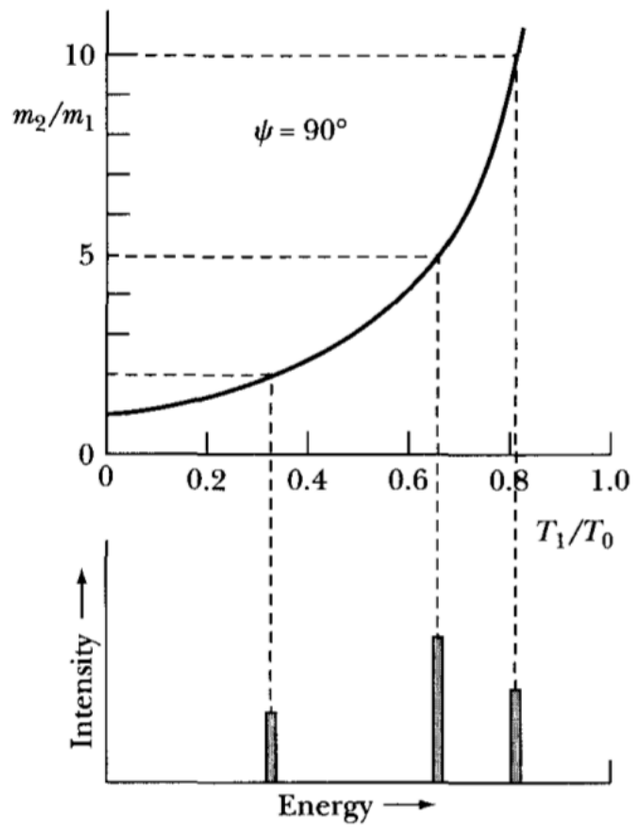


FIGURE 9-15 Results of particles of mass m_1 and energy T_0 being scattered from particles of various masses m_2 at angle $\psi = 90°$. Bottom: Histogram of number of particles detected within an energy range ΔT. Top: Curve giving scattered energy T_1 in terms of T_0 as a function of the mass ratio m_2/m_1.

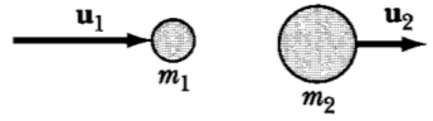
This graph is a **histogram** that plots the number of particles detected within a range of energy ΔT at the energy T. This particle histogram shows that three energy groups were observed in the particles detected at $\psi = 90°$. The upper portion of the figure shows a curve giving the scattered energy T_1 in terms of T_0 as a function of the mass ratio m_2/m_1 (Equation 9.87b). The curve can be used to determine the mass m_2 of the particle from which one of the incident particles was scattered to fall into one of the three energy groups. Thus, the energy group with $T_1 \simeq 0.8\ T_0$ results from the scattering by target particles width mass $m_2 = 10m_1$, and the other two groups result from target masses $5m_1$ and $2m_1$.

The measurement of the energies of scattered particles is therefore a method of *qualitative analysis* of the target material. Indeed, this method is useful in practice when the incident beam consists of particles (protons, say) that have been given high velocities in an accelerator of some sort.

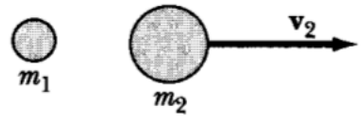
If the detector is capable of precise energy measurements, the method yields accurate information regarding the composition of the target. **Quantitative analysis** can also be made from the intensities of the groups if the cross sections are known (see the following section). Applying this technique has been useful in determining the composition of air pollution.

EXAMPLE 9.7

In a head-on elastic collision of two particles with masses m_1 and m_2, the initial velocities are $\mathbf{u}_1$ and $\mathbf{u}_2 = \alpha u1$ ($\alpha > 0$). If the initial kinetic energies of the two particles are equal in the LAB system, find the conditions on u_1/u_2 and m_1/m_2 so that m_1 will be at rest in the LAB system after the collision. See Figure 9-16.



Before collision



After collision

FIGURE 9-16 Example 9.7. Velocities are indicated for two particles of different masses in a head-on elastic collision before and after the collision.

Solution. Because the initial kinetic energies are equal, we have

$$\frac{1}{2}m_1 u_1^2 = \frac{1}{2}m_2 u_2^2 = \frac{1}{2}\alpha^2 m_2 u_1^2$$

or

$$\frac{m_1}{m_2} = \alpha^2 \tag{9.93}$$

If m_1 is at rest after the collision, the conservation of energy requires

$$\frac{1}{2}m_1 u_1^2 + \frac{1}{2}m_2 u_2^2 = \frac{1}{2}m_2 v_2^2$$

or

$$m_1 u_1^2 = \frac{1}{2} m_2 v_2^2 \tag{9.94}$$

The conservation of linear momentum states that

$$m_1 \mathbf{u}_1 + m_2 \mathbf{u}_2 = (m_1 + \alpha m_2)\mathbf{u}_1 = m_2 \mathbf{v}_2 \tag{9.95}$$

Substituting $\mathbf{v}_2$ from Equation 9.95 into Equation 9.94 gives

$$m_1 u_1^2 = \frac{1}{2} m_2 \left(\frac{m_1 + \alpha m_2}{m_2} \right)^2 u_1^2$$

or

$$m_1 = \frac{1}{2} m_2 \left(\frac{m_1}{m_2} + \alpha \right)^2 \tag{9.96}$$

Substituting $m_1/m_2 = \alpha^2$ from Equation 9.93 gives

$$2\alpha^2 = (\alpha^2 + \alpha)^2$$

with the result

$$\alpha = \sqrt{2} - 1 = 0.414$$

so that

$$\alpha^2 = 0.172$$

and

$$\frac{m_1}{m_2} = \alpha^2 = 0.172$$

$$\frac{u_2}{u_1} = \alpha = 0.414$$

Because $\alpha > 0$, both particles are traveling in the same direction; the collision is shown in Figure 9-16.

EXAMPLE 9.8

Particles of mass m_1 elastically scatter from particles of mass m_2 at rest, (a) At what LAB angle should a magnetic spectrometer be set to detect particles that lose one-third of their momentum? (b) Over what range m_1/m_2 is this possible? (c) Calculate the scattering angle for $m_1/m_2 = 1$.

Solution. We have

$$m_1 v_1 = \frac{2}{3} m_1 u_1 \quad \text{and} \quad v_1 = \frac{2}{3} u_1$$

Using Equations 9.82 and 9.87a, we have

$$\frac{T_1}{T_0} = \frac{v_1^2}{u_1^2} = \left(\frac{2}{3}\right)^2 = 1 - \frac{2m_1 m_2}{(m_1 + m_2)^2}(1 - \cos\theta) \tag{9.97}$$

This equation can be solved for $\cos\theta$, yielding

$$\cos\theta = 1 - \frac{5(m_1 + m_2)^2}{18 m_1 m_2} = 1 - y \tag{9.98}$$

where

$$y = \frac{5(m_1 + m_2)^2}{18 m_1 m_2} \tag{9.99}$$

But we need ψ , which can be obtained from Equation 9.69.

$$\tan\psi = \frac{\sin\theta}{\cos\theta + m_1/m_2} = \frac{\sqrt{2y - y^2}}{1 - y + m_1/m_2} \tag{9.100}$$

where we have used Equation 9.98 for $\cos\theta$ and found $\sin\theta = \sqrt{2y - y^2}$.

Because $\tan\psi$ must be a real number, only values for m_1/m_2 where $2 - y \geq 0$ are possible. Therefore,

$$2 - \frac{5(m_1 + m_2)^2}{18 m_1 m_2} \geq 0 \tag{9.101}$$

which can be reduced to

$$-5\left(\frac{m_1}{m_2}\right)^2 + 26\left(\frac{m_1}{m_2}\right) - 5 \geq 0$$

or

$$-5x^2 + 26x - 5 \geq 0 \tag{9.102}$$

where $x = m_1/m_2$. The solutions for x when Equation 9.102 is equal to zero are $x = 1/5, 5$. Substitution verifies that

$$\frac{1}{5} \leq \frac{m_1}{m_2} \leq 5$$

satisfies Equation 9.101, but values of m_1/m_2 outside this range do not.

Substituting $m_1/m_2 = 1$ into Equation 9.99 gives

$$y = \frac{5(m_1 + m_2)^2}{18 m_1 m_2} = \frac{5\left(\frac{m_1}{m_2} + 1\right)^2}{18 m_1/m_2}$$

$$= \frac{5(1 + 1)^2}{18} = \frac{10}{9}$$

and substituting for y into Equation 9.100 gives $\psi = 48°$

9.8 Inelastic Collisions

When two particles interact, many results are possible, depending on the forces involved. In the previous two sections, we were restricted to elastic collisions. But, in general, multiparticles may be produced if large changes of energy are involved. For example, when a proton collides with some nuclei, energy may be released. In addition, the proton may be absorbed, and the collision may produce a neutron or alpha particle instead. All these possibilities are handled with the same methods: conservation of energy and linear momentum. We continue to restrict our considerations to the same particles in the final system as were considered in the initial system. In general, the conservation of energy is

$$Q + \frac{1}{2}m_1u_1^2 + \frac{1}{2}m_2u_2^2 = \frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 \tag{9.103}$$

where Q is called the Q-value and represents the energy loss or gain in the collision.

$Q = 0$: Elastic collision, kinetic energy is conserved
$Q > 0$: Exoergic collision, kinetic energy is gained
$Q < 0$: Endoergic collision, kinetic energy is lost

An inelastic collision is an example of an endoergic collision. The kinetic energy may be converted to mass-energy, as, for example, in a nuclear collision. Or it may be lost as heat energy, as, for example, by frictional forces in a collision. The collisions of all macroscopic bodies are endoergic (inelastic) to some degree. Two silly putty balls with equal masses and speeds striking head-on may come to a complete stop, a totally inelastic collision. Even two billiard balls colliding do not completely conserve kinetic energy; some small fraction of the initial kinetic energy is converted to heat.

A measure of the inelasticity of two bodies colliding may be considered by referring to a direct head-on collision (see Figure 9-17) in which no rotations are involved (translational kinetic energy only).

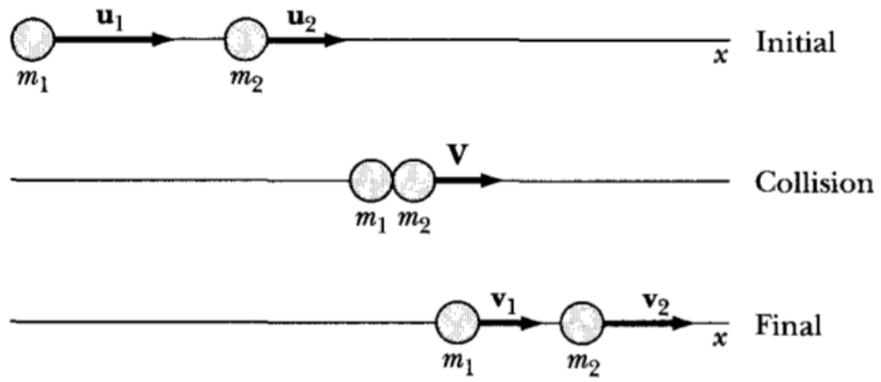


FIGURE 9-17 Direct head-on collision between two bodies indicating the initial conditions, the collision, and the resulting situation.

Newton found experimentally that the ratio of the relative initial velocities to the relative final velocities was nearly constant for any two bodies. This ratio, called the **coefficient of restitution** (ε), is defined by

$$\varepsilon = \frac{|v_2 - v_1|}{|u_2 - u_1|} \tag{9.104}$$

This is sometimes called Newton's rule. For a perfectly elastic collision, $\varepsilon = 1$; and for a totally inelastic collision, $\varepsilon = 0$. Values of ε have the limits 0 and 1.

We must be careful when applying Equation 9.104 to oblique collisions, because Newton's rule applies only to the velocity components along the normal (aa') to the plane of contact (bb') between the two bodies, as shown in Figure 9-18. For smooth surfaces, the velocity components along the plane of contact are hardly changed by the collision.

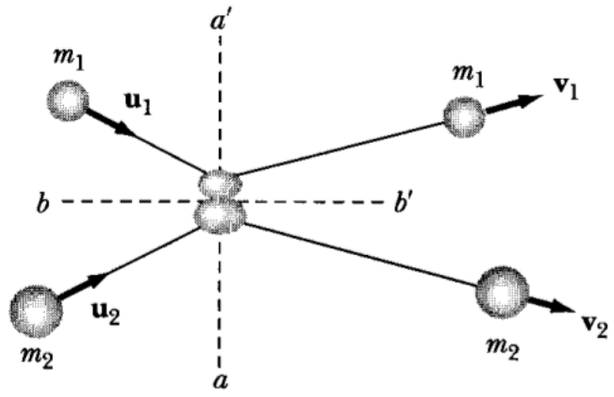


FIGURE 9-18 An oblique collision between two bodies. For smooth surfaces, the velocity components along the plane of contact bb' are hardly changed by the collision.

EXAMPLE 9.9

For an elastic head-on collision described in Sections 9.6 and 9.7, show that $\varepsilon = 1$. The mass m_2 is initially at rest.

Solution. Because the final velocities are along the same direction as $\mathbf{u}_1$, we state the conservation of linear momentum and energy as

$$m_1 u_1 = m_1 v_1 + m_2 v_2 \tag{9.105}$$

$$\frac{1}{2} m_1 u_1^2 = \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 \tag{9.106}$$

We solve Equation 9.105 for v_2 and substitute into the equation for ε

$$\varepsilon = \frac{v_2 - v_1}{u_1} = \frac{\dfrac{m_1 u_1 - m_1 v_1}{m_2} - v_1}{u_1} = \frac{m_1}{m_2} - \frac{m_1}{m_2}\frac{v_1}{u_1} - \frac{v_1}{u_1} \tag{9.107}$$

We can find the ratio v_1/u_1 from Equation 9.106 after substituting for v_2 from Equation 9.105:

$$\frac{1}{2} m_1 u_1^2 = \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 \left(\frac{m_1 u_1 - m_1 v_1}{m_2} \right)^2$$

$$m_1 u_1^2 = m_1 v_1^2 + \frac{m_1^2}{m_2} (u_1^2 + v_1^2 - 2 u_1 v_1)$$

Dividing by $m_1 u_1^2$ and letting $x = v_1/u_1$ gives

$$1 = x^2 + \frac{m_1}{m_2} (1 + x^2 - 2x)$$

Collecting terms,

$$\left(1 + \frac{m_1}{m_2} \right) x^2 - \frac{2 m_1}{m_2} x + \left(\frac{m_1}{m_2} - 1 \right) = 0$$

Using the quadratic equation to solve for x. we find

$$x = 1$$

and

$$x = \frac{\frac{m_1}{m_2} - 1}{\frac{m_1}{m_2} + 1}$$

The solution x = 1 is trivial ($v_1 = u_1$, $v_2 = 0$), so we substitute the other solution for x into Equation 9.107:

$$\varepsilon = \frac{m_1}{m_2} - \frac{\frac{m_1}{m_2}\left(\frac{m_1}{m_2} - 1\right)}{\frac{m_1}{m_2} + 1} + \frac{\left(\frac{m_1}{m_2} - 1\right)}{\frac{m_1}{m_2} + 1}$$

$$\varepsilon = \frac{\frac{m_1^2}{m_2^2} + \frac{m_1}{m_2} - \frac{m_1^2}{m_2^2} + \frac{m_1}{m_2} - \frac{m_1}{m_2} + 1}{\frac{m_1}{m_2} + 1} = 1$$

During a collision (elastic or inelastic), the forces involved may act over a very short period of time and are called impulsive forces. A hammer striking a nail and two billiard balls colliding are examples of impulsive forces. Newton's Second Law is still valid throughout the time period Δt of the collision:

$$\mathbf{F} = \frac{d}{dt}(m\mathbf{v}) \tag{9.108}$$

After multiplying by dt and integrating, we have

$$\int_{t_1}^{t_2} \mathbf{F}\, dt \equiv \mathbf{P} \tag{9.109}$$

where $\Delta t = t_2 - t_1$.

Equation 9.109 defines the term **impulse P**. The impulse may be measured experimentally by the change of momentum. An ideal impulse represented by no displacement during the collision would be caused by an infinite force acting during an infinitesimal time.

EXAMPLE 9.10

Consider a rope of mass per unit length ρ and length a suspended just above a table as shown in Figure 9-19. If the rope is released from rest at the top, find the force on the table when a length x of the rope has dropped to the table.

Solution. We have a gravitational force of mg = ρ xg because the rope lies on the table, but we need to consider the impulsive force as well.

$$F = \frac{dp}{dt} \qquad (9.110)$$

During the time interval dt, the mass of rope equal to ρ (vdt) drops to the table. The change in momentum imparted to the table is

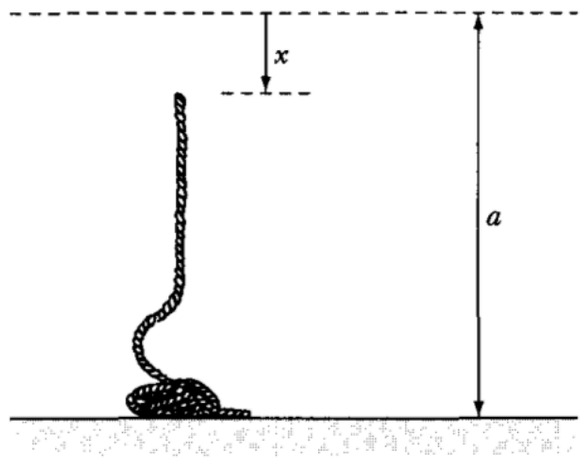


FIGURE 9-19 Example 9.10. A rope of length *a* is released while suspended just above a table. We want to find the force *F* on the table after the rope has dropped a distance *x*.

$$dp = (\rho v\, dt)\, v = \rho v^2\, dt$$

and

$$\frac{dp}{dt} = \rho v^2 = F_{\text{impulse}} \qquad (9.111)$$

The velocity v is related to x at time thy v2 = 2gx, because each part of the remaining rope is under constant acceleration g.

$$F_{\text{impulse}} = \rho v^2 = 2\rho x g \qquad (9.112)$$

The total force is the sum of the gravitational and impulsive forces:

$$F = F_g + F_{\text{impulse}} = 3\rho x g \qquad (9.113)$$

which is equivalent to the weight of a length 3x of the rope.

9.9 Scattering Cross Sections

In the preceding sections, we derived various relationships connecting the initial state of a moving particle with the final states of the original particle and a struck particle. Only kinematic relationships were involved; that is, no attempt was made to predict a scattering angle or a final velocity - only equations connecting these quantities were obtained. We now look more closely at the collision process and investigate the scattering in the event that the particles interact with a specified force field. Consider the situation depicted in Figure 9-20, which illustrates such a collision in the LAB coordinate system when a repulsive force exists between m_1 and m_2. The particle m_1 approaches the vicinity of m_2 in such a way that if there were no force acting between the particles, m_1 would pass m_2 with a distance of closest approach b. The quantity b is called the **impact parameter**. If the velocity of m_1 is u_1, then the impact parameter b determines the angular momentum l of particle m_1 about m_2:

$$l = m_1 u_1 b \tag{9.114}$$

We may express u_1 in terms of the incident energy T_0 by using Equation 9.78:

$$l = b\sqrt{2m_1 T_0} \tag{9.115}$$

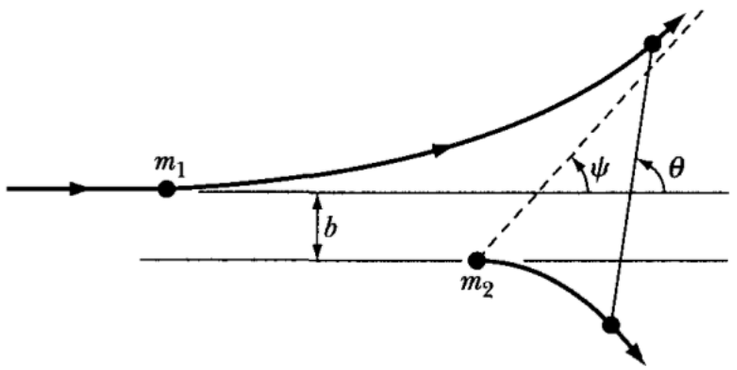


FIGURE 9-20 Particle m_1 approaches m_2, initially at rest, in the LAB system, and the repulsive force between particles results in scattering. Had m_1 continued in a straight line, its closest distance of approach to m_2 would have been b, the impact parameter.

Evidently, for a given energy T_0, the angular momentum and hence the scattering angle θ (or ψ) is uniquely specified by the impact parameter b if the force law is known.

In the scattering of atomic or nuclear particles, we can neither choose nor measure directly the impact parameter. We are therefore reduced, in such situations, to speaking in terms of the probability for scattering at various angles θ .

We now consider the distribution of scattering angles that result from collisions with various impact parameters. To accomplish this, let us assume that we have a narrow beam of particles, each having mass m_1 and energy T_0. We direct this beam toward a small region of space containing a collection of particles, each of which has mass m_2 and is at rest (in the LAB system). We define the **intensity** (or **flux density**) I of the incident particles as the number of particles passing in unit time through a unit area normal to the direction of the beam. If we assume that the force law between m_1 and m_2 falls off with distance sufficiently rapidly. then after an encounter, the motion of a scattered particle asymptotically approaches a straight line with a well-defined angle θ between the initial and final directions of motion. We now define a **differential scattering cross section** $\sigma(\theta)$ in the CM system for the scattering into an element of solid angle $d\Omega'$ at a particular CM angle θ :

$$\sigma(\theta) = \frac{\begin{pmatrix}\text{Number of interactions per target particle that} \\ \text{lead to scattering into } d\Omega' \text{ at the angle } \theta\end{pmatrix}}{\text{Number of incident particles per unit area}} \tag{9.116}$$

If dN is the number of particles scattered into $d\Omega'$ per unit time, then the probability of scattering into $d\Omega'$ for a unit area of the incident beam is

$$\sigma(\theta)\, d\Omega' = \frac{dN}{I} \tag{9.117a}$$

We sometimes write, alternatively,

$$\sigma(\theta) = \frac{d\sigma}{d\Omega'} = \frac{1}{I}\frac{dN}{d\Omega'} \tag{9.117b}$$

The fact that $\sigma(\theta)$ has the dimensions of area per steradian gives rise to the term *cross section*. If the scattering has axial symmetry (as for central forces), we can immediately perform the integration over the azimuthal angle to obtain 2π, and then the element of solid angle $d\Omega'$ is given by

$$d\Omega' = 2\pi \sin\theta \, d\theta \tag{9.118}$$

If we return, for the moment, to the equivalent one-body problem discussed in the preceding chapter, we can consider the scattering of a particle of mass μ by a force center. For such a case, Figure 9-21 shows that the number of particles with impact parameters within a range db at a distance b must correspond to the number of particles scattered into the angular range $d\theta$ at an angle θ .

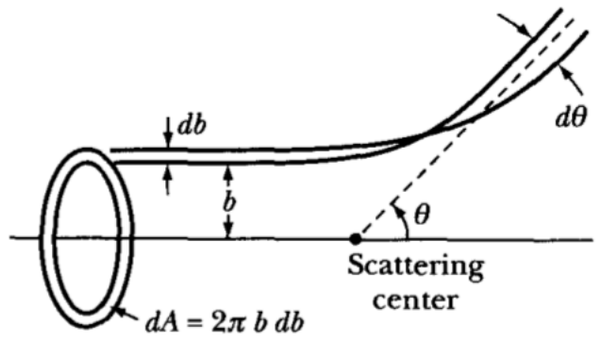


FIGURE 9-21 The equivalent one-body problem has mass μ scattered by a force center in the CM system. Particles within a range db around impact parameter b scatter into the angular range $d\theta$ at the angle θ.

Therefore,

$$I \cdot 2\pi b \, db = -I \cdot \sigma(\theta) \cdot 2\pi \sin\theta \, d\theta \tag{9.119}$$

where db/$d\theta$ is negative, because we assume that the force law is such that the amount of angular deflection decreases (monotonically) with increasing impact parameter. Hence,

$$\sigma(\theta) = \frac{b}{\sin\theta}\left|\frac{db}{d\theta}\right| \tag{9.120}$$

We can obtain the relationship between the impact parameter b and the scattering angle θ by using Figure 9-22.

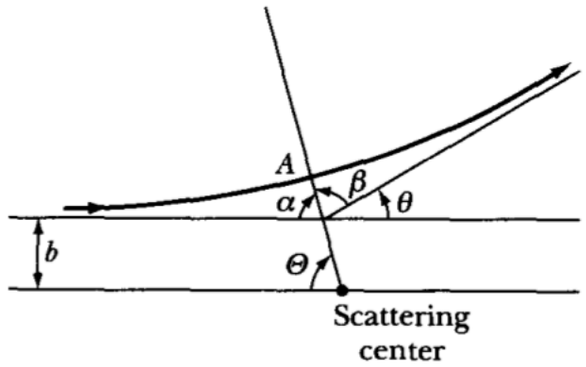


FIGURE 9-22 The geometry of particle scattering in a central-force field. Point *A* is the distance of closest approach.

In the preceding chapter, we found (in Equation 8.31) that the change in angle θ for a particle of mass μ moving in a central-force field was given by

$$\Delta\Theta = \int_{r_{\min}}^{r_{\max}} \frac{(l/r^2)\,dr}{\sqrt{2\mu[E - U - (l^2/2\mu r^2)]}} \tag{9.121}$$

The angles α and β are therefore equal and, in fact, are equal to Θ . Thus,

$$\theta = \pi - 2\Theta \tag{9.122}$$

For the case that $r_{max} = \infty$, the angle Θ is given b

$$\Theta = \int_{r_{\min}}^{\infty} \frac{(b/r^2)\,dr}{\sqrt{1 - (b^2/r^2) - (U/T_0')}} \tag{9.123}$$

where use has been made of the one-body equivalent of Equation 9.115:

$$l = b\sqrt{2\mu T_0'}$$

where, as in Equation 9.79, $T_0' = \frac{1}{2}\mu u_1^2$. We have also used $E = T_0'$ because the total energy E must equal the kinetic energy T_0' at $r = \infty$ where U= 0. The value of r_{min} is a root of the radical in the denominator in Equations 9.121 or 9.123 - that is, r_{min} is a turning point of the motion and corresponds to the distance of closest approach of the particle to the force center. Thus, Equations 9.122 and 9.123 give the dependence of the scattering angle θ on the impact parameter b. Once we know b = b(θ) for a given potential U(r) and a given value of T_0' , we can calculate the differential scattering cross section from Equation 9.120. This procedure leads to the scattering cross section in the CM system, because we have been considering m_2 as a fixed force center. If $m_2 \gg m_1$, the cross section so obtained is very close to the LAB system cross section; but if m_1 cannot be considered negligible compared with m_2, the proper transformation of solid angles must be made. We now obtain the general relations.

Because the total number of particles scattered into a unit solid angle must be the same in the LAB system as in the CM system, we have

$$\sigma(\theta)\, d\Omega' = \sigma(\psi)\, d\Omega$$
$$\sigma(\theta) \cdot 2\pi \sin\theta d\theta = \sigma(\psi) \cdot 2\pi \sin\psi d\psi \tag{9.124}$$

where θ and ψ represent the *same* scattering angle but measured in the CM or LAB system, respectively, and where $d\Omega'$ and $d\Omega$, represent the *same* element, of solid angle but measured in the CM or LAB system, respectively. Therefore, $\sigma(\theta)$ and $\sigma(\psi)$ are the differential cross sections for the scattering in the CM and LAB systems, respectively. Thus,

$$\sigma(\psi) = \sigma(\theta) \cdot \frac{\sin\theta}{\sin\psi} \frac{d\theta}{d\psi} \tag{9.125}$$

The derivative $d\theta/d\psi$ can be evaluated by first referring to Figure 9-11 a and writing, from the sine law (and using Equations 9.63 and 9.65b),

$$\frac{\sin(\theta - \psi)}{\sin \psi} = \frac{m_1}{m_2} \equiv x \tag{9.126}$$

We set the differential dx= 0 and find

$$dx = 0 = \frac{\partial x}{\partial \psi} d\psi + \frac{\partial x}{\partial \theta} d\theta$$

which gives, after taking the partial derivatives and collecting terms,

$$\frac{d\theta}{d\psi} = \frac{\sin(\theta - \psi) \cos \psi}{\cos(\theta - \psi) \sin \psi} + 1$$

Expanding sin($\theta - \psi$) and simplifying, we have

$$\frac{d\theta}{d\psi} = \frac{\sin \theta}{\cos(\theta - \psi) \sin \psi}$$

and so

$$\sigma(\psi) = \sigma(\theta) \cdot \frac{\sin^2 \theta}{\cos(\theta - \psi) \sin^2 \psi} \tag{9.127}$$

Multiplying both sides of Equation 9.126 by cos ψ and then adding cos($\theta - \psi$)to both sides, we have

$$\frac{\sin(\theta - \psi) \cos \psi}{\sin \psi} + \cos(\theta - \psi) = x \cos \psi + \cos(\theta - \psi)$$

Expanding sin($\theta - \psi$) and cos($\theta - \psi$) on the left-hand side, we obtain

$$\frac{\sin \theta}{\sin \psi} = x \cos \psi + \cos(\theta - \psi)$$

Substituting this result into Equation 9.127,

$$\sigma(\psi) = \sigma(\theta) \cdot \frac{[x \cos \psi + \cos(\theta - \psi)]^2}{\cos(\theta - \psi)}, \quad (x < 1) \tag{9.128}$$

And from Equation 9.126, we have

$$\cos(\theta - \psi) = \sqrt{1 - x^2 \sin^2 \psi}$$

Hence,

$$\boxed{\sigma(\psi) = \sigma(\theta) \cdot \frac{\left[x \cos \psi + \sqrt{1 - x^2 \sin^2 \psi}\right]^2}{\sqrt{1 - x^2 \sin^2 \psi}}} \tag{9.129}$$

Equation 9.126 can be used to write

$$\boxed{\theta = \sin^{-1}(x \sin \psi) + \psi} \tag{9.130}$$

Equations 9.129 and 9.130 therefore specify the cross section entirely in terms of the angle ψ . For the general case (i.e., for an arbitrary value of x), the evaluation of $\sigma(\psi)$ is complicated. Tables exist, however, so the particular cases can be computed with relative ease.

The transformation represented by Equations 9.129 and 9.130 assumes a simple form for two cases. For $x = m_1/m_2 = 1$, we have from Equation 9.71, $\theta = 2\psi$, and Equation 9.129 becomes

$$\sigma(\psi) = \sigma(\theta)|_{\theta=2\psi} \cdot 4 \cos \psi, \quad m_1 = m_2 \tag{9.131}$$

and for $m_1 \ll m_2$, $x \simeq 0$, and $\theta \simeq \psi$, so that

$$\sigma(\psi) \cong \sigma(\theta)|_{\theta=\psi}, \quad m_1 \ll m_2 \tag{9.132}$$

EXAMPLE 9.11

Consider molecules of radius R_1 moving toward the right with identical velocities scattering from dust particles of radius R_2 that are at rest. Consider both as hard spheres and find the differential and total cross sections.

Solution. The dust particles are at rest, and we will solve this scattering problem in the LAB system. Consider the geometry of scattering in Figure 9-23. The particles with impact parameter b will be scattered at angle ψ . Similarly to Equation 9.119, incident particles entering within a range of impact parameters db scatter into an angular range $d\psi$, and we have

$$2\pi b\, db = -\sigma(\psi) \cdot 2\pi \sin\psi\, d\psi \tag{9.133}$$

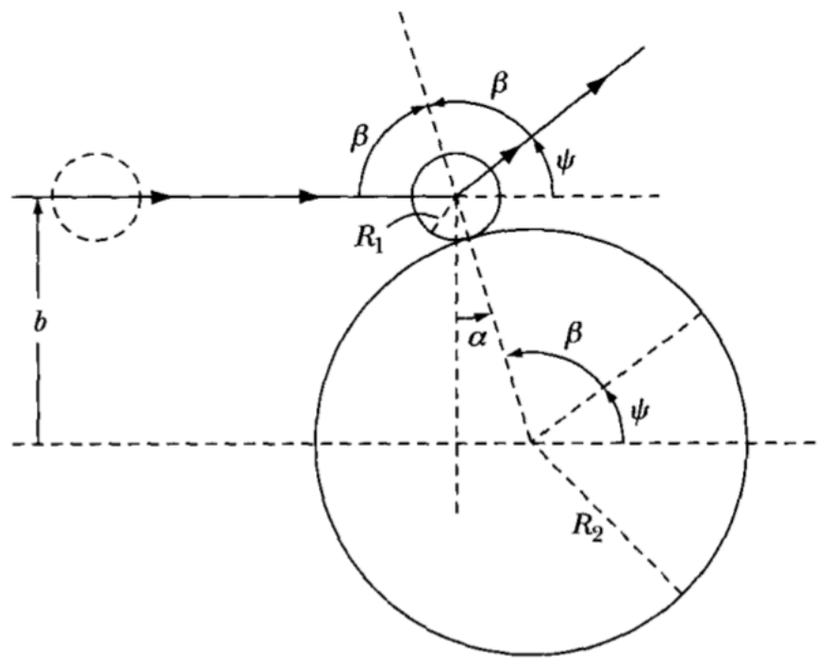


FIGURE 9-23 A molecule of radius R_1 approaches a dust particle of radius R_2 from the left and scatters at angle ψ.

In order to find the differential cross section $\sigma(\psi)$, we have to find the relationship between the impact parameter b and scattering angle ψ . We see from Figure 9-23 that $b = (R_1 + R_2)\cos\alpha$, so we need to first find the relationship between angles α and ψ .

Look closely at Figure 9-23 to see that $2\beta + \psi = \pi, \alpha = \psi + \beta - \pi/2$, and

$$\alpha = \psi + \left(\frac{\pi}{2} - \frac{\psi}{2}\right) - \frac{\pi}{2} = \frac{\psi}{2} \tag{9.134}$$

Before using Equation 9.133, we need to find the differential db. We have

and

$$b = (R_1 + R_2)\cos\alpha = (R_1 + R_2)\cos\frac{\psi}{2}$$

$$db = -\frac{(R_1 + R_2)}{2}\sin\frac{\psi}{2}\,d\psi.$$

We now insert the terms into Equation 9.133 to find

$$-\frac{2\pi}{2}(R_1 + R_2)^2\cos\frac{\psi}{2}\sin\frac{\psi}{2}\,d\psi = -\sigma(\psi)\cdot 2\pi\sin\psi\,d\psi$$

If we use the identity, $\sin\psi = 2\sin(\psi/2)\cos(\psi/2)$, we finally find

$$\sigma(\psi) = \frac{1}{4}(R_1 + R_2)^2 \tag{9.135}$$

We first note that the differential scattering cross section is isotropic, the scattering is the same in every direction. This is somewhat surprising, because differential scattering cross sections normally have an angular dependence.

We will discuss the total cross section in the next section, but briefly it is proportional to the probability that any scattering takes place. In order to determine the total cross section, we must integrate Equation 9.135 over all possible ψ. Note that we have already done so over the azimuthal angle to determine Equation 9.118. We have

$$\sigma_t = \int\frac{d\sigma}{d\psi}d\Omega = \int\sigma(\psi)\,d\Omega$$

$$= \int_0^{\pi}\frac{1}{4}(R_1 + R_2)^2 2\pi\sin\psi\,d\psi$$

$$= \frac{\pi}{2}(R_1 + R_2)^2\int_0^{\pi}\sin\psi\,d\psi = -\frac{\pi}{2}(R_1 + R_2)^2\cos\psi\Big|_0^{\pi}$$

$$\sigma_t = \pi(R_1 + R_2)^2$$

This is precisely what we would expect for the total cross section for scattering of two hard spheres. The maximum area occurs when the molecule and dust particle have just a glancing blow, angle $\alpha = 0$. The impact parameter will be simply b = $R_1 + R_2$, and the area is πb^2 .

9.10 Rutherford Scattering Formula

One of the most important problems that makes use of the formulas developed in the preceding section is the scattering of charged particles in a Coulomb or electrostatic field. The potential for this case is

$$U(r) = \frac{k}{r} \tag{9.136}$$

where $k = q_1 q_2 / 4\pi\varepsilon_0$, with q1 and q2 the amounts of charge that the two particles carry (k may be either positive or negative, depending on whether the charges are of the same or opposite sign; k > 0 corresponds to a repulsive force and k < 0 to an attractive force). Equation 9.123 then becomes

$$\Theta = \int_{r_{\min}}^{\infty} \frac{(b/r)\,dr}{\sqrt{r^2 - (k/T_0')r - b^2}} \tag{9.137}$$

which can be integrated to obtain (see the integration of Equation 8.38):

$$\cos\Theta = \frac{(\kappa/b)}{\sqrt{1 + (\kappa/b)^2}} \tag{9.138}$$

where

$$\kappa \equiv \frac{k}{2T_0'}$$

Equation 9.138 can be rewritten as

$$b^2 = \kappa^2 \tan^2\Theta$$

But Equation 9.122 states that $\Theta = \pi/2 - \theta/2$, so

$$b = \kappa \cot(\theta/2) \tag{9.139}$$

Thus,

$$\frac{db}{d\theta} = -\frac{\kappa}{2}\frac{1}{\sin^2(\theta/2)}$$

Equation 9.120 thus becomes

$$\sigma(\theta) = \frac{\kappa^2}{2}\cdot\frac{\cot(\theta/2)}{\sin\theta\sin^2(\theta/2)}$$

Now,

$$\sin\theta = 2\sin(\theta/2)\cos(\theta/2)$$

Hence,

$$\sigma(\theta) = \frac{\kappa^2}{4}\cdot\frac{1}{\sin^4(\theta/2)}$$

or

$$\boxed{\sigma(\theta) = \frac{k^2}{(4T_0')^2}\cdot\frac{1}{\sin^4(\theta/2)}} \qquad (9.140)$$

which is the Rutherford scattering formula and demonstrates the dependence of the CM scattering cross section on the inverse fourth power of $\sin(\theta/2)$. Note that $\sigma(\theta/2)$ is independent of the sign of k, so that the form of the scattering distribution is the same for an attractive force as for a repulsive one. It is also rather remarkable that the quantum-mechanical treatment of Coulomb scattering leads to exactly the same result as does the classical derivation. This is indeed a fortunate circumstance because, if it were otherwise, the disagreement at this early stage between classical theory and experiment might have seriously delayed the progress of nuclear physics.

For the case $m_1 = m_2$, Equation 9.79 states that $T_0' = T_0/2$, so that

$$\sigma(\theta) = \frac{k^2}{4T_0^2}\cdot\frac{1}{\sin^4(\theta/2)}, \qquad m_1 = m_2 \qquad (9.141)$$

Or, from Equation 9.131,

$$\sigma(\psi) = \frac{k^2}{T_0^2} \frac{\cos\psi}{\sin^4\psi}, \qquad m_1 = m_2 \tag{9.142}$$

NOTE: This form of the scattering law was verified for the interaction of a particles and heavy nuclei by the experiments of H. Geiger and E. Marsden

NOTE: Bohr showed that the identity of the results is a consequence of the $1/r^2$ nature of the force; it cannot be expected for any other type of force law

All the preceding discussion applies to the calculation of *differential* scattering cross sections. If it is desired to know the probability that *any* interaction *whatsoever* will take place, it is necessary to integrate $\sigma(\theta)$ [or $\sigma(\psi)$] over all possible scattering angles. The resulting quantity is called the **total scattering cross section** (σ_t) and is equal to the effective area of the target particle for producing a scattering event:

$$\sigma_t = \int_{4\pi} \sigma(\theta)\, d\Omega' = 2\pi \int_0^{\pi} \sigma(\theta) \sin\theta\, d\theta \tag{9.143}$$

where the integration over θ runs from 0 to π. The *total* cross section is the same in the LAB as in the CM system. If we wish to express the total cross section in terms of an integration over the LAB quantities.

$$\sigma_t = \int \sigma(\psi)\, d\Omega$$

then if $m_1 < m_2$, ψ also runs from 0 to π. If $m_1 \geq m_2$, ψ runs only up to ψ_{max} (given by Equation 9.77), and we have

$$\sigma_t = 2\pi \int_0^{\psi_{max}} \sigma(\psi) \sin\psi\, d\psi \tag{9.144}$$

If we attempt to calculate σ_t , for the case of Rutherford scattering, we find that the result is infinite.This occurs because the Coulomb potential, which varies as 1/r, falls off so slowly that, as the impact parameter b becomes indefinitely large, the decrease in scattering angle is too slow to prevent the integral from diverging. We have, however, pointed out in Example 8.6 that the Coulomb field of a real atomic nucleus is screened by the surrounding electrons so that the potential is effectively cut off at large distances. The evaluation of the scattering cross section for a screened Coulomb potential according to the classical theory is quite complicated and is not discussed here; the quantum-mechanical treatment is actually easier for this case.

9.11 Rocket Motion

The motion of a simple rocket is an interesting application of elementary Newtonian dynamics and might have been covered in Chapter 2. However, we want to include more complicated rockets with exhaust masses and multiple rocket stages, so we deferred the discussion to this chapter on systems of particles. The two cases we examine are rocket motion in free space and the vertical ascent of rockets under gravity. The first case requires an application of the conservation of linear momentum. The second case requires a more complicated application of Newton's Second Law.

Rocket Motion in Free Space

We assume here that the rocket (space ship) moves under the influence of no external forces. We choose a closed system in which Newton's Second Law can be applied. In outer space, the motion of the space ship must depend entirely on its own energy. It moves by the reaction of ejecting mass at high velocities. To conserve linear momentum, the space ship will have to move in the opposite direction. The diagram of the space ship motion is shown in Figure 9-24. At some time t, the instantaneous total mass of the space ship in m, and the instantaneous speed of the space ship is v with respect to an inertial reference system. We assume that all motion is in the x direction and eliminate the vector notation.

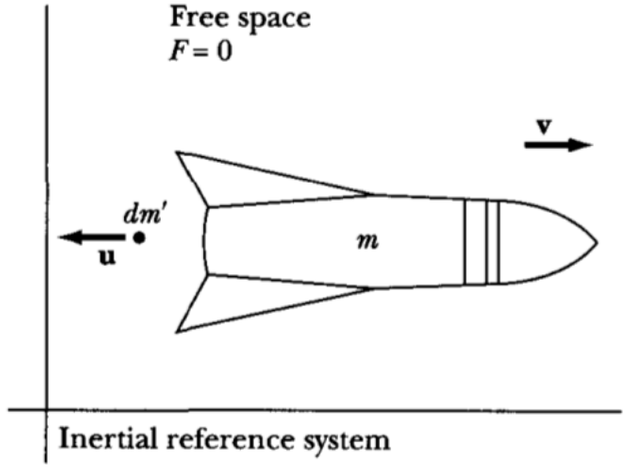


FIGURE 9-24 A rocket moves in free space at velocity **v**. In the time interval *dt*, a mass *dm'* is ejected from the rocket engine with velocity **u** with respect to the rocket ship.

During a time interval dt, a positive mass dm' is ejected from the rocket engine with a speed -u with respect to the space ship. Immediately after the mass dm' is ejected, the space ship's mass and speed are m - dm' and v + dv, respectively.

$$\text{Initial momentum} = mv \quad (\text{at time } t) \tag{9.145}$$

$$\text{Final momentum} = \underbrace{(m - dm')(v + dv)}_{\text{space ship less } dm'} + \underbrace{dm'(v - u)}_{\text{rocket exhaust } dm'} \quad (\text{at time } t + dt) \tag{9.146}$$

Notice that the speed of the ejected mass dm' with respect to the reference system is v - u. The conservation of linear momentum requires that Equations 9.145 and 9.146 be equal. There are no external forces ($F_{ext} = 0$).

$$\begin{aligned} p_{\text{initial}} &= p_{\text{final}} \\ p(t) &= p(t + dt) \\ mv &= (m - dm')(v + dv) + dm'(v - u) \\ mv &= mv + m\,dv - v\,dm' - dm'\,dv + v\,dm' - u\,dm' \\ m\,dv &= u\,dm' \end{aligned} \tag{9.147}$$

$$dv = u\frac{dm'}{m} \tag{9.148}$$

where we have neglected the product of two differentials dm'dv. We have considered dm' to be a positive mass ejected from the space ship. The change in mass of the space ship itself is dm, where

$$dm = -dm' \tag{9.149}$$

and

$$dv = -u\frac{dm}{m} \tag{9.150}$$

because dm must be negative. Let m_0 and v_0 be the initial mass and speed of the space ship, respectively, and integrate Equation 9.150 to its final values m and v.

$$\int_{v_0}^{v} dv = -u\int_{m_0}^{m}\frac{dm}{m}$$

$$v - v_0 = u\ln\left(\frac{m_0}{m}\right) \tag{9.151}$$

$$v = v_0 + u\ln\left(\frac{m_0}{m}\right) \tag{9.152}$$

The exhaust velocity u is assumed constant. Thus, to maximize the space ship's speed, we need to maximize the exhaust velocity u and the ratio m_0/m.

Because the terminal speed is limited by the ratio m_0/m, engineers have constructed multistage rockets. The minimum mass (less fuel) of the space ship is limited by structural material. However, if the fuel container itself is jettisoned after its fuel has been burned, the mass of the remaining space ship is even less. The space ship can contain two or more fuel containers, each of which can be jettisoned.

For example, let

m_0 = Initial total mass of space ship

$m_1 = m_a + m_b$

m_a = Mass of first-stage payload

m_b = Mass of first-stage fuel containers, etc.

v_1 = Terminal speed of first stage of "burnout" after all fuel is burned

$$v_1 = v_0 + u \ln\left(\frac{m_0}{m_1}\right) \tag{9.153}$$

At burnout, the terminal speed v_1 of the first stage is reached, and the mass m_b is released into space. Next, the second-stage rocket ignites with the same exhaust velocity, and we have

m_a = Initial total mass of space ship second stage

$m_2 = m_c + m_d$

m_c = Mass of second-stage payload

m_d = Mass of second-stage fuel container, etc.

v_1 = Initial speed of second stage

v_2 = Terminal speed of second stage at burnout

$$v_2 = v_1 + u \ln\left(\frac{m_a}{m_2}\right) \tag{9.154}$$

$$v_2 = v_0 + u \ln\left(\frac{m_0 m_a}{m_1 m_2}\right) \tag{9.155}$$

The product $(m_0 m_a / m_1 m_2)$ can be made much larger than just m_0/m_1. Multistage rockets are more commonly used in ascent under gravity than in free space.

We have seen that the space ship is propelled as a result of the conservation of linear momentum. But engineers and scientists like to refer to the force term as rocket "thrust." If we multiply Equation 9.150 by m and divide by dt, we have

$$m\frac{dv}{dt} = -u\frac{dm}{dt} \tag{9.156}$$

Since the left side of this equation "appears" as ma(force), the right side is called thrust:

$$\text{Thrust} \equiv -u\frac{dm}{dt} \tag{9.157}$$

Because dm/dt is negative, the thrust is actually positive.

Vertical Ascent Under Gravity

The actual motion of a rocket attempting to leave Earth's gravitational field is quite complicated. For analytical purposes, we begin by making several assumptions. The rocket will have only vertical motion, with no horizontal component. We neglect air resistance and assume that the acceleration of gravity is constant with height. We also assume that the burn rate of the fuel is constant. All these factors that are neglected can reasonably be included with a numerical analysis by computer.

We can use the results of the previous case of rocket motion in free space, but we no longer have $F_{ext} = 0$. The geometry is shown in Figure 9-25.

We again have dm' as positive, with dm = -dm'. The external force F_{ext} is

$$F_{ext} = \frac{d}{dt}(mv)$$

or

$$F_{ext}dt = d(mv) = dp = p(t + dt) - p(t) \tag{9.158}$$

over a small differential time.

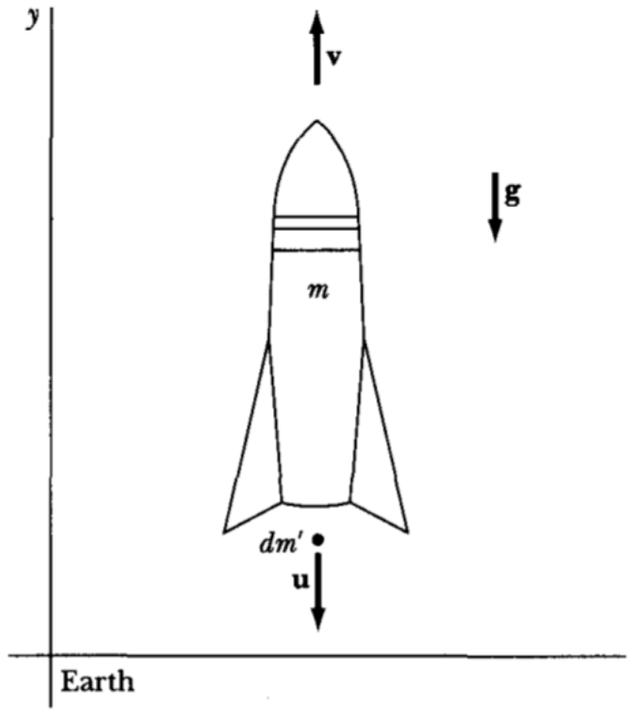


FIGURE 9-25 A rocket in vertical ascent under Earth's gravity. Mass dm' is ejected from the rocket engine with velocity **u** with respect to the rocket ship.

For the space ship system, we found the initial and final momenta in Equations 9.145-9.150. We now use the results leading up to Equation 9.150 to obtain

$$p(t + dt) - p(t) = m dv + u dm \tag{9.159}$$

In free space, $F_{ext} = 0$, but in ascent, $F_{ext} = -mg$. Combining Equations 9.158 and 9.159 gives

$$\begin{aligned} F_{ext}\,dt = -mg\,dt &= m\,dv + u\,dm \\ -mg &= m\dot{v} + u\dot{m} \end{aligned} \tag{9.160}$$

Because the fuel burn rate is constant, let

$$\dot{m} = \frac{dm}{dt} = -\alpha, \qquad \alpha > 0 \tag{9.161}$$

and Equation 9.160 becomes

$$dv = \left(-g + \frac{\alpha}{m}u\right)dt$$

This equation, however, has three unknowns (v, m, t), so we use Equation 9.161 to eliminate time, giving

$$dv = \left(\frac{g}{\alpha} - \frac{u}{m}\right)dm \tag{9.162}$$

Assume the initial and final values of the velocity to be 0 and v respectively and of the mass m_0 and m respectively, so that

$$\int_0^v dv = \int_{m_0}^m \left(\frac{g}{\alpha} - \frac{u}{m}\right) dm$$

$$v = -\frac{g}{\alpha}(m_0 - m) + u \ln\left(\frac{m_0}{m}\right) \tag{9.163}$$

We can integrate Equation 9.161 to find the time:

$$\int_{m_0}^m dm = -\alpha \int_0^t dt$$

$$m_0 - m = \alpha t \tag{9.164}$$

Equation 9.163 becomes

$$v = -gt + u \ln\left(\frac{m_0}{m}\right) \tag{9.165}$$

We could continue with Equation 9.163 and integrate once more to determine the height of the rocket, but we leave that to Example 9.13. Such integrations are tedious, and the problem is more easily handled by computer methods. Even at burnout, the rocket will continue rising because it still has an upward velocity. Eventually, with the preceding assumptions, the gravitational force will stop the rocket (because we assumed a constant g not decreasing with height).

An interesting situation occurs if the exhaust velocity u is not sufficiently great to make v in Equation 9.165 positive. In this case, the rocket would remain on the ground. This situation occurs because of the limits of integration we assumed leading to Equation 9.163. We would need to burn off sufficient fuel before the rocket thrust would lift it off the ground. Of course, rockets are not designed this way; they are made to lift off as the rockets reach a full bum rate.

EXAMPLE 9.12

Consider the first stage of a Saturn brocket used for the Apollo moon program. The initial mass is 2.8 x 10^6 kg, and the mass of the first-stage fuel is 2.1 x 10^6 kg. Assume a mean thrust of 37 x 10^6 N. The exhaust velocity is 2600 m/s. Calculate the final speed of the first stage at burnout. Also calculate the vertical height at burnout.

Solution. From the thrust (Equation 9.157), we can determine the fuel burn rate:

$$\frac{dm}{dt} = \frac{\text{thrust}}{-u} = \frac{37 \times 10^6\,\text{N}}{-2600\ \text{m/s}} = -1.42 \times 10^4\,\text{kg/s}$$

The final rocket mass is (2.8 x l0^6 kg - 2.1 x l0^6 kg) or 0.7 x 10^6 kg. We can determine the rocket speed at burnout (v_b) using Equation 9.163.

$$v_b = -\frac{9.8\ \text{m/s}^2(2.1 \times 10^6\,\text{kg})}{1.42 \times 10^4\,\text{kg/s}} + (2600\ \text{m/s}) \ln\left[\frac{2.8 \times 10^6\,\text{kg}}{0.7 \times 10^6\,\text{kg}}\right]$$

$$v_b = 2.16 \times 10^3\,\text{m/s}$$

The time to burnout t_b,, from Equation 9.164, is

$$t_b = \frac{m_0 - m}{\alpha} = \frac{2.1 \times 10^6\,\text{kg}}{1.42 \times 10^4\,\text{kg/s}} = 148\ \text{s}$$

or about 2.5 min.

It can be shown that the height at burnout y_b is:

$$y_b = ut_b - \frac{1}{2}gt_b^2 - \frac{mu}{\alpha}\ln\left(\frac{m_0}{m}\right) \tag{9.166}$$

$$y_b = (2600\ \text{m/s})(148\ \text{s}) - \frac{1}{2}(9.8\ \text{m/s}^2)\cdot(148\ \text{s})^2$$

$$- \frac{(0.7\times 10^6\ \text{kg})\cdot(2600\ \text{m/s})}{1.42\times 10^4\ \text{kg/s}}\ln\left(\frac{2.8\times 10^6\ \text{kg}}{0.7\times 10^6\ \text{kg}}\right)$$

$$y_b = 9.98\times 10^4\ \text{m} \approx 100\ \text{km}$$

The actual height is only about two-thirds of this value.

EXAMPLE 9.13

A sounding rocket leaves Earth's surface under gravity, typically in a vertical direction and returns to Earth. The exhaust velocity is u, and the constant fuel burn rate is α . Let the initial mass be m_0 and the mass at fuel burnout be m_f. Calculate the altitude and speed of the rocket at fuel burnout in terms of u, α , mf, m0, and g.

Solution. We determine the time T at burnout from Equation 9.164, T = (m_0 - m_f)/ α. We integrate over the velocity, Equation 9.165, to find H_{bo}, the height at fuel burn out.

$$H_{bo} = \int_0^T v\,dt = \int_0^T\left[-gt + u\ln\left(\frac{m_0}{m}\right)\right]dt$$

We use Equation 9.161, dt= -(dm)/ α , for the last integral and integrate over dm.

$$H_{bo} = -g\int_0^T dt + \frac{u}{\alpha}\int_{m_0}^{m_f}\ln\left(\frac{m}{m_0}\right)dm$$

We integrate the last term using the definite integral,

$$\int ln\, x\, dx = x\, ln\, x - x$$

to obtain after collecting terms.

$$H_{bo} = -\frac{g(m_0 - m_f)^2}{2\alpha^2} + \frac{u}{\alpha}\left[m_f \ln\left(\frac{m_f}{m_0}\right) + m_0 - m_f\right] \tag{9.167}$$

If we insert the numbers from the last example, we find the same answer for the burnout height.

The speed at burnout can be determined directly from Equation 9.165.

$$v_{bo} = -gT + u \ln\left(\frac{m_0}{m_f}\right)$$

$$= -\frac{g(m_0 - m_f)}{\alpha} + u \ln\left(\frac{m_0}{m_f}\right) \tag{9.168}$$

CHAPTER 10 - Motion in a Noninertial Reference Frame

10.1 Introduction

The advantage of choosing an inertial reference frame to describe dynamic processes was made evident in the discussions in Chapters 2 and 7. It is always possible to express the equations of motion for a system in an inertial frame. But there are types of problems for which these equations would be extremely complex, and it becomes easier to treat the motion of the system in a noninertial frame of reference.

To describe, for example, the motion of a particle on or near the surface of Earth, it is tempting to do so by choosing a coordinate system fixed with respect to Earth. We know, however, that Earth undergoes a complicated motion, compounded of many different rotations (and hence accelerations) with respect to an inertial reference frame identified with the "fixed" stars. Earth's coordinate system is, therefore, a *noninertial* frame of reference; and, although the solutions to many problems can be obtained to the desired degree of accuracy by ignoring this distinction, many important effects result from the noninertial nature of the Earth coordinate system.

In fact, we have already studied noninertial systems when we studied ocean tides (Section 5.5). Tidal forces due to Earth-Moon and Sun-Earth orbits are observed on Earth's surface, which is a noninertial system. Space does not allow us further study in this chapter of this interesting subject, but quite reasonable accounts can be found elsewhere.

In analyzing the motion of rigid bodies in the following chapter, we also find it convenient to use noninertial reference frames and therefore make use of much of the development presented here.

10.2 Rotating Coordinate Systems

Let us consider two sets of coordinate axes. Let one set be the "fixed" or inertial axes, and let the other be an arbitrary set that may be in motion with respect to the inertial system. We designate these axes as the "fixed" and "rotating" axes, respectively. We use x'_i as coordinates in the fixed system and x_i as coordinates in the rotating system. If we choose some point P, as in Figure 10-1, we have

$$\mathbf{r}' = \mathbf{R} + \mathbf{r} \tag{10.1}$$

where **r**' is the radius vector of P in the fixed system and **r** is the radius vector of P in the rotating system. The vector **R** locates the origin of the rotating system in the fixed system.

We may always represent an arbitrary infinitesimal displacement by a pure rotation about some axis called the **instantaneous axis of rotation**. For example, the instantaneous motion of a disk rolling down an inclined plane can be described as a rotation about the point of contact between the disk and the plane. Therefore, if the x, system undergoes an infinitesimal rotation $\delta\boldsymbol{\theta}$, corresponding to some arbitrary infinitesimal displacement, the motion of P (which, for the moment, we consider to be at rest in the x_i system) can be described in terms of Equation 1.106 as

$$(d\mathbf{r})_{\text{fixed}} = d\boldsymbol{\theta} \times \mathbf{r} \tag{10.2}$$

where the designation "fixed" is explicitly included to indicate that the quantity d**r** is measured in the x'_i, or fixed, coordinate system. Dividing this equation by dt, the time interval during which the infinitesimal rotation takes place, we obtain the time of rate change of **r** as measured in the fixed coordinate system:

$$\left(\frac{d\mathbf{r}}{dt}\right)_{\text{fixed}} = \frac{d\boldsymbol{\theta}}{dt} \times \mathbf{r} \tag{10.3}$$

or, because the angular velocity of the rotation is

$$\boldsymbol{\omega} \equiv \frac{d\boldsymbol{\theta}}{dt} \tag{10.4}$$

We have

$$\left(\frac{d\mathbf{r}}{dt}\right)_{\text{fixed}} = \boldsymbol{\omega} \times \mathbf{r} \qquad (\text{for } P \text{ fixed in } x_i \text{ system}) \tag{10.5}$$

This same result was determined in Section 1.15.

If we allow the point P to have a velocity $(d\mathbf{r}/dt)_{\text{rotating}}$ with respect to the x_i system, this velocity must be added to to $\boldsymbol{\omega} \times \mathbf{r}$ to obtain the time rate of change of $\mathbf{r}$ in the fixed system:

$$\left(\frac{d\mathbf{r}}{dt}\right)_{\text{fixed}} = \left(\frac{d\mathbf{r}}{dt}\right)_{\text{rotating}} + \boldsymbol{\omega} \times \mathbf{r} \tag{10.6}$$

EXAMPLE 10.1

Consider a vector $\mathbf{r} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$ in the rotating system. Let the fixed and rotating systems have the same origin. Find $\dot{\mathbf{r}}'$ in the fixed system by direct differentiation if the angular velocity of the rotating system is $\boldsymbol{\omega}$ in the fixed system.

Solution, We begin by taking the time derivative directly

$$\begin{aligned}\left(\frac{d\mathbf{r}}{dt}\right)_{\text{fixed}} &= \frac{d}{dt}\left(\sum_i \mathbf{x}_i\,\mathbf{e}_i\right) \\ &= \sum_i (\dot{x}_i\mathbf{e}_i + x_i\dot{\mathbf{e}}_i)\end{aligned} \tag{10.7}$$

The first term is simply $\dot{\mathbf{r}}_r$ in the rotating system, but what are the $\dot{\mathbf{e}}_i$?

$$\dot{\mathbf{r}}_r = \left(\frac{d\mathbf{r}}{dt}\right)_{\text{rotating}}$$

$$\left(\frac{d\mathbf{r}}{dt}\right)_{\text{fixed}} = \dot{\mathbf{r}}_r + \sum_i x_i \dot{\mathbf{e}}_i \tag{10.8}$$

Look at Figure 10-2 and examine which components of ω_i tend to rotate $\mathbf{e}_1$.

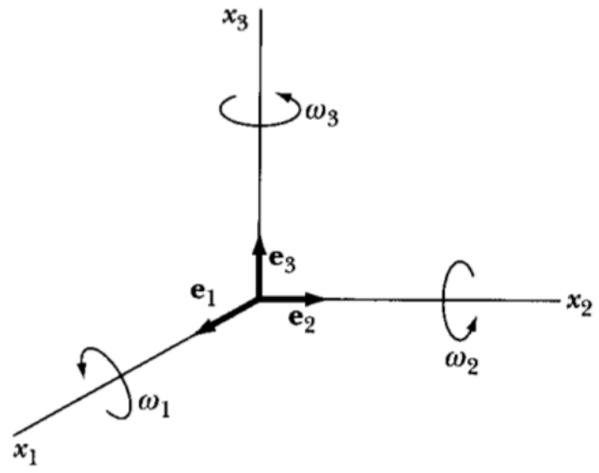


FIGURE 10-2 The angular velocity components ω_i rotate the system around the $\mathbf{e}_i$ axis, so that, for example, ω_3 tends to rotate $\mathbf{e}_1$ toward the $+\mathbf{e}_2$ direction.

We see that ω_2 tends to rotate $\mathbf{e}_1$ toward the $-\mathbf{e}_3$ direction and that ω_3 tends to rotate $\mathbf{e}_1$ toward the $+\mathbf{e}_2$ direction. We therefore have

$$\frac{d\mathbf{e}_1}{dt} = \omega_3 \mathbf{e}_2 - \omega_2 \mathbf{e}_3 \tag{10.9a}$$

Similarly, we have

$$\frac{d\mathbf{e}_2}{dt} = -\omega_3 \mathbf{e}_1 + \omega_1 \mathbf{e}_3 \tag{10.9b}$$

$$\frac{d\mathbf{e}_3}{dt} = \omega_2 \mathbf{e}_1 - \omega_1 \mathbf{e}_2 \tag{10.9c}$$

In each case, the direction of the time derivative of the unit vector must be perpendicular to the unit vector in order not to change its magnitude.

Equations 10.9a-c can be written

$$\dot{\mathbf{e}}_i = \boldsymbol{\omega} \times \mathbf{e}_i \tag{10.10}$$

and Equation 10.8 becomes

$$\begin{aligned}\left(\frac{d\mathbf{r}}{dt}\right)_{\text{fixed}} &= \dot{\mathbf{r}}_r + \sum_i \boldsymbol{\omega} \times x_i \mathbf{e}_i \\ &= \dot{\mathbf{r}}_r + \boldsymbol{\omega} \times \mathbf{r}\end{aligned} \tag{10.11}$$

which is the same result as Equation 10.6.

Although we choose the displacement vector **r** for the derivation of Equation 10.6, the validity of this expression is not limited to the vector **r**. In fact, for an arbitrary vector **Q**, we have

$$\boxed{\left(\frac{d\mathbf{Q}}{dt}\right)_{\text{fixed}} = \left(\frac{d\mathbf{Q}}{dt}\right)_{\text{rotating}} + \boldsymbol{\omega} \times \mathbf{Q}} \tag{10.12}$$

Equation 10.12 is an important result

We note, for example, that the angular acceleration $\dot{\boldsymbol{\omega}}$ is the same in both the fixed and rotating systems:

$$\left(\frac{d\boldsymbol{\omega}}{dt}\right)_{\text{fixed}} = \left(\frac{d\boldsymbol{\omega}}{dt}\right)_{\text{rotating}} + \boldsymbol{\omega} \times \boldsymbol{\omega} \equiv \dot{\boldsymbol{\omega}} \tag{10.13}$$

because $\boldsymbol{\omega} \times \boldsymbol{\omega}$ vanishes and $\dot{\boldsymbol{\omega}}$ designates the common value in the two systems.

Equation 10.12 may now be used to obtain the expressions for the velocity of the point P as measured in the fixed coordinate system. From Equation 10.1, we have

$$\left(\frac{d\mathbf{r}'}{dt}\right)_{\text{fixed}} = \left(\frac{d\mathbf{R}}{dt}\right)_{\text{fixed}} + \left(\frac{d\mathbf{r}}{dt}\right)_{\text{fixed}} \tag{10.14}$$

so that

$$\left(\frac{d\mathbf{r}'}{dt}\right)_{\text{fixed}} = \left(\frac{d\mathbf{R}}{dt}\right)_{\text{fixed}} + \left(\frac{d\mathbf{r}}{dt}\right)_{\text{rotating}} + \boldsymbol{\omega} \times \mathbf{r} \tag{10.15}$$

If we define

$$\mathbf{v}_f \equiv \dot{\mathbf{r}}_f \equiv \left(\frac{d\mathbf{r}'}{dt}\right)_{\text{fixed}} \tag{10.16a}$$

$$\mathbf{V} \equiv \dot{\mathbf{R}}_f \equiv \left(\frac{d\mathbf{R}}{dt}\right)_{\text{fixed}} \tag{10.16b}$$

$$\mathbf{v}_r \equiv \dot{\mathbf{r}}_r \equiv \left(\frac{d\mathbf{r}}{dt}\right)_{\text{rotating}} \tag{10.16c}$$

we may write

$$\boxed{\mathbf{v}_f = \mathbf{V} + \mathbf{v}_r + \boldsymbol{\omega} \times \mathbf{r}} \tag{10.17}$$

where

$\mathbf{v}_f$ = Velocity relative to the fixed axes

$\mathbf{V}$ = Linear velocity of the moving origin

$\mathbf{v}_r$ = Velocity relative to the rotating axes

$\boldsymbol{\omega}$ = Angular velocity of the rotating axes

$\boldsymbol{\omega} \times \mathbf{r}$ = Velocity due to the rotation of the moving axes

10.3 Centrifugal and Coriolis Forces

We have seen that Newton's equation **F** = m**a** is valid only in an inertial frame of reference. The expression for the force on a particle can therefore be obtained from

$$\mathbf{F} = m\mathbf{a}_f = m\left(\frac{d\mathbf{v}_f}{dt}\right)_{\text{fixed}} \tag{10.18}$$

where the differentiation must be carried out with respect to the fixed system.

Differentiating Equation 10.17, we have

$$\left(\frac{d\mathbf{v}_f}{dt}\right)_{\text{fixed}} = \left(\frac{d\mathbf{V}}{dt}\right)_{\text{fixed}} + \left(\frac{d\mathbf{v}_r}{dt}\right)_{\text{fixed}} + \dot{\boldsymbol{\omega}} \times \mathbf{r} + \boldsymbol{\omega} \times \left(\frac{d\mathbf{r}}{dt}\right)_{\text{fixed}} \tag{10.19}$$

We denote the first term by $\ddot{\mathbf{R}}_f$:

$$\ddot{\mathbf{R}}_f \equiv \left(\frac{d\mathbf{V}}{dt}\right)_{\text{fixed}} \tag{10.20}$$

The second term can be evaluated by substituting $\mathbf{v}_r$ for **Q** in Equation 10.12:

$$\begin{aligned}\left(\frac{d\mathbf{v}_r}{dt}\right)_{\text{fixed}} &= \left(\frac{d\mathbf{v}_r}{dt}\right)_{\text{rotating}} + \boldsymbol{\omega} \times \mathbf{v}_r \\ &= \mathbf{a}_r + \boldsymbol{\omega} \times \mathbf{v}_r\end{aligned} \tag{10.21}$$

where $\mathbf{a}_r$ is the acceleration in the rotating coordinate system. The last term in Equation 10.19 can be obtained directly from Equation 10.6:

$$\begin{aligned}\boldsymbol{\omega} \times \left(\frac{d\mathbf{r}}{dt}\right)_{\text{fixed}} &= \boldsymbol{\omega} \times \left(\frac{d\mathbf{r}}{dt}\right)_{\text{rotating}} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) \\ &= \boldsymbol{\omega} \times \mathbf{v}_r + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})\end{aligned} \tag{10.22}$$

Combining Equations 10.18-10.22, we obtain

$$\mathbf{F} = m\mathbf{a}_f = m\ddot{\mathbf{R}}_f + m\mathbf{a}_r + m\dot{\boldsymbol{\omega}} \times \mathbf{r} + m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) + 2m\boldsymbol{\omega} \times \mathbf{v}_r \tag{10.23}$$

To an observer in the rotating coordinate system, however, the effective force on a particle is given by

$$\mathbf{F}_{\text{eff}} \equiv m\mathbf{a}_r \tag{10.24}$$

$$= \mathbf{F} - m\ddot{\mathbf{R}}_f - m\dot{\boldsymbol{\omega}} \times \mathbf{r} - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) - 2m\boldsymbol{\omega} \times \mathbf{v}_r \tag{10.25}$$

The first term, $\mathbf{F}$, is the sum of the forces acting on the particle as measured in the fixed inertial system. The second ($-m\ddot{\mathbf{R}}_f$) and third ($-m\dot{\boldsymbol{\omega}} \times \mathbf{r}$) terms result because of the translational and angular acceleration, respectively, of the moving coordinate system relative to the fixed system.

The quantity $-m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})$ is the usual *centrifugal force* term and reduces to $m\omega^2 r$ for the case in which bi is normal to the radius vector. Note that the minus sign implies that the centrifugal force is directed *outward* from the center of rotation (Figure 10-3).

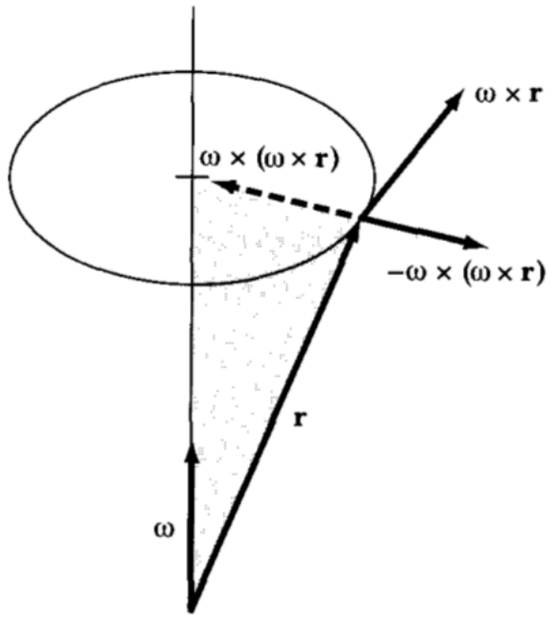


FIGURE 10-3 Diagram indicating that the vector $-\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})$ points outward, away from the axis of rotation along $\boldsymbol{\omega}$. The term $-m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})$ is the usual centrifugal force.

The last term in Equation 10.25 is a totally new quantity that arises from the motion of the particle in the rotating coordinate system. This term is called the **Coriolis force**.

Note that the Coriolis force does indeed arise from the motion of the particle, because the force is proportional to v, and hence vanishes if there is no motion.

Because we have used (on several occasions) the term *centrifugal force* and have now introduced the Coriolis force, we must now inquire about the physical meaning of these quantities. It is important to realize that the centrifugal and Coriolis forces are not forces in the usual sense of the word; they have been introduced in an artificial manner as a result of our arbitrary requirement that we be able to write an equation resembling Newton's equation that is at the same time valid in a noninertial reference frame; that is, the equation

$$\mathbf{F} = m\mathbf{a}_f$$

is valid only in an inertial frame. If, in a rotating reference frame, we wish to write (let $\ddot{\mathbf{R}}_f$ and $\dot{\boldsymbol{\omega}}$ be zero for simplicity)

$$\mathbf{F}_{\text{eff}} = m\mathbf{a}_r$$

then we can express such an equation in terms of the real force $m\mathbf{a}_f$ as

$$\mathbf{F}_{\text{eff}} = m\mathbf{a}_f + (\text{noninertial terms})$$

where the "noninertial terms" are identified as the centrifugal and Coriolis "forces." Thus, for example, if a body rotates about a fixed force center, the only real force on the body is the force of attraction toward the force center (and gives rise to the *centripetal* acceleration). An observer moving with the rotating body, however, measures this central force and also notes that the body does not fall toward the force center. To reconcile this result with the requirement that the net force on the body vanish, the observer must postulate an additional force - the centrifugal force. But the "requirement" is artificial; it arises solely from an attempt to extend the form of Newton's equation to a noninertial system, and this can be done only by introducing a fictitious "correction force." The same comments apply for the Coriolis force; this "force" arises when an attempt is made to describe motion relative to the rotating body.

Despite their artificiality, the concepts of centrifugal and Coriolis forces are useful. To describe the motion of a particle relative to a body rotating with respect to an inertial reference frame is a complicated matter. But the problem can be made relatively easy by the simple expedient of introducing the "noninertial forces," which then allows the use of an equation of motion resembling Newton's equation.

EXAMPLE 10.2

A student is performing measurements with a hockey puck on a large merry-go-round with a smooth (frictionless) horizontal, flat surface. The merry-go-round has a constant angular velocity ω and rotates counterclockwise as seen from above, (a) Find the effective force on the hockey puck after it is given a push, (b) Plot the path for various initial directions and velocities of the puck as observed by the person on the merry-go-round that pushes the puck.

Solution. The first three terms for $\mathbf{F}_{\text{eff}}$ in Equation 10.25 are zero, so the effective force as observed by the person on the merry-go-round is

$$\mathbf{F}_{\text{eff}} = -m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) - 2m\boldsymbol{\omega} \times \mathbf{v}_r \tag{10.26}$$

We have taken the frictional force to be zero. Remember that $\mathbf{v}_r$ is the velocity as measured by the observer on the rotating surface. The effective acceleration is

$$\mathbf{a}_{\text{eff}} = \frac{\mathbf{F}_{\text{eff}}}{m} = -\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) - 2\boldsymbol{\omega} \times \mathbf{v}_r \tag{10.27}$$

The velocity and position are given by integration, in turn, of the acceleration.

$$\mathbf{v}_{\text{eff}} = \int \mathbf{a}_{\text{eff}}\, dt \tag{10.28a}$$

$$\mathbf{r}_{\text{eff}} = \int \mathbf{v}_{\text{eff}}\, dt \tag{10.28b}$$

We put the origin of our rotating coordinate system at the center of the merry-go-round. We will need the initial positions and velocities of the puck to plot the motion. For this example, we let the radius of the merry-go-round be R and the velocities be in units of ωR. The initial position of the puck will always be at an (x, y) position of (—0.5R, 0).

We perform a numerical calculation to determine the motion and show the results for several directions and values of the initial velocity in Figure 10-4.

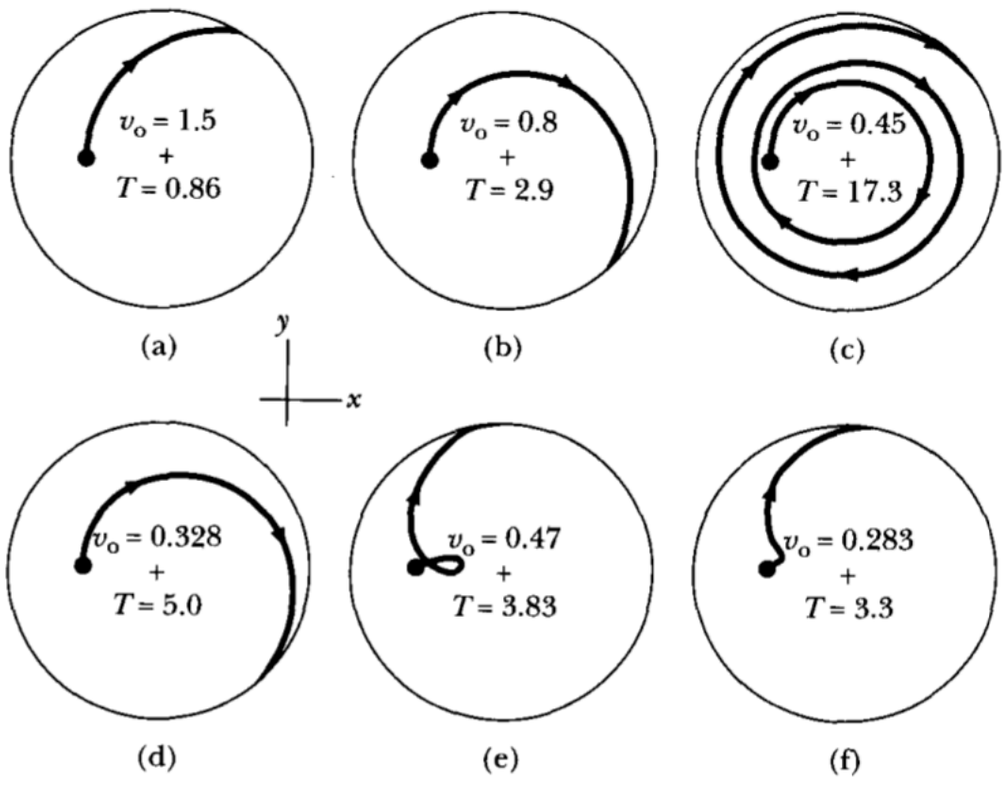


FIGURE 10-4 The motion of the hockey puck of Example 10.2 as observed in the rotating system for various initial directions and velocities v_0 at the times T noted. The angular velocity ω(1 rad/s) is out of the page.

For purposes of calculation, we let ω = 1 rad/s and R = 1 m, so the units of v_0 (initial speed) and T (time for puck to slide off the surface) shown in Figure 10-4 are in m/s and s, respectively. For parts (a)-(d), the initial velocity is in the +y-direction, and the initial speed decreases in each succeeding view. In (a), the puck slides off quickly. For (b) and (d), the puck slides off at similar positions, but note the differences in initial speeds as well as the time it takes the puck to reach the edge.

For a speed intermediate between these two speeds, as seen in (c), the puck may make several paths around the merry-go-round; at some speed, the puck must stay on. The last two views show the initial velocity at an angle of 45° to the x-axis. In (e), the puck loops around its path along the way to exiting the merry-go-round, and in (f), it changes direction rather abruptly.

The real challenge is to perform such experiments to compare the actual paths in the fixed and rotating coordinate systems with the computer calculations.

10.4 Motion Relative to the Earth

The motion of Earth with respect to an inertial reference frame is dominated by Earth's rotation about its own axis. The effects of the other motions (e.g., the revolution about the Sun and the motion of the solar system with respect to the local galaxy) are small by comparison. If we place the fixed inertial frame x'y'z' at the center of Earth and the moving reference frame xyz on the surface of Earth, we can describe the motion of a moving object close to the surface of Earth as shown in Figure 10-5.

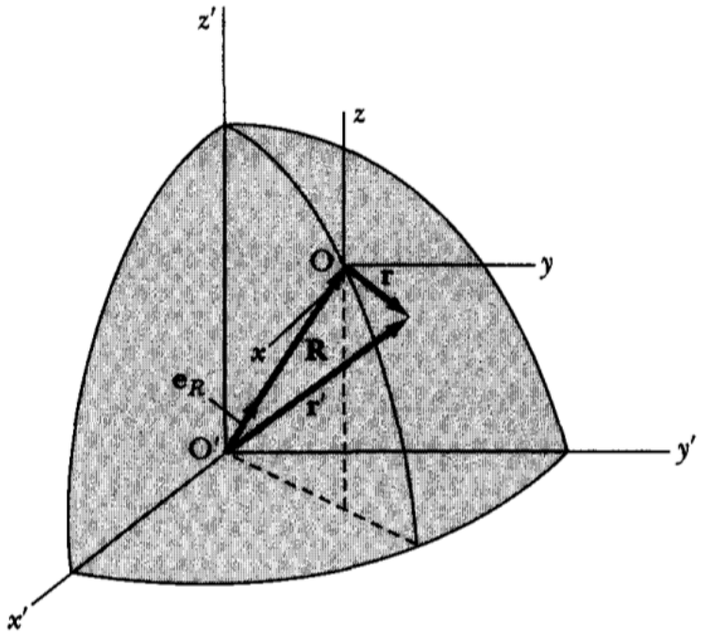


FIGURE 10-5 In order to study the motion of an object near Earth's surface, we place a fixed inertial frame $x'y'z'$ at the center of Earth and the moving frame xyz on Earth's surface.

We then apply Equation 10.25 to the dynamical motion. We denote the forces as measured in the fixed inertial system as $\mathbf{F} = \mathbf{S} + m\mathbf{g}_0$, where $\mathbf{S}$ represents the sum of the external forces (e.g., impulse, electromagnetic, friction) other than gravitation, and $m\mathbf{g}_0$ represents the gravitational attraction to Earth. In this case, $\mathbf{g}_0$ represents Earth's gravitational field vector (Equation 5.3),

$$\mathbf{g}_0 = -G\frac{M_E}{R^2}\mathbf{e}_R \tag{10.29}$$

where M_E is the mass of Earth, R is the radius of Earth, and the unit vector $\mathbf{e}_R$ is a unit vector along the direction of $\mathbf{R}$ in Figure 10-5. We are assuming Earth is spherical and isotropic and that $\mathbf{R}$ originates from the center of Earth. The acceleration of gravity varies over the surface of Earth due to Earth's oblateness, density nonuniformities, and altitude. We choose, at the present time, not to add this complexity to motion relative to Earth, but we have previously pointed out that effects such as these can be considered in due course by performing computer calculations.

The effective force $\mathbf{F}_{\text{eff}}$ as measured in the moving system placed on the surface of Earth becomes, from Equation 10.25,

$$\mathbf{F}_{\text{eff}} = \mathbf{S} + m\mathbf{g}_0 - m\ddot{\mathbf{R}}_f - m\dot{\boldsymbol{\omega}} \times \mathbf{r} - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) - 2m\boldsymbol{\omega} \times \mathbf{v}_r \tag{10.30}$$

We let Earth's angular velocity $\boldsymbol{\omega}$ be along the inertial system's z'-direction ($\mathbf{e}'z$). The value of ω is 7.3×10^{-5} rad/s, which is a relatively slow rotation, but it is 365 times greater than the rotation frequency of Earth about the Sun. The value of $\boldsymbol{\omega}$ is practically constant in time, and the term $\dot{\boldsymbol{\omega}} \times \mathbf{r}$ will be neglected.

According to Equation 10.12, we have for the third term above,

$$\ddot{\mathbf{R}}_f = \boldsymbol{\omega} \times \dot{\mathbf{R}}_f$$

$$\ddot{\mathbf{R}}_f = \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{R}) \tag{10.31}$$

Equation 10.30 now becomes

$$\mathbf{F}_{\text{eff}} = \mathbf{S} + m\mathbf{g}_0 - m\boldsymbol{\omega} \times [\boldsymbol{\omega} \times (\mathbf{r} + \mathbf{R})] - 2m\boldsymbol{\omega} \times \mathbf{v}_r \tag{10.32}$$

The second and third terms (divided by m) are what we experience (and measure) on the surface of Earth as the effective **g**, and we will henceforth denote it as **g**. Its value is

$$\mathbf{g} = \mathbf{g}_0 - \boldsymbol{\omega} \times [\boldsymbol{\omega} \times (\mathbf{r} + \mathbf{R})] \tag{10.33}$$

The second term of Equation 10.33 is the centrifugal force. Because we are limiting our present consideration to motion near the surface of Earth, we have $r \ll R$, and the to $\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{R})$ term totally dominates the centrifugal force. For situations far away from the surface of Earth, we would have to consider both the variation of g with altitude as well as the to $\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r})$ term. The centrifugal force is responsible for the oblateness of Earth. Earth is not really a solid spheroid; it is more like a strongly viscous liquid with a solid crust. Because of Earth's rotation. Earth has deformed so that its equatorial radius is 21.4 km greater than its polar radius, and the acceleration of gravity is 0.052 m/s^2 greater at the poles than at the equator. The surface of calm ocean water is perpendicular to **g**, not $\mathbf{g}_0$ and on the average, the plane of Earth's surface is also perpendicular to **g**.

We rewrite Equation 10.32 in simpler terms as

$$\boxed{\mathbf{F}_{\text{eff}} = \mathbf{S} + m\mathbf{g} - 2m\boldsymbol{\omega} \times \mathbf{v}_r} \tag{10.34}$$

It is this equation that we will use to discuss the motion of objects close to the surface of Earth.

But first, let's return to the effective **g** of Equation 10.33. The period of a pendulum determines the magnitude of **g**, and the direction of a plumb bob in equilibrium determines the direction of **g**. The value of $\omega^2 R$ is 0.034 m/s^2, and this is a significant enough amount (0.35%) of the magnitude of **g** to be considered. We determined the direction of the centrifugal term $\boldsymbol{\omega} \times [\boldsymbol{\omega} \times (\mathbf{r} + \mathbf{R})]$ in Figure 10-3 (where the **r** is our **r**' of Figure 10-5).

The direction of the centrifugal term ($\boldsymbol{\omega} \times [\boldsymbol{\omega} \times (\mathbf{r} + \mathbf{R})]$) is outward from the axis of the rotating Earth. The direction of a plumb bob will include the centrifugal term. Because of this fact, the direction of **g** at a given point is in general slightly different from the true vertical (defined as the direction of the line connecting the point with the center of Earth; see Problem 10-12). The situation is represented schematically (with considerable exaggeration) in Figure 10-6.

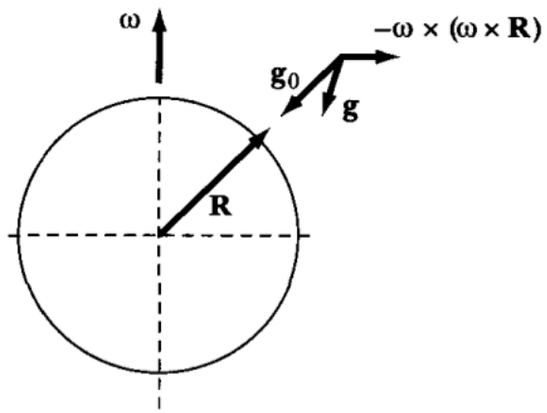


FIGURE 10-6 Near Earth's surface the terms $\mathbf{g}_0$ (Earth's gravitational field vector) and $-\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{R})$ (main centrifugal term) make up the effective **g** (other smaller terms have been neglected). The effect of the centrifugal term on **g** is exaggerated here.

Coriolis Force Effects

The angular velocity vector $\boldsymbol{\omega}$, which represents Earth's rotation about its axis, is directed in a northerly direction. Therefore, in the Northern Hemisphere, $\boldsymbol{\omega}$ has a component ω_z directed *outward* along the local vertical. If a particle is projected in a horizontal plane (in the local coordinate system at the surface of Earth) with a velocity $\mathbf{v}_r$, then the Coriolis force $-2m\boldsymbol{\omega} \times \mathbf{v}_r$ has a component in the plane of magnitude $2m\omega_z v_r$ directed toward the right of the particle's motion (see Figure 10-7), and a deflection from the original direction of motion results.

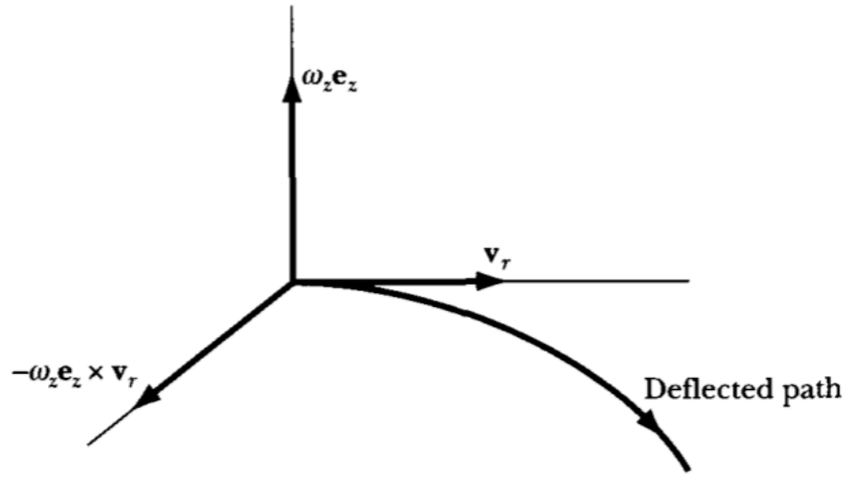


FIGURE 10-7 In the Northern Hemisphere, a particle projected in a horizontal plane will be directed toward the right of the particle's motion. In the Southern Hemisphere, the direction will be to the left.

Because the magnitude of the horizontal component of the Coriolis force is proportional to the vertical component of ω, the portion of the Coriolis force producing deflections depends on the latitude, being a maximum at the North Pole and zero at the equator. In the Southern Hemisphere, the component ω_z is directed inward along the local vertical, and hence all deflections are in the opposite sense from those in the Northern Hemisphere.

During the naval engagement near the Falkland Islands early in World War I, the British gunners were surprised to see their accurately aimed salvos falling 100 yards to the left of the German ships. The designers of the sighting mechanisms were well aware of the Coriolis deflection and had carefully taken this into account, but they apparently were under the impression that all sea battles took place near 50°N latitude and never near 50°S latitude. The British shots, therefore, fell at a distance from the targets equal to twice the Coriolis deflection.

Perhaps the most noticeable effect of the Coriolis force is that on the air masses. As air flows from high-pressure regions to low pressure, the Coriolis force deflects the air toward the right in the Northern Hemisphere, producing cyclonic motion (Figure 10-8).

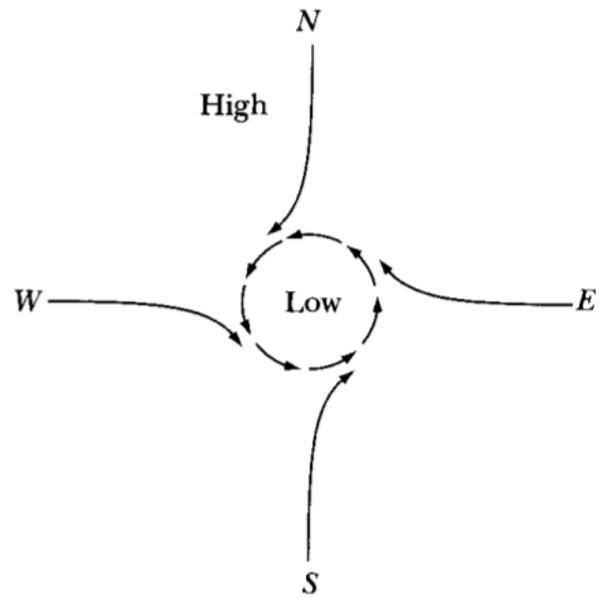


FIGURE 10-8 The Coriolis force deflects air in the Northern Hemisphere to the right producing cyclonic motion.

The air rotates with high pressure on the right and low pressure on the left. The high pressure prevents the Coriolis force from deflecting the air masses farther to the right, resulting in a counterclockwise flow of air In the temperate regions, the airflow does not tend to be along the pressure gradients, but rather along the pressure isobars due to the Coriolis force and the associated centrifugal force of the rotation.

Near the equatorial regions, the sun heating Earth's surface causes hot surface air to rise. In the Northern Hemisphere, this results in cooler air moving in a southerly direction toward the equator. The Coriolis force deflects this moving air to the right, resulting in the trade winds, which provide a breeze toward the southwest in the Northern Hemisphere and toward the northwest in the Southern Hemisphere. Note that this particular effect does not occur at the *equator* because of the directions of $\boldsymbol{\omega}$ and the air's surface $\mathbf{v}$.

The actual motion of air masses is much more complicated than the simple picture described here, but the qualitative features of cyclonic motion and the trade winds are correctly given by considering the effects of the Coriolis force. The motion of water in whirlpools is (at least in principle) a similar situation, but in actuality, other factors (various perturbations and residual angular momentum) dominate the Coriolis force, and whirlpools are found with both directions of flow.

Even under laboratory conditions, it is extremely difficult to isolate the Coriolis effect. (Reports of water in flush toilets and bathtubs circulating in opposite directions as cruise ships cross the equator are most likely highly exaggerated.)

EXAMPLE l0.5

Find the horizontal deflection from the plumb line caused by the Coriolis force acting on a particle falling freely in Earth's gravitational field from a height h above Earth's surface.

Solution. We use Equation 10.34 with the applied forces $\mathbf{S} = 0$. If we set $\mathbf{F}_{\text{eff}} = m\mathbf{a}_r$, we can solve for the acceleration of the particle in the rotating coordinate system fixed on Earth.

The acceleration due to gravity **g** is the effective one and is along the plumb line.

We choose a z-axis directed vertically outward (along -**g**) from the surface of Earth. With this definition of $\mathbf{e}_z$, we complete the construction of a right-hand coordinate system by specifying that $\mathbf{e}_x$ be in a southerly and $\mathbf{e}_y$ in an easterly direction, as in Figure 10-9. We make the approximation that the distance of fall is sufficiently small that g remains constant during the process.

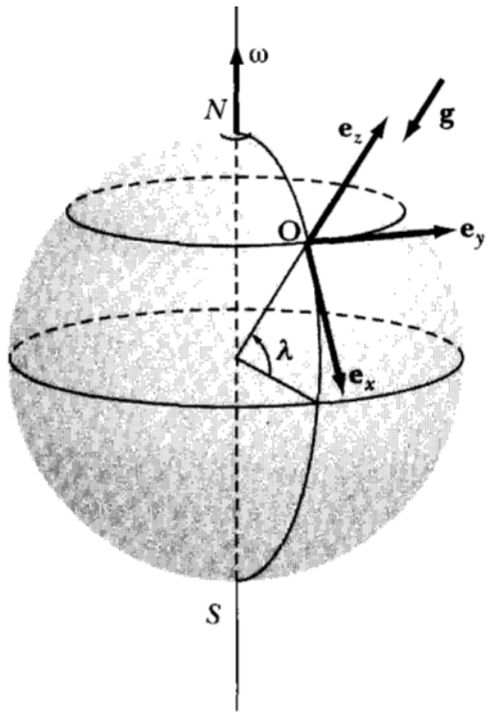


FIGURE 10-9 The coordinate system on Earth's surface for finding the horizontal deflection of a falling particle from the plumb line caused by the Coriolis force. The vector $\mathbf{e}_x$ is in the southerly direction, and $\mathbf{e}_y$ is in the easterly direction.

Because we have chosen the origin O of the rotating coordinate system to lie in the Northern Hemisphere, we have

$$\omega_x = -\omega \cos \lambda$$

$$\omega_y = 0$$

$$\omega_z = \omega \sin \lambda$$

Although the Coriolis force produces small velocity components in the $\mathbf{e}_y$, and $\mathbf{e}_x$ directions, we can certainly neglect $\dot{x}$ and $\dot{y}$ compared with $\dot{z}$, the vertical velocity. Then, approximately,

$$\dot{x} \cong 0$$

$$\dot{y} \cong 0$$

$$\dot{z} \cong -gt$$

where we obtain $\dot{z}$ by considering a fall from rest. Therefore, we have

$$\boldsymbol{\omega} \times \mathbf{v}_r \cong \begin{vmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \\ -\omega \cos \lambda & 0 & \omega \sin \lambda \\ 0 & 0 & -gt \end{vmatrix}$$

$$\cong -(\omega g t \cos \lambda)\mathbf{e}_y$$

The components of $\mathbf{g}$ are

$$g_x = 0$$

$$g_y = 0$$

$$g_z = -g$$

so the equations for the components of $\mathbf{a}_r$ (neglecting terms in ω^2) become

$$(\mathbf{a}_r)_x = \ddot{x} \cong 0$$

$$(\mathbf{a}_r)_y = \ddot{y} \cong 2\omega g t \cos\lambda$$

$$(\mathbf{a}_r)_z = \ddot{z} \cong -g$$

Thus, the effect of the Coriolis force is to produce an acceleration in the $\mathbf{e}_y$, or easterly, direction. Integrating $\ddot{y}$ twice, we have

$$y(t) \cong \frac{1}{3}\omega g t^3 \cos\lambda$$

where y = 0 and $\dot{y} = 0$ at t = 0. The integration of $\dot{z}$ yields the familiar result for the distance of fall,

$$z(t) \cong z(0) - \frac{1}{2}gt^2$$

and the time of fall from a height h = z(0) is given by

$$t \cong \sqrt{2h/g}$$

Hence the result for the eastward deflection d of a particle dropped from rest at a height h and at a northern latitude λ is

$$d \cong \frac{1}{3}\omega \cos\lambda \sqrt{\frac{8h^3}{g}} \tag{10.35}$$

An object dropped from a height of 100 m at latitude 45° is deflected approximately 1.55 cm (neglecting the effects of air resistance).

EXAMPLE 10.4

To demonstrate the power of the Coriolis method for obtaining the equations of motion in a noninertial reference frame, rework the last example but use only the formalism previously developed - the theory of central-force motion.

Solution. If we release a particle of small mass from a tower of height h above Earth's surface, the path the particle describes is a conic section - an ellipse with $\varepsilon \simeq 1$ and with one focus very close to Earth's center. If r is Earth's radius and λ the (northern) latitude, then at the moment of release, the particle has a horizontal velocity in the eastward direction:

$$v_{\text{hor}} = r\omega \cos \lambda = (R + h)\omega \cos \lambda$$

and the angular momentum about the polar axis is

$$l = mrv_{\text{hor}} = m(R + h)^2\omega \cos \lambda \tag{10.36}$$

The equation of the path is

$$\frac{\alpha}{r} = 1 - \varepsilon \cos \theta \tag{10.37}$$

if we measure θ from the initial position of the particle (see Figure 10-10).

Notice that there is a change of sign between Equation 10.37 and Equation 8.41 due to the different origins for θ in the two cases.

At t = 0, we have

$$\frac{\alpha}{R + h} = 1 - \varepsilon$$

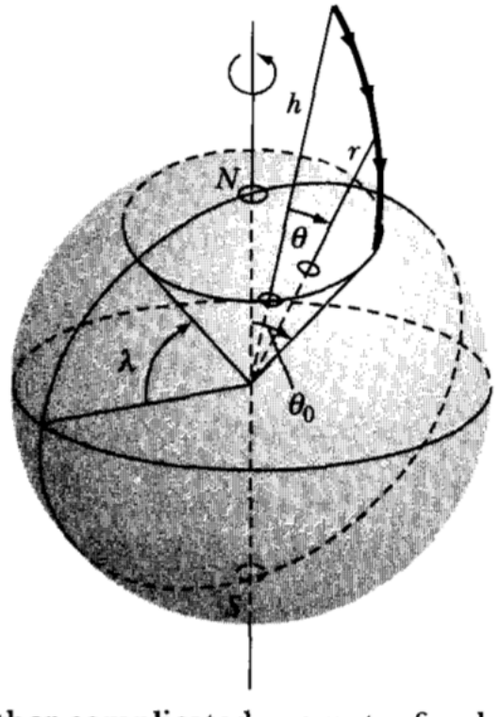


FIGURE 10-10 The rather complicated geometry for describing the motion of a falling particle in a noninertial system using central-force motion.

so Equation 10.37 can be written as

$$r = \frac{(1 - \varepsilon)(R + h)}{1 - \varepsilon \cos \theta} \tag{10.38}$$

From Equation 8.12 for the areal velocity, we can write

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

Thus, the time t required to describe an angle θ is

$$t = \frac{m}{l} \int_0^\theta r^2 d\theta$$

Substituting into this expression the value of l from Equation 10.36 and r from Equation 10.38, we find

$$t = \frac{1}{\omega \cos \lambda} \int_0^\theta \left(\frac{1 - \varepsilon}{1 - \varepsilon \cos \theta} \right)^2 d\theta \tag{10.39}$$

If we let $\theta = \theta_0$ when the particle has reached Earth's surface (r = R), then Equation 10.38 becomes

$$\frac{R}{R + h} = \frac{1 - \varepsilon}{1 - \varepsilon \cos \theta_0}$$

or, inverting,

$$\begin{aligned} 1 + \frac{h}{R} &= \frac{1 - \varepsilon \cos \theta_0}{1 - \varepsilon} \\ &= \frac{1 - \varepsilon[1 - 2 \sin^2(\theta_0/2)]}{1 - \varepsilon} \\ &= 1 + \frac{2\varepsilon}{1 - \varepsilon} \sin^2 \frac{\theta_0}{2} \end{aligned} \tag{10.40}$$

from which we have

$$\frac{h}{R} = \frac{2\varepsilon}{1-\varepsilon}\sin^2\frac{\theta_0}{2}$$

Because the path described by the particle is almost vertical, little change occurs in the angle θ between the position of release and the point at which the particle reaches the surface of Earth; θ_0 is therefore small and $\sin(\theta_0/2)$ can be approximated by its argument:

$$\frac{h}{R} \cong \frac{\varepsilon\theta_0^2}{2(1-\varepsilon)} \tag{10.41}$$

If we expand the integrand in Equation 10.39 by the same method used to obtain Equation 10.40, we find

$$t = \frac{1}{\omega\cos\lambda}\int_0^{\theta}\frac{d\theta}{\left\{1+[2\varepsilon/(1-\varepsilon)]\sin^2(\theta/2)\right\}^2}$$

and because θ is small, we have

$$t \cong \frac{1}{\omega\cos\lambda}\int_0^{\theta}\frac{d\theta}{[1+\varepsilon\theta^2/2(1-\varepsilon)]^2}$$

Substituting for $\varepsilon/2(1-\varepsilon)$ from Equation 10.41 and writing $t(\theta=\theta_0)=T$ for the total time of fall, we obtain

$$\begin{aligned} T &\cong \frac{1}{\omega\cos\lambda}\int_0^{\theta_0}\frac{d\theta}{\left[1+(h\theta^2/R\theta_0^2)\right]^2} \\ &\cong \frac{1}{\omega\cos\lambda}\int_0^{\theta_0}\left(1-\frac{2h}{R\theta_0^2}\theta^2\right)d\theta \\ &= \frac{1}{\omega\cos\lambda}\left(1-\frac{2h}{3R}\right)\theta_0 \end{aligned}$$

Solving for θ_0 , we find

$$\theta_0 \cong \frac{\omega T \cos\lambda}{1 - 2h/3R} \cong \omega T \cos\lambda\left(1 + \frac{2h}{3R}\right)$$

During the time of fall T, Earth turns through an angle ωT, so the point on Earth directly beneath the initial position of the particle moves toward the east by an amount $R\omega T\cos\lambda$. During the same time, the particle is deflected toward the east by an amount $R\theta_0$. Thus, the net easterly deviation d is

$$d = R\theta_0 - R\omega T\cos\lambda$$

$$= \frac{2}{3}h\omega T\cos\lambda$$

and using $T \simeq \sqrt{2h/g}$ as in the preceding example, we have, finally,

$$d \cong \frac{1}{3}\omega\cos\lambda\sqrt{\frac{8h^3}{g}}$$

which is identical with the result obtained previously (Equation 10.35).

EXAMPLE 10.5

The effect of the Coriolis force on the motion of a pendulum produces a precession, or rotation with time of the plane of oscillation. Describe the motion of this system, called a *Foucault pendulum*.

Solution. To describe this effect, let us select a set of coordinate axes with origin at the equilibrium point of the pendulum and z-axis along the local vertical. We are interested only in the rotation of the plane of oscillation - that is, we wish to consider the motion of the pendulum bob in the x-y plane (the horizontal plane). We therefore limit the motion to oscillations of small amplitude, with the horizontal excursions small compared with the length of the pendulum. Under this condition, $\dot{z}$ is small compared with $\dot{x}$ and $\dot{y}$ and can be neglected.

The equation of motion is

$$\mathbf{a}_r = \mathbf{g} + \frac{\mathbf{T}}{m} - 2\boldsymbol{\omega} \times \mathbf{v}_r \tag{10.42}$$

where **T**/m is the acceleration produced by the force of tension **T** in the pendulum suspension (Figure 10-11).

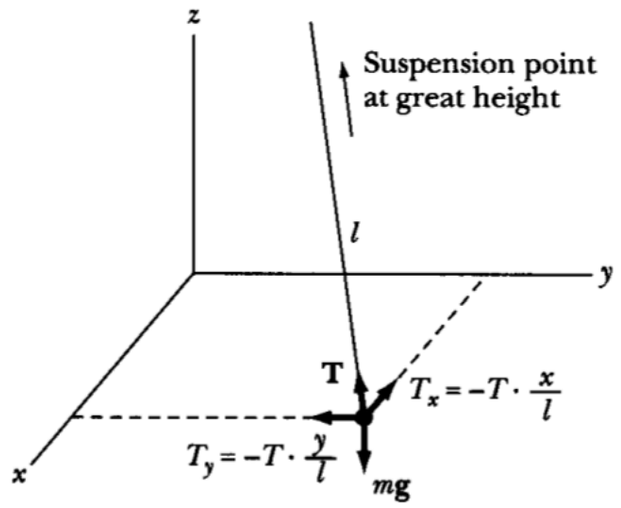


FIGURE 10-11 Geometry for the Foucault pendulum. The acceleration **g** vector is along the $-z$-direction, and the tension **T** is separated into x-, y-, and z-components.

We therefore have, approximately.

$$\left.\begin{aligned} T_x &= -T \cdot \frac{x}{l} \\ T_y &= -T \cdot \frac{y}{l} \\ T_z &\cong T \end{aligned}\right\} \tag{10.43}$$

As before,

$$g_x = 0$$

$$g_y = 0$$

$$g_z = -g$$

and

$$\omega_x = -\omega \cos \lambda$$
$$\omega_y = 0$$
$$\omega_z = \omega \sin \lambda$$

with

$$(\mathbf{v}_r)_x = \dot{x}$$
$$(\mathbf{v}_r)_y = \dot{y}$$
$$(\mathbf{v}_r)_z = \dot{z} \cong 0$$

Therefore,

$$\boldsymbol{\omega} \times \mathbf{v} \cong \begin{vmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \\ -\omega \cos \lambda & 0 & \omega \sin \lambda \\ \dot{x} & \dot{y} & 0 \end{vmatrix}$$

so that

$$\left.\begin{aligned} (\boldsymbol{\omega} \times \mathbf{v}_r)_x &\cong -\dot{y}\, \omega \sin \lambda \\ (\boldsymbol{\omega} \times \mathbf{v}_r)_y &\cong \dot{x}\, \omega \sin \lambda \\ (\boldsymbol{\omega} \times \mathbf{v}_r)_z &\cong -\dot{y}\, \omega \cos \lambda \end{aligned}\right\} \tag{10.44}$$

Thus, the equations of interest are

$$\left.\begin{aligned} (\mathbf{a}_r)_x = \ddot{x} &\cong -\frac{T}{m} \cdot \frac{x}{l} + 2\dot{y}\omega \sin \lambda \\ (\mathbf{a}_r)_y \cong \ddot{y} &\cong -\frac{T}{m} \cdot \frac{y}{l} - 2\dot{x}\omega \sin \lambda \end{aligned}\right\} \tag{10.45}$$

For small displacements, $T \simeq mg$. Defining $\alpha^2 = T/ml \simeq g/l$, and writing $\omega_z = \omega \sin \lambda$, we have

$$\left.\begin{aligned} \ddot{x} + \alpha^2 x &\cong 2\omega_z \dot{y} \\ \ddot{y} + \alpha^2 y &\cong -2\omega_z \dot{x} \end{aligned}\right\} \tag{10.46}$$

We note that the equation for $\ddot{x}$ contains a term in $\dot{y}$ and that the equation for $\ddot{y}$ contains a term in $\dot{x}$. Such equations are called **coupled equations**. A solution for this pair of coupled equations can be effected by adding the first of the above equations to i times the second:

$$(\ddot{x} + i\ddot{y}) + \alpha^2(x + iy) \cong -2\omega_z(i\dot{x} - \dot{y}) = -2i\omega_z(\dot{x} + i\dot{y})$$

If we write

$$q \cong x + iy$$

we then have

$$\ddot{q} + 2i\omega_z\dot{q} + \alpha^2 q \cong 0$$

This equation is identical with the equation that describes damped oscillations (Equation 3.35), except that here the term corresponding to the damping factor is purely imaginary. The solution (see Equation 3.37) is

$$q(t) \cong \exp[-i\omega_z t]\left[A\exp\left(\sqrt{-\omega_z^2 - \alpha^2}\, t\right) + B\exp\left(-\sqrt{-\omega_z^2 - \alpha^2}\, t\right)\right] \tag{10.47}$$

If Earth were not rotating, so that $\omega_z = 0$, then the equation for q would become

$$\ddot{q}' + \alpha^2 q' \cong 0, \qquad \omega_z = 0$$

from which it is seen that α corresponds to the oscillation frequency of the pendulum. This frequency is clearly much greater than the angular frequency of Earth's rotation. Therefore, $\alpha \gg \omega_z$, and the equation for q(t) becomes

$$q(t) \cong e^{-i\omega_z t}(Ae^{i\alpha t} + Be^{-i\alpha t}) \tag{10.48}$$

We can interpret this equation more easily if we note that the equation for q' has the solution

$$q'(t) = x'(t) + iy'(t) = Ae^{i\alpha t} + Be^{-i\alpha t}$$

Thus,

$$q(t) = q'(t) \cdot e^{-iw_z t}$$

or

$$\begin{aligned} x(t) + iy(t) &= [(x'(t) + iy'(t)] \cdot e^{-i\omega_z t} \\ &= (x' + iy')(\cos \omega_z t - i \sin \omega_z t) \\ &= (x' \cos \omega_z t + y' \sin \omega_z t) + i(-x' \sin \omega_z t + y' \cos \omega_z t) \end{aligned}$$

Equating real and imaginary parts,

$$\left.\begin{aligned} x(t) &= x' \cos \omega_z t + y' \sin \omega_z t \\ y(t) &= -x' \sin \omega_z t + y' \cos \omega_z t \end{aligned}\right\}$$

We can write these equations in matrix form as

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} \cos \omega_z t & \sin \omega_z t \\ -\sin \omega_z t & \cos \omega_z t \end{pmatrix} \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix} \tag{10.49}$$

from which (x, y) may be obtained from (x', y') by the application of a rotation matrix of the familiar form

$$\boldsymbol{\lambda} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \tag{10.50}$$

Thus, the angle of rotation is $\theta = \omega_z t$, and the plane of oscillation of the pendulum therefore rotates with a frequency $\omega_z = \omega \sin \lambda$. The observation of this rotation gives a clear demonstration of the rotation of Earth.

CHAPTER 11 - Dynamics of Rigid Bodies

11.1 Introduction

We define a rigid body as a collection of particles whose relative distances are constrained to remain absolutely fixed. Such bodies do not exist in nature, because the ultimate component particles composing every body (the atoms) are always undergoing some relative motion like vibrations. This motion, however, is microscopic, and it therefore usually may be ignored when describing the macroscopic motion of the body. However, macroscopic displacement within the body (such as elastic deformations) can take place. For many bodies of interest, we can safely neglect the changes in size and shape caused by such deformations and obtain equations of motion valid to a high degree of accuracy. We use here the idealized concept of a rigid body as a collection of discrete particles or as a continuous distribution of matter interchangeably. The only change is the replacement of summations over particles by integrations over mass density distributions. The equations of motion are equally valid for either viewpoint.

We have studied rigid bodies in introductory physics and have already seen examples in this book of hoops and cylinders rolling down inclined planes. We also know how to find the center of mass of various rigid objects (Section 9.2). Such problems can be handled with concepts already presented including rotational inertia, angular velocity and momentum, and torque. We can use these techniques to solve many problems, such as some simple examples of planar motion in Section 11.2. When we allow complete three-dimensional motion, the mathematical complexity considerably escalates. The classic example, of course, is of the cat, who invariably lands on its feet after being dropped (under a carefully controlled experimental situation) with its feet initially pointing upwards

We have learned to describe the motion of a body by the sum of two independent motions - a linear translation of some point of the body plus a rotation about that point.

If the point is chosen to be the center of mass of the body, then such a separation of the motion into two parts allows the use of the development in Chapter 9, which indicates that the angular momentum (see Equation 9.39) can be separated into portions relating to the motion of the center of mass and to the motion *around* the center of mass.

This is an example of Chasles' Theorem, which is even more general than the above statement - it says that the line of translation and the axis of rotation can be made to coincide.

It is the general rotation that increases the complexity. We will find it useful to have two coordinate systems, one an inertial coordinate system (fixed) and the other a coordinate system fixed with respect to the body. Six quantities must be specified to denote the position of the body. We normally use three coordinates to describe the position of the center of mass (which can often conveniently be made to coincide with the origin of the body coordinate system) and three independent angles that give the orientation of the body coordinate system with respect to the fixed (or inertial) system. In this chapter, we use the designation body system in place of the term rotating system used in the preceding chapter. The term fixed system will be retained. The three independent angles are normally taken to be the **Eulerian angles**, described in Section 11.8.

Unfortunately, the mathematical level increases in this chapter We will find it prudent to introduce tensor and matrix algebra in order to describe the complete motion of simple looking dynamical systems like rotating tops (either free or in a gravitational field), dumbbells, gyroscopes, flywheels, and automobile wheels out-of-balance. We will use the dumbbell, because of its simplicity, as our system of interest as we introduce the needed mathematics.

11.2 Simple Planar Motion

We have already solved the problem of a disk rolling down an inclined plane (see Examples 6.5 and 7.9, and Figure 6-7). Several end-of-chapter problems in Chapter 7 concerned simple rigid bodies. We discussed center of mass in Section 9.2, angular momentum of a system of particles in Section 9.4, and the energy of the system in Section 9.5.

We restrict ourselves in this section to the motion of a rigid body in a plane and present examples as a review of our introductory physics.

Example 11.1

A string attached to the ceiling is wrapped around a homogenous cylinder of mass M and radius R (see Figure 11-1).

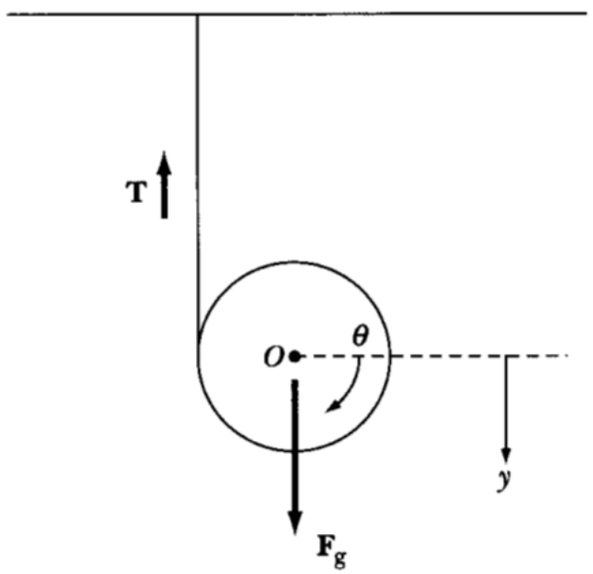


FIGURE 11-1 Example 11-1. A string attached to the ceiling is wrapped around a cylinder. The cylinder is released from rest.

At time t= 0, the cylinder is dropped from rest and rotates as the string unwinds. Find the tension T in the string, the linear and angular accelerations of the cylinder, and the angular velocity about the cylinder's center.

Solution. The center of mass moves due to the sum of the forces, which are all in the vertical direction. We let y point downward.

$$M\ddot{y} = F_g - T = Mg - T \tag{11.1}$$

where the center of mass acceleration is $\ddot{y}$, and we have used $F_g = Mg$. The rotation about the cylinder's center of mass at O is due to the tension T.

$$\tau = RT = I\ddot{\theta} \tag{11.2}$$

where τ is the torque about O, and I is the rotational inertia of the cylinder ($MR^2/2$). We let y = 0 and θ = 0 at t = 0 when the cylinder is released.

Then we have $y = R\theta$, $\dot{y} = V = R\dot{\theta}$, and $\ddot{y} = R\ddot{\theta}$. We can combine these relations with Equations 11.1 and 11.2 to determine the acceleration.

$$\ddot{y} = g - \frac{T}{M} = g - \frac{I\ddot{\theta}}{MR} = g - \frac{MR^2\ddot{y}}{2MR^2} = g - \frac{\ddot{y}}{2}$$

which gives $\ddot{y} = 2g/3$ for the acceleration, and the angular acceleration, $\alpha = \ddot{\theta} = \ddot{y}/R = 2g/3R$

The tension T is then found from Equation 11.2 to be

$$T = \frac{I\ddot{\theta}}{R} = \frac{MR^2\ddot{y}}{2R^2} = \frac{M}{2}\frac{2g}{3} = Mg/3 \tag{11.3}$$

The angular velocity is $\omega = \dot{\theta} = V/R$. We integrate $\ddot{y}$ to obtain $V = \dot{y} = 2gt/3$ and $\omega = \dot{y} = 2gt/3R$

EXAMPLE 11.2

A *physical* or *compound* pendulum is a rigid body that oscillates due to its own weight about a horizontal axis that does not pass through the center of mass of the body (Figure 11-2).

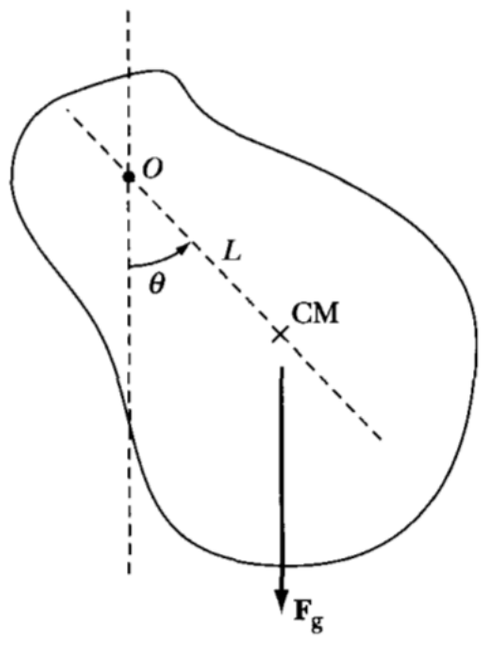


FIGURE 11-2 Example 11-2. The physical or compound pendulum. The body rotates about an axis passing through *O*. The body rotates due to the gravitational force acting at the center of mass.

For small oscillations, find the frequency and period of oscillation if the mass of the body is M and the *radius of gyration* is k.

Solution. We use the Lagrangian method to solve this example, although we could just as easily solve for the torque to find the equation of motion. The rotation axis passes through the point O of the body. The radius of gyration is defined such that the rotational inertia I about the given axis of rotation (O in this case) is given by $I = Mk^2$.

The kinetic energy of rotation and the potential energy is

$$T = \frac{1}{2} I\dot{\theta}^2$$

$$U = -MgL\cos\theta = -MgL\left(1 - \frac{\theta^2}{2}\right)$$

where we have defined the zero of the potential energy to be at point O and have used the small angle approximation for $\cos\theta$. We find the Lagrangian function and take the appropriate derivatives to form the Lagrange equation of motion. The generalized coordinate is clearly θ .

$$L = T - U = \frac{1}{2} I\dot{\theta}^2 + MgL\left(1 - \frac{\theta^2}{2}\right)$$

$$\frac{\partial L}{\partial \theta} = -MgL\theta$$

$$\frac{\partial L}{\partial \dot{\theta}} = I\dot{\theta}$$

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} = I\ddot{\theta}$$

The Lagrange equation of motion is

$$\ddot{\theta} + \frac{MgL}{I}\theta = 0$$

We have seen this equation several times, and the angular frequency is given by $\omega^2 = MgL/I$. From this we find the frequency ν and period T,

$$\nu = \frac{\omega}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{MgL}{I}} = \frac{1}{2\pi}\sqrt{\frac{MgL}{Mk^2}} = \frac{1}{2\pi}\sqrt{\frac{gL}{k^2}}$$

$$T = \frac{1}{\nu} = 2\pi\sqrt{\frac{k^2}{gL}}$$

Now that we have briefly reviewed our previous study of rigid body motion, let's proceed to the more general cases. For this we will need the inertia tensor.

11.3 Inertia Tensor

We now direct our attention to a rigid body composed of n particles of masses m_α , $\alpha = 1, 2, 3, \ldots, n$. If the body rotates with an instantaneous angular velocity $\boldsymbol{\omega}$ about some point fixed with respect to the body coordinate system and if this point moves with an instantaneous linear velocity **V** with respect to the fixed coordinate system, then the instantaneous velocity of the α^{th} particle in the fixed system can be obtained by using Equation 10.17. But we are now considering a rigid body, so

$$\mathbf{v}_r = \left(\frac{d\mathbf{r}}{dt}\right)_{\text{rotating}} \equiv 0$$

Therefore,

$$\boxed{\mathbf{v}_\alpha = \mathbf{V} + \boldsymbol{\omega} \times \mathbf{r}_\alpha} \qquad (11.4)$$

where the subscript f, denoting the fixed coordinate system, has been deleted from the velocity $\mathbf{v}_\alpha$ it now being understood that all velocities are measured in the fixed system.

All velocities with respect to the rotating or body system now vanish because the body is rigid.

Because the kinetic energy of the α^{th} particle is given by

$$T_\alpha = \frac{1}{2} m_\alpha v_\alpha^2$$

we have, for the total kinetic energy.

$$T = \frac{1}{2} \sum_\alpha m_\alpha (\mathbf{V} + \boldsymbol{\omega} \times \mathbf{r}_\alpha)^2$$

Expanding the squared term, we find

$$T = \frac{1}{2} \sum_\alpha m_\alpha V^2 + \sum_\alpha m_\alpha \mathbf{V} \cdot \boldsymbol{\omega} \times \mathbf{r}_\alpha + \frac{1}{2} \sum_\alpha m_\alpha (\boldsymbol{\omega} \times \mathbf{r}_\alpha)^2 \tag{11.5}$$

This is a general expression for the kinetic energy and is valid for any choice of the origin from which the vectors $\mathbf{r}_\alpha$ are measured. But if we make the origin of the body coordinate system coincide with the center of mass of the object, a considerable simplification results. First, we note that in the second term on the right-hand side of this equation neither $\mathbf{V}$ nor $\boldsymbol{\omega}$ is characteristic of the α^{th} particle, and therefore, these quantities may be taken outside the summation:

$$\sum_\alpha m_\alpha \mathbf{V} \cdot \boldsymbol{\omega} \times \mathbf{r}_\alpha = \mathbf{V} \cdot \boldsymbol{\omega} \times \left(\sum_\alpha m_\alpha \mathbf{r}_\alpha \right)$$

But now the term

$$\sum_\alpha m_\alpha \mathbf{r}_\alpha = M\mathbf{R}$$

is the center-of-mass vector (see Equation 9.3), which vanishes in the body system because the vectors $\mathbf{r}_\alpha$ are measured from the center of mass. The kinetic energy can then be written as

$$T = T_{\text{trans}} + T_{\text{rot}}$$

where

$$T_{\text{trans}} = \frac{1}{2}\sum_{\alpha} m_\alpha V^2 = \frac{1}{2}MV^2 \tag{11.6a}$$

$$T_{\text{rot}} = \frac{1}{2}\sum_{\alpha} m_\alpha (\boldsymbol{\omega} \times \mathbf{r}_\alpha)^2 \tag{11.6b}$$

T_{trans} and T_{rot} designate the translational and rotational kinetic energies, respectively. Thus, the kinetic energy separates into two independent parts.

The rotational kinetic energy term can be evaluated by noting that

$$(\mathbf{A} \times \mathbf{B})^2 = (\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{A} \times \mathbf{B})$$

$$= A^2B^2 - (\mathbf{A} \cdot \mathbf{B})^2$$

Therefore,

$$T_{\text{rot}} = \frac{1}{2}\sum_{\alpha} m_\alpha \left[\omega^2 r_\alpha^2 - (\boldsymbol{\omega} \cdot \mathbf{r}_\alpha)^2\right] \tag{11.7}$$

We now express T_{rot} by using the components ω_i and $r_{\alpha,i}$ of the vectors $\boldsymbol{\omega}$ and $\mathbf{r}_\alpha$. We also note that $\mathbf{r}_\alpha = (x_{\alpha,1}, x_{\alpha,2}, x_{\alpha,3})$ in the body system, so we can write $r_{\alpha,i} = x_{\alpha,i}$. Thus,

$$T_{\text{rot}} = \frac{1}{2}\sum_{\alpha} m_\alpha \left[\left(\sum_i \omega_i^2\right)\left(\sum_k x_{\alpha,k}^2\right) - \left(\sum_i \omega_i x_{\alpha,i}\right)\left(\sum_j \omega_j x_{\alpha,j}\right)\right] \tag{11.8}$$

Now, we can write $\omega_i = \sum_j \omega_j \delta_{ij}$, so that

$$T_{\text{rot}} = \frac{1}{2}\sum_{\alpha}\sum_{i,j} m_\alpha \left[\omega_i \omega_j \delta_{ij}\left(\sum_k x_{\alpha,k}^2\right) - \omega_i \omega_j x_{\alpha,i} x_{\alpha,j}\right]$$

$$= \frac{1}{2}\sum_{i,j} \omega_i \omega_j \sum_{\alpha} m_\alpha \left(\delta_{ij}\sum_k x_{\alpha,k}^2 - x_{\alpha,i} x_{\alpha,j}\right) \tag{11.9}$$

If we define the ij[th] element of the sum over α to be I_{ij}

$$\boxed{I_{ij} \equiv \sum_{\alpha} m_{\alpha} \left(\delta_{ij} \sum_{k} x_{\alpha,k}^2 - x_{\alpha,i}\, x_{\alpha,j} \right)} \tag{11.10}$$

then we have

$$\boxed{T_{\text{rot}} = \frac{1}{2} \sum_{i,j} I_{ij} \omega_i \omega_j} \tag{11.11}$$

This equation in its most restricted form becomes

$$T_{\text{rot}} = \frac{1}{2} I \omega^2 \tag{11.12}$$

where I is the (scalar) rotational inertia (moment of inertia) about the axis of rotation. This equation will be recognized as the familiar expression for the rotational kinetic energy given in elementary treatments.

The nine terms I_{ij} constitute the elements of a quantity we designated by {**I**}. In form, {**I**} is similar to a 3 x 3 matrix. It is the proportionality factor between the rotational kinetic energy and the angular velocity and has the dimensions (mass) x (length)^2. Because {**I**} relates two quite different physical quantities, we expect that it is a member of a somewhat higher class of functions than has heretofore been encountered. Indeed, {**I**} is a tensor and is known as the inertia tensor. The true test of a tensor lies in its behavior under a coordinate transformation (see Section 11.7). Note, however, that T_{rot} can be calculated without regard to any of the special properties of tensors, by using Equation 11.9, which completely specifies the necessary operations.

The elements of {**I**} can be obtained directly from Equation 11.10. We write the elements in a 3 x 3 array for clarity:

$$\{\mathbf{I}\} = \begin{Bmatrix} \sum_\alpha m_\alpha (x_{\alpha,2}^2 + x_{\alpha,3}^2) & -\sum_\alpha m_\alpha x_{\alpha,1} x_{\alpha,2} & -\sum_\alpha m_\alpha x_{\alpha,1} x_{\alpha,3} \\ -\sum_\alpha m_\alpha x_{\alpha,2} x_{\alpha,1} & \sum_\alpha m_\alpha (x_{\alpha,1}^2 + x_{\alpha,3}^2) & -\sum_\alpha m_\alpha x_{\alpha,2} x_{\alpha,3} \\ -\sum_\alpha m_\alpha x_{\alpha,3} x_{\alpha,1} & -\sum_\alpha m_\alpha x_{\alpha,3} x_{\alpha,2} & \sum_\alpha m_\alpha (x_{\alpha,1}^2 + x_{\alpha,2}^2) \end{Bmatrix} \tag{11.13a}$$

Equation 11.10 is a compact way to write the inertia tensor components, but Equation 11.13a is an imposing equation. By using components {x^, ya, z^)
instead of (x„i, Xa_2, Xa.3) and letting 1^ = x^ + ^ + ^, Equation 11.13a can be
written as

$$\{\mathbf{I}\} = \begin{Bmatrix} \sum_\alpha m_\alpha (r_\alpha^2 - x_\alpha^2) & -\sum_\alpha m_\alpha x_\alpha y_\alpha & -\sum_\alpha m_\alpha x_\alpha z_\alpha \\ -\sum_\alpha m_\alpha y_\alpha x_\alpha & \sum_\alpha m_\alpha (r_\alpha^2 - y_\alpha^2) & -\sum_\alpha m_\alpha y_\alpha z_\alpha \\ -\sum_\alpha m_\alpha z_\alpha x_\alpha & -\sum_\alpha m_\alpha z_\alpha y_\alpha & \sum_\alpha m_\alpha (r_\alpha^2 - z_\alpha^2) \end{Bmatrix} \tag{11.13b}$$

which is less imposing and more recognizable. We continue, however, with the $x_{\alpha,i}$ notation because of its utility.

The diagonal elements, I_{11}, I_{22}, and I_{33}, are called the **moments of inertia** about the x_1-, x_2-, and x_3-axes, respectively, and the negatives of the off-diagonal elements I_{12}, I_{13}, and so forth, are termed the **products of inertia**. It should be clear that the inertia tensor is symmetric; that is,

$$I_{ij} = I_{ji} \tag{11.14}$$

and, therefore, that there are only six independent elements in $\{\mathbf{I}\}$. Furthermore, the inertia tensor is composed of additive elements; the inertia tensor for a body can be considered to be the sum of the tensors for the various portions of the body. Therefore, if we consider a body as a continuous distribution of matter with mass density $\rho = \rho(\mathbf{r})$, then

$$\boxed{I_{ij} = \int_V \rho(\mathbf{r}) \left(\delta_{ij} \sum_k x_k^2 - x_i x_j \right) dv} \tag{11.15}$$

where $dv = dx_1 dx_2 dx_3$ is the element of volume at the position defined by the vector $\mathbf{r}$, and where V is the volume of the body.

EXAMPLE 11.3

Calculate the inertia tensor of a homogeneous cube of density ρ, mass M, and side of length b. Let one corner be at the origin, and let three adjacent edges lie along the coordinate axes (Figure 11-3).

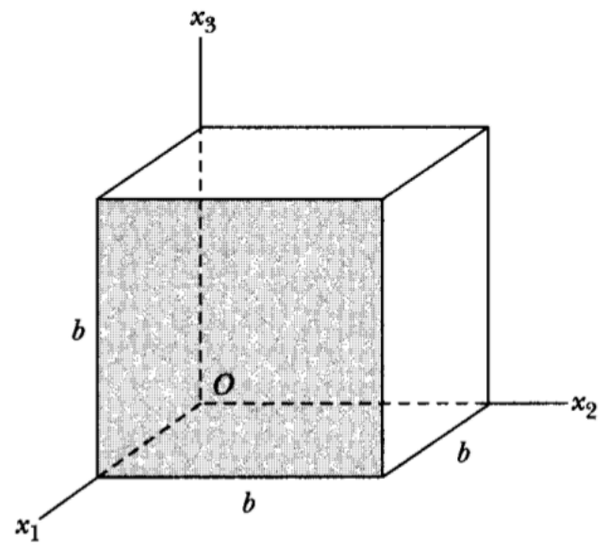


FIGURE 11-3 Example 11-3. A homogeneous cube of sides b with the origin at one corner.

For this choice of the coordinate axes, it should be obvious that the origin does not lie at the center of mass; we return to this point .

Solution. According to Equation 11.15, we have

$$I_{11} = \rho \int_0^b dx_3 \int_0^b dx_2 (x_2^2 + x_3^2) \int_0^b dx_1$$

$$= \frac{2}{3} \rho b^5 = \frac{2}{3} M b^2$$

$$I_{12} = -\rho \int_0^b x_1 dx_1 \int_0^b x_2 dx_2 \int_0^b dx_3$$

$$= -\frac{1}{4}\rho b^5 = -\frac{1}{4} Mb^2$$

It should be easy to see that all the diagonal elements are equal and, furthermore, that all the off-diagonal elements are equal. If we define $\beta = Mb^2$, we have

$$\left.\begin{aligned} I_{11} = I_{22} = I_{33} &= \frac{2}{3}\beta \\ I_{12} = I_{13} = I_{23} &= -\frac{1}{4}\beta \end{aligned}\right\}$$

The moment-of-inertia tensor then becomes

$$\{\mathbf{I}\} = \begin{Bmatrix} \frac{2}{3}\beta & -\frac{1}{4}\beta & -\frac{1}{4}\beta \\ -\frac{1}{4}\beta & \frac{2}{3}\beta & -\frac{1}{4}\beta \\ -\frac{1}{4}\beta & -\frac{1}{4}\beta & \frac{2}{3}\beta \end{Bmatrix}$$

We shall continue the investigation of the moment-of-inertia tensor for the cube in later sections.

11.4 Angular Momentum

With respect to some point O fixed in the body coordinate system, the angular momentum of the body is

$$\mathbf{L} = \sum_\alpha \mathbf{r}_\alpha \times \mathbf{p}_\alpha \tag{11.16}$$

The most convenient choice for the position of the point O depends on the particular problem. Only two choices are important: (a) if one or more points of the body are fixed (in the fixed coordinate system), O is chosen to coincide with one such point (as in the case of the rotating top. Section 11.11); (b) if no point of the body is fixed, O is chosen to be the center of mass.

Relative to the body coordinate system, the linear momentum $\mathbf{p}_\alpha$ is

$$\mathbf{p}_\alpha = m_\alpha \mathbf{v}_\alpha = m_\alpha \boldsymbol{\omega} \times \mathbf{r}_\alpha$$

Hence, the angular momentum of the body is

$$\mathbf{L} = \sum_\alpha m_\alpha \mathbf{r}_\alpha \times (\boldsymbol{\omega} \times \mathbf{r}_\alpha) \tag{11.17}$$

The vector identity

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{A}) = A^2 \mathbf{B} - \mathbf{A}(\mathbf{A} \cdot \mathbf{B})$$

can be used to express **L**:

$$\boxed{\mathbf{L} = \sum_\alpha m_\alpha \left[r_\alpha^2 \boldsymbol{\omega} - \mathbf{r}_\alpha (\mathbf{r}_\alpha \cdot \boldsymbol{\omega}) \right]} \tag{11.18}$$

The same technique we used to write T_{rot} in tensor form can now be applied here. But the angular momentum is a vector, so for the i^{th} component, we write

$$\begin{aligned} L_i &= \sum_\alpha m_\alpha \left(\omega_i \sum_k x_{\alpha,k}^2 - x_{\alpha,i} \sum_j x_{\alpha,j} \omega_j \right) \\ &= \sum_\alpha m_\alpha \sum_j \left(\omega_j \delta_{ij} \sum_k x_{\alpha,k}^2 - \omega_j x_{\alpha,i} x_{\alpha,j} \right) \\ &= \sum_j \omega_j \sum_\alpha m_\alpha \left(\delta_{ij} \sum_k x_{\alpha,k}^2 - x_{\alpha,i} x_{\alpha,j} \right) \end{aligned} \tag{11.19}$$

The summation over a can be recognized (see Equation 11.10) as the ij^{th} element of the inertia tensor. Therefore,

$$\boxed{L_i = \sum_j I_{ij}\omega_j} \tag{11.20a}$$

or, in tensor notation,

$$\mathbf{L} = \{\mathbf{I}\} \cdot \boldsymbol{\omega} \tag{11.20b}$$

Thus, the inertia tensor relates a sum over the components of the angular velocity vector to the i^{th} component of the angular momentum vector. This may at first seem a somewhat unexpected result; for, if we consider a rigid body for which the inertia tensor has nonvanishing off-diagonal elements, then even if $\boldsymbol{\omega}$ is directed along, say, the x_1-direction, $\boldsymbol{\omega} = (\omega, 0, 0)$, the angular momentum vector in general has nonvanishing components in all three directions: $\mathbf{L} = (L_1, L_2, L_3)$; that is, the angular momentum vector does not in general have the same direction as the angular velocity vector. (It should be emphasized that this statement depends on $I_{ij} \neq 0$ for $i \neq j$; we return to this point in the next section.)

As an example of $\boldsymbol{\omega}$ and $\mathbf{L}$ not being collinear, consider the rotating dumbbell in Figure 11-4.

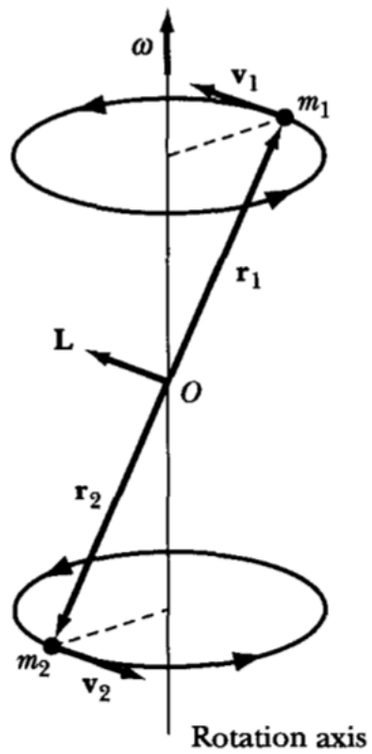


FIGURE 11-4 A dumbbell connected by masses m_1 and m_2 at the ends of its shaft. Note that $\boldsymbol{\omega}$ is not along the shaft, and that $\boldsymbol{\omega}$ and $\mathbf{L}$ are not collinear.

(We consider the shaft connecting m_1 and m_2 to be weightless and extensionless.) The relation connecting $\mathbf{r}_\alpha, \mathbf{v}_\alpha$ and $\boldsymbol{\omega}$ is

$$\mathbf{v}_\alpha = \boldsymbol{\omega} \times \mathbf{r}_\alpha$$

and the relation connecting $\mathbf{r}_\alpha, \mathbf{v}_\alpha$ and $\mathbf{L}$ is

$$\mathbf{L} = \sum_\alpha m_\alpha \mathbf{r}_\alpha \times \mathbf{v}_\alpha$$

It should be clear that $\boldsymbol{\omega}$ is directed along the axis of rotation and that $\mathbf{L}$ is perpendicular to the line connecting m_1 and m_2.

We note, for this example, that the angular-momentum vector $\mathbf{L}$ does not remain constant in time but rotates with an angular velocity ω in such a way that it traces out a cone whose axis is the axis of rotation. Therefore $\dot{\mathbf{L}} \neq 0$. But Equation 9.31 states that

$$\dot{\mathbf{L}} = \mathbf{N} \tag{11.21}$$

where $\mathbf{N}$ is the external torque applied to the body. Thus, to keep the dumbbell rotating as in Figure 11-4, we must constantly apply a torque.

We can obtain another result from Equation 11.20a by multiplying L_i by $\frac{1}{2}\omega_i$ and summing over i:

$$\frac{1}{2}\sum_i \omega_i L_i = \frac{1}{2}\sum_{i,j} I_{ij}\omega_i\omega_j = T_{\text{rot}} \tag{11.22a}$$

where the second equality is just Equation 11.11. Thus,

$$\boxed{T_{\text{rot}} = \frac{1}{2}\boldsymbol{\omega}\cdot\mathbf{L}} \tag{11.22b}$$

Equations 11.20b and 11.22b illustrate two important properties of tensors. The product of a tensor and a vector yields a vector, as in

$$\mathbf{L} = \{\mathbf{I}\} \cdot \boldsymbol{\omega}$$

and the product of a tensor and two vectors yields a scalar, as in

$$T_{\text{rot}} = \frac{1}{2}\boldsymbol{\omega} \cdot \mathbf{L} = \frac{1}{2}\boldsymbol{\omega} \cdot \{\mathbf{I}\} \cdot \boldsymbol{\omega}$$

We shall not, however, have occasion to use tensor equations in this form. We use only the summation (or integral) expressions as in Equations 11.11, 11.15, and 11.20a.

EXAMPLE 11.4

Consider the pendulum shown in Figure 11-5 composed of a rigid rod of length b with a mass m_1 at its end.

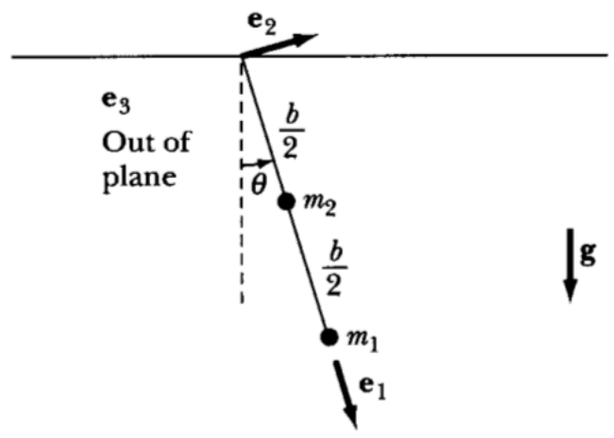


FIGURE 11-5 Example 11.4. A rigid rod rotating as a pendulum has a mass m_1 at its end and another mass m_2 halfway.

Another mass (m_2) is placed halfway down the rod. Find the frequency of small oscillations if the pendulum swings in a plane.

Solution. We use the methods of this chapter to analyze the system. Let the fixed and body systems have their origin at the pendulum pivot point. Let $\mathbf{e}_1$ be along the rod, $\mathbf{e}_2$ be in the plane, and $\mathbf{e}_3$ be out of the plane (Figure 11-5). The angular velocity is

$$\boldsymbol{\omega} = \omega_3 \mathbf{e}_3 = \dot{\theta}\mathbf{e}_3 \tag{11.23}$$

We use Equation 11.10 to find the inertia tensor. All the mass is along $\mathbf{e}_1$, with $x_{1,1} = b$ and $x_{2,1} = b/2$. All other components of $x_{\alpha,k}$ equal zero.

$$I_{ij} = m_1(\delta_{ij} x_{1,1}^2 - x_{1,i} x_{1,j}) + m_2(\delta_{ij} x_{2,1}^2 - x_{2,i} x_{2,j}) \tag{11.24}$$

The inertia tensor. Equation 11.13a, becomes

$$\{\mathbf{I}\} = \begin{Bmatrix} 0 & 0 & 0 \\ 0 & m_1 b^2 + m_2 \dfrac{b^2}{4} & 0 \\ 0 & 0 & m_1 b^2 + m_2 \dfrac{b^2}{4} \end{Bmatrix} \tag{11.25}$$

We determine the angular momentum from Equation 11.20a:

$$\left.\begin{aligned} L_1 &= 0 \\ L_2 &= 0 \\ L_3 &= I_{33}\omega_3 = \left(m_1 b^2 + m_2 \frac{b^2}{4}\right)\dot{\theta} \end{aligned}\right\} \tag{11.26}$$

The only external force is gravity, which causes a torque $\mathbf{N}$ on the system. Because $\dot{\mathbf{L}} = \mathbf{N}$, we have

$$\left(m_1 b^2 + m_2 \frac{b^2}{4}\right)\ddot{\theta}\mathbf{e}_3 = \sum_\alpha \mathbf{r}_\alpha \times \mathbf{F}_\alpha \tag{11.27}$$

Because the gravitational force is down,

$$\mathbf{g} = g\cos\theta\,\mathbf{e}_1 - g\sin\theta\,\mathbf{e}_2$$

Thus,

$$\mathbf{r}_1 \times \mathbf{F}_1 = b\mathbf{e}_1 \times (\cos\theta\mathbf{e}_1 - \sin\theta\mathbf{e}_2)\, m_1 g = -m_1 g b \sin\theta\mathbf{e}_3$$

$$\mathbf{r}_2 \times \mathbf{F}_2 = \frac{b}{2}\mathbf{e}_1 \times (\cos\theta\mathbf{e}_1 - \sin\theta\mathbf{e}_2)\, m_2 g = -m_2 g \frac{b}{2}\sin\theta\mathbf{e}_3$$

Equation 11.27 becomes

$$b^2\left(m_1 + \frac{m_2}{4}\right)\ddot{\theta} = -bg\sin\theta\left(m_1 + \frac{m_2}{2}\right) \tag{11.28}$$

and the frequency of small oscillations is

$$\omega_0^2 = \frac{m_1 + \dfrac{m_2}{2}}{m_1 + \dfrac{m_2}{4}}\,\frac{g}{b} \tag{11.29}$$

We can check Equation 11.29 by noting that $\omega_0^2 \approx g/b$ for $m_1 \gg m_2$ and $\omega_0^2 \approx 2g/b$ for $m_2 \gg m_1$ as it should.

This example could have just as easily been solved by finding the kinetic energy from Equation 11.22a and using Lagrange's equations of motion. We would then have

$$T_{\text{rot}} = \frac{1}{2}\omega_3 L_3 = \frac{1}{2}\omega_3^2 I_{33}$$

$$= \frac{1}{2}\left(m_1 b^2 + m_2\frac{b^2}{4}\right)\dot{\theta}^2 \tag{11.30}$$

$$U = -m_1 g b\cos\theta - m_2 g\frac{b}{2}\cos\theta \tag{11.31}$$

Where U= 0 at the origin. The equation of motion (Equation 11.28) follows directly from a straightforward application of the Lagrangian technique.

11.5 Principal Axes of Inertia

It should be clear that a considerable simplification in the expressions for T and **L** would result if the inertia tensor consisted only of diagonal elements. If we could write

$$I_{ij} = I_i \delta_{ij} \tag{11.32}$$

then the inertia tensor would be

$$\{\mathbf{I}\} = \begin{Bmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{Bmatrix} \tag{11.33}$$

We would then have

$$L_i = \sum_j I_i \delta_{ij} \omega_j = I_i \omega_i \tag{11.34}$$

and

$$T_{\text{rot}} = \frac{1}{2} \sum_{i,j} I_i \delta_{ij} \omega_i \omega_j = \frac{1}{2} \sum_i I_i \omega_i^2 \tag{11.35}$$

Thus, the condition that {**I**} have only diagonal elements provides quite simple expressions for the angular momentum and the rotational kinetic energy. We now determine the conditions under which Equation 11.32 becomes the description of the inertia tensor. This involves finding a set of body axes for which the products of inertia (i.e., the off-diagonal elements of {**I**}) vanish. We call such axes the **principal axes of inertia**.

If a body rotates around a principal axis, both the angular velocity and the angular momentum are, according to Equation 11.34, directed along this axis.

Then, if I is the rotational inertia (moment of inertia) about this axis, we can write

$$\mathbf{L} = I\boldsymbol{\omega} \tag{11.36}$$

Equating the components of **L** in Equations 11.20a and 11.36, we have

$$\left.\begin{aligned} L_1 &= I\omega_1 = I_{11}\omega_1 + I_{12}\omega_2 + I_{13}\omega_3 \\ L_2 &= I\omega_2 = I_{21}\omega_1 + I_{22}\omega_2 + I_{23}\omega_3 \\ L_3 &= I\omega_3 = I_{31}\omega_1 + I_{32}\omega_2 + I_{33}\omega_3 \end{aligned}\right\} \tag{11.37}$$

Or, collecting terms, we obtain

$$\left.\begin{aligned} (I_{11} - I)\omega_1 + I_{12}\omega_2 + I_{13}\omega_3 &= 0 \\ I_{21}\omega_1 + (I_{22} - I)\omega_2 + I_{23}\omega_3 &= 0 \\ I_{31}\omega_1 + I_{32}\omega_2 + (I_{33} - I)\omega_3 &= 0 \end{aligned}\right\} \tag{11.38}$$

The condition that these equations have a nontrivial solution is that the determinant of the coefficients vanish:

$$\begin{vmatrix} (I_{11} - I) & I_{12} & I_{13} \\ I_{21} & (I_{22} - I) & I_{23} \\ I_{31} & I_{32} & (I_{33} - I) \end{vmatrix} = 0 \tag{11.39}$$

The expansion of this determinant leads to the **secular** or **characteristic equation** for **I**, which is a cubic. Each of the three roots corresponds to a moment of inertia about one of the principal axes. These values, I_1, I_2, and I_3, are called the **principal moments of inertia**. If the body rotates about the axis corresponding to the principal moment I_1, then Equation 11.36 becomes $\mathbf{L} = I_1\boldsymbol{\omega}$ - that is, both $\boldsymbol{\omega}$ and **L** are directed along this axis. The direction of $\boldsymbol{\omega}$ with respect to the body coordinate system is then the same as the direction of the principal axis corresponding to I_1. Therefore, we can determine the direction of this principal axis by substituting I_1 for I in Equation 11.38 and determining the ratios of the components of the angular-velocity vector: $\omega_1 : \omega_2 : \omega_3$ We thereby determine the direction cosines of the axis about which the moment of inertia is I_1. The directions corresponding to I_2 and I_3 can be found in a similar fashion. That the principal axes determined in this manner are indeed real and orthogonal is proved in Section 11.7; these results also follow from the more general considerations given in Section 12.4.

The fact that the diagonalization procedure just described yields only the ratios of the components of $\boldsymbol{\omega}$ is no handicap, because the ratios completely determine the direction of each of the principal axes, and it is only the directions of these axes that are required. Indeed, we would not expect the *magnitudes* of the ω_i to be determined, because the actual rate of the body's angular motion cannot be specified by the geometry alone. We are free to impress on the body any magnitude of the angular velocity we wish.

For most of the problems encountered in rigid-body dynamics, the bodies are of some regular shape, so we can determine the principal axes merely by examining the symmetry of the body. For example, any body that is a solid of revolution (e.g., a cylindrical rod) has one principal axis that lies along the symmetry axis (e.g., the center line of the cylindrical rod), and the other two axes are in a plane perpendicular to the symmetry axis. It should be obvious that because the body is symmetrical, the choice of the angular placement of these other two axes is arbitrary. If the moment of inertia along the symmetry axis is I_1, then $I_2 = I_3$ for a solid of revolution - that is, the secular equation has a double root.

If a body has $I_1 = I_2 = I_3$, it is termed a **spherical top**; if $I_1 = I_2 \neq I_3$, it is termed a **symmetric top**; if the principal moments of inertia are all distinct, it is termed an **asymmetric top**. If a body has $I_1 = 0$, $I_2 = I_3$, as, for example, two point masses connected by a weightless shaft, or a diatomic molecule, it is called a r**otor**.

EXAMPLE 11.5

Find the principal moments of inertia and the principal axes for the cube in Example 11.3.

Solution. In Example 11.3, we found that the moment-of-inertia tensor for a cube (with origin at one corner) had nonzero off-diagonal elements. Evidently, the coordinate axes chosen for that calculation were not principal axes. If, for example, the cube rotates about the x_3-axis, then $\boldsymbol{\omega} = \omega_3 \mathbf{e}_3$ and the angular momentum vector **L** (see Equation 11.37) has the components

$$L_1 = -\frac{1}{4}\beta\omega_3$$

$$L_2 = -\frac{1}{4}\beta\omega_3$$

$$L_3 = \frac{2}{3}\beta\omega_3$$

Thus,

$$\mathbf{L} = Mb^2\omega_3\left(-\frac{1}{4}\mathbf{e}_1 - \frac{1}{4}\mathbf{e}_2 + \frac{2}{3}\mathbf{e}_3\right)$$

which is not in the same direction as $\boldsymbol{\omega}$.

To find the principal moments of inertia, we must solve the secular equation

$$\begin{vmatrix} \frac{2}{3}\beta - I & -\frac{1}{4}\beta & -\frac{1}{4}\beta \\ -\frac{1}{4}\beta & \frac{2}{3}\beta - I & -\frac{1}{4}\beta \\ -\frac{1}{4}\beta & -\frac{1}{4}\beta & \frac{2}{3}\beta - I \end{vmatrix} = 0 \tag{11.40}$$

The value of a determinant is not affected by adding (or subtracting) any row (or column) from any other row (or column). Equation 11.40 can be solved more easily if we subtract the first row from the second:

$$\begin{vmatrix} \frac{2}{3}\beta - I & -\frac{1}{4}\beta & -\frac{1}{4}\beta \\ -\frac{11}{12}\beta + I & \frac{11}{12}\beta - I & 0 \\ -\frac{1}{4}\beta & -\frac{1}{4}\beta & \frac{2}{3}\beta - I \end{vmatrix} = 0$$

We can factor $(\frac{11}{12}\beta - I)$ from the second row:

$$\left(\frac{11}{12}\beta - I\right)\begin{vmatrix} \frac{2}{3}\beta - I & -\frac{1}{4}\beta & -\frac{1}{4}\beta \\ -1 & 1 & 0 \\ -\frac{1}{4}\beta & -\frac{1}{4}\beta & \frac{2}{3}\beta - I \end{vmatrix} = 0$$

Expanding, we have

$$\left(\frac{11}{12}\beta - I\right)\left[\left(\frac{2}{3}\beta - I\right)^2 - \frac{1}{8}\beta^2 - \frac{1}{4}\beta\left(\frac{2}{3}\beta - I\right)\right] = 0$$

which can be factored to obtain

$$\left(\frac{1}{6}\beta - I\right)\left(\frac{11}{12}\beta - I\right)\left(\frac{11}{12}\beta - I\right) = 0$$

Thus, we have the following roots, which give the principal moments of inertia:

$$I_1 = \frac{1}{6}\beta, \quad I_2 = \frac{11}{12}\beta, \quad I_3 = \frac{11}{12}\beta$$

The diagonalized moment-of-inertia tensor becomes

$$\{\mathbf{I}\} = \begin{Bmatrix} \frac{1}{6}\beta & 0 & 0 \\ 0 & \frac{11}{12}\beta & 0 \\ 0 & 0 & \frac{11}{12}\beta \end{Bmatrix} \tag{11.41}$$

Because two of the roots are identical, $I_2 = I_3$, the principal axis associated with I_1 must be an axis of symmetry.

To find the direction of the principal axis associated with I_1, we substitute for I in Equation 11.38 the value $I = I_1 = \frac{1}{6}\beta$:

$$\left.\begin{aligned}\left(\frac{2}{3}\beta - \frac{1}{6}\beta\right)\omega_{11} - \frac{1}{4}\beta\omega_{21} - \frac{1}{4}\beta\omega_{31} &= 0\\ -\frac{1}{4}\beta\omega_{11} + \left(\frac{2}{3}\beta - \frac{1}{6}\beta\right)\omega_{21} - \frac{1}{4}\beta\omega_{31} &= 0\\ -\frac{1}{4}\beta\omega_{11} - \frac{1}{4}\beta\omega_{21} + \left(\frac{2}{3}\beta - \frac{1}{6}\beta\right)\omega_{31} &= 0\end{aligned}\right\}$$

where the second subscript 1 on the ω_i signifies that we are considering the principal axis associated with I_1. Dividing the first two of these equations by $\beta/4$, we have

$$\left.\begin{aligned}2\omega_{11} - \omega_{21} - \omega_{31} &= 0\\ -\omega_{11} + 2\omega_{21} - \omega_{31} &= 0\end{aligned}\right\} \tag{11.42}$$

Subtracting the second of these equations from the first, we find $\omega_{11} = \omega_{21}$. Using this result in either of the Equations 11.42, we obtain $\omega_{11} = \omega_{21} = \omega_{31}$ and the desired ratios are

$$\omega_{11}:\omega_{21}:\omega_{31} = 1:1:1$$

Therefore, when the cube rotates about an axis that has associated with it the moment of inertia $I_1 = \frac{1}{6}\beta = \frac{1}{6}Mb^2$, the projections of $\boldsymbol{\omega}$ on the three coordinate axes are all equal. Hence, this principal axis corresponds to the diagonal of the cube.

Because the moments I_2 and I_3 are equal, the orientation of the principal axes associated with these moments is arbitrary; they need only lie in a plane normal to the diagonal of the cube.

11.6 Moments of Inertia for Different Body Coordinate Systems

For the kinetic energy to be separable into translational and rotational portions (see Equation 11.6), it is, in general, necessary to choose a body coordinate system whose origin is the center of mass of the body. For certain geometrical shapes, it may not always be convenient to compute the elements of the inertia tensor using such a coordinate system. We therefore consider some other set of coordinate axes X_i, also fixed with respect to the body and having the same orientation as the x_i-axes but with an origin Q that does not correspond with the origin O (located at the center of mass of the body coordinate system). Origin Q may be located either within or outside the body under consideration.

The elements of the inertia tensor relative to the X_i-axes can be written as

$$J_{ij} = \sum_{\alpha} m_{\alpha} \left(\delta_{ij} \sum_{k} X_{\alpha,k}^2 - X_{\alpha,i} X_{\alpha,j} \right) \tag{11.43}$$

If the vector connecting Q with O is **a**, then the general vector **R** (Figure 11-6) can be written as

$$\mathbf{R} = \mathbf{a} + \mathbf{r} \tag{11.44}$$

with components

$$X_i = a_i + x_i \tag{11.45}$$

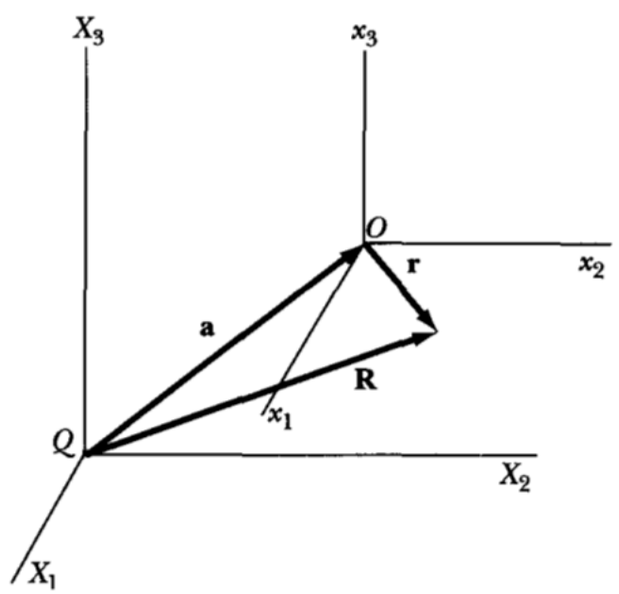


FIGURE 11-6 The coordinate axes X_i are fixed in the body and have the same orientation as the x_i-axes, but its origin Q is not located at the origin O (at center of mass of body system).

Using Equation 11.45, the tensor element J_{ij} becomes

$$J_{ij} = \sum_\alpha m_\alpha \left(\delta_{ij} \sum_k (x_{\alpha,k} + a_k)^2 - (x_{\alpha,i} + a_i)(x_{\alpha,j} + a_j) \right)$$

$$= \sum_\alpha m_\alpha \left(\delta_{ij} \sum_k x_{\alpha,k}^2 - x_{\alpha,i} x_{\alpha,j} \right)$$

$$+ \sum_\alpha m_\alpha \left(\delta_{ij} \sum_k (2 x_{\alpha,k} a_k + a_k^2) - (a_i x_{\alpha,j} + a_j x_{\alpha,i} + a_i a_j) \right) \tag{11.46}$$

Identifying the first summation as I_{ij}, we have, on regrouping,

$$J_{ij} = I_{ij} + \sum_\alpha m_\alpha \left(\delta_{ij} \sum_k a_k^2 - a_i a_j \right) + \sum_\alpha m_\alpha \left(2\delta_{ij} \sum_k x_{\alpha,k} a_k - a_i x_{\alpha,j} - a_j x_{\alpha,i} \right) \tag{11.47}$$

But each term in the last summation involves a sum of the form

$$\sum_\alpha m_\alpha x_{\alpha,k}$$

We know, however, that because O is located at the center of mass,

$$\sum_\alpha m_\alpha \mathbf{r}_\alpha = 0$$

or, for the k^{th} component.

$$\sum_\alpha m_\alpha x_{\alpha,k} = 0$$

Therefore, all such terms in Equation 11.47 vanish and we have

$$J_{ij} = I_{ij} + \sum_\alpha m_\alpha \left(\delta_{ij} \sum_k a_k^2 - a_i a_j \right) \tag{11.48}$$

But

$$\sum_\alpha m_\alpha = M \quad \text{and} \quad \sum_k a_k^2 \equiv a^2$$

Solving for I_{ij}, we have the result

$$\boxed{I_{ij} = J_{ij} - M(a^2\delta_{ij} - a_i a_j)} \tag{11.49}$$

which allows the calculation of the elements I_{ij} of the desired inertia tensor (with origin at the center of mass) once those with respect to the X_i-axes are known. The second term on the right-hand side of Equation 11.49 is the inertia tensor referred to the origin Q for a point mass M.

Equation 11.49 is the general form of **Steiner's parallel-axis theorem**, the simplified form of which is given in elementary treatments. Consider, for example. Figure 11-7.

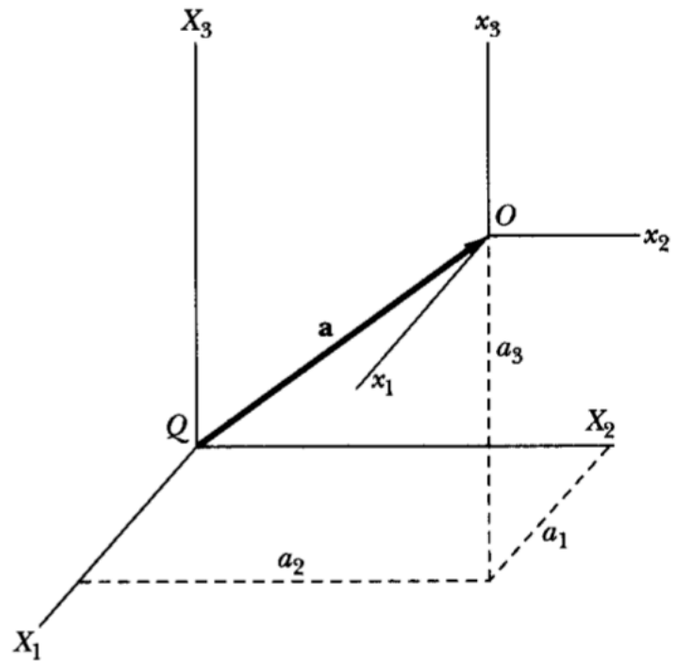


FIGURE 11-7 The elements I_{ij} in the x_i-axes are related to those (J_{ij}) in the X_{ij}-axes by Equation 11.49. The vector **a** connects the origin Q with the origin O.

Element I_{11} is

$$I_{11} = J_{11} - M[(a_1^2 + a_2^2 + a_3^2)\delta_{11} - a_1^2]$$

$$= J_{11} - M(a_2^2 + a_3^2)$$

which states that the difference between the elements is equal to the mass of the body multiplied by the square of the distance between the parallel axes (in this case, between the x_i- and X_i-axes).

EXAMPLE 11.6

Find the inertia tensor of the cube of Example 11.3 in a coordinate system with origin at the center of mass.

Solution. In Example 11.3, with the origin at the corner of the cube, we found the inertia tensor to be

$$\{\mathbf{J}\} = \begin{Bmatrix} \frac{2}{3}Mb^2 & -\frac{1}{4}Mb^2 & -\frac{1}{4}Mb^2 \\ -\frac{1}{4}Mb^2 & \frac{2}{3}Mb^2 & -\frac{1}{4}Mb^2 \\ -\frac{1}{4}Mb^2 & -\frac{1}{4}Mb^2 & \frac{2}{3}Mb^2 \end{Bmatrix} \tag{11.50}$$

We may now use Equation 11.49 to obtain the inertia tensor $\{\mathbf{I}\}$ referred to a coordinate system with origin at the center of mass. In keeping with the notation of this section, we call the new axes x_i with origin O and call the previous axes X_i with origin Q at one corner of the cube (Figure 11-8).

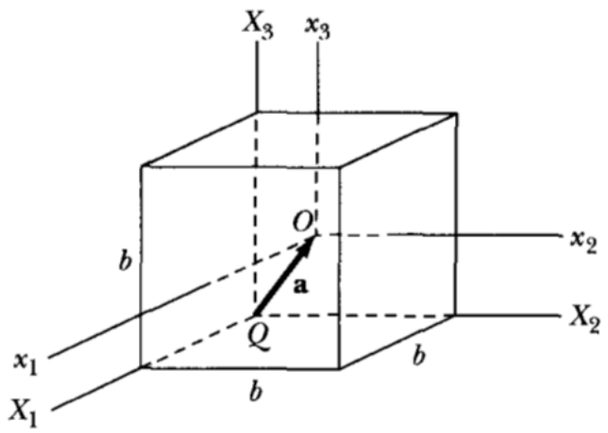


FIGURE 11-8 Example 11.6. The X_i-axes have their origin Q at one corner of a cube of sides b. The system x_i has its origin O at the cube's center of mass.

The center of mass of the cube is at the point (b/2, b/2, b/2) in the X_i coordinate system, and the components of the vector **a** therefore are

$$a_1 = a_2 = a_3 = b/2$$

From Equation 11.50, we have

$$\left.\begin{aligned} J_{11} = J_{22} = J_{33} &= \frac{2}{3} Mb^2 \\ J_{12} = J_{13} = J_{23} &= -\frac{1}{4} Mb^2 \end{aligned}\right\}$$

And applying Equation 11.49, we find

$$\begin{aligned} I_{11} &= J_{11} - M(a^2 - a_1^2) \\ &= J_{11} - M(a_2^2 + a_3^2) \\ &= \frac{2}{3} Mb^2 - \frac{1}{2} Mb^2 = \frac{1}{6} Mb^2 \end{aligned}$$

and

$$\begin{aligned} I_{12} &= J_{12} - M(-a_1 a_2) \\ &= -\frac{1}{4} Mb^2 + \frac{1}{4} Mb^2 = 0 \end{aligned}$$

Altogether, we have

$$\begin{aligned} I_{11} = I_{22} = I_{33} &= \frac{1}{6} Mb^2 \\ I_{12} = I_{13} = I_{23} &= 0 \end{aligned}$$

The inertia tensor is therefore diagonal:

$$\{\mathbf{I}\} = \begin{Bmatrix} \frac{1}{6} Mb^2 & 0 & 0 \\ 0 & \frac{1}{6} Mb^2 & 0 \\ 0 & 0 & \frac{1}{6} Mb^2 \end{Bmatrix} \tag{11.51}$$

If we factor out the common term $\frac{1}{6}Mb^2$ from this expression, we can write

$$\{\mathbf{I}\} = \frac{1}{6} Mb^2 \{\mathbf{1}\} \tag{11.52}$$

where $\{\mathbf{1}\}$ is the unit tensor:

$$\{\mathbf{1}\} \equiv \begin{Bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Bmatrix} \tag{11.53}$$

Thus, we find that, for the choice of the origin at the center of mass of the cube, the principal axes are perpendicular to the faces of the cube. Because, from a physical standpoint, nothing distinguishes any one of these axes from another, the principal moments of inertia are all equal for this case. We note further that, as long as we maintain the origin at the center of mass, then the inertia tensor is the same for any orientation of the coordinate axes and these axes are equally valid principal axes.

11.7 Further Properties of the Inertia Tensor

Before attacking the problems of rigid-body dynamics by obtaining the general equations of motion, we should consider the fundamental importance of some of the operations we have been discussing. Let us begin by examining the properties of the inertia tensor under coordinate transformations.

We have already obtained the fundamental relation connecting the inertia tensor and the angular momentum and angular velocity vectors (Equation 11.20), which we can write as

$$L_k = \sum_l I_{kl} \omega_l \tag{11.54a}$$

Because this is a vector equation, in a coordinate system rotated with respect to the system for which Equation 11.54a applies, we must have an entirely analogous relation,

$$L'_i = \sum_j I'_{ij}\omega'_j \tag{11.54b}$$

where the primed quantities all refer to the rotated system. Both **L** and $\boldsymbol{\omega}$ obey the standard transformation equation for vectors (Equation L8):

$$x_i = \sum_j \lambda^t_{ij} x'_j = \sum_j \lambda_{ji} x'_j$$

We can therefore write

$$L_k = \sum_m \lambda_{mk} L'_m \tag{11.55a}$$

and

$$\omega_l = \sum_j \lambda_{jl}\omega'_j \tag{11.55b}$$

If we substitute Equations 11.55a and b into Equation 11.54a, we obtain

$$\sum_m \lambda_{mk} L'_m = \sum_l I_{kl} \sum_j \lambda_{jl}\omega'_j \tag{11.56}$$

Next, we multiply both sides of this equation by λ_{ik} and sum over k:

$$\sum_m \left(\sum_k \lambda_{ik}\lambda_{mk}\right) L'_m = \sum_j \left(\sum_{k,l} \lambda_{ik}\lambda_{jl} I_{kl}\right)\omega'_j \tag{11.57}$$

The term in parentheses on the left-hand side is just δ_{im}, so performing the summation over m we obtain

$$L'_i = \sum_j \left(\sum_{k,l} \lambda_{ik}\lambda_{jl} I_{kl}\right)\omega'_j \tag{11.58}$$

For this equation to be identical with Equation 11.54b, we must have

$$I'_{ij} = \sum_{k,l} \lambda_{ik}\lambda_{jl} I_{kl} \tag{11.59}$$

This is therefore the rule that the inertia tensor must obey under a coordinate transformation.

Equation 11.59 is, in fact, the general rule specifying the manner in which any second-rank tensor must transform. For a tensor {**T**} of arbitrary rank, the statement is

$$T'_{abcd\ldots} = \sum_{i,j,k,l,\ldots} \lambda_{ai}\lambda_{bj}\lambda_{ck}\lambda_{dl}\ldots T_{ijkl\ldots} \tag{11.60}$$

Note that a tensor of the first rank transforms as

$$T'_a = \sum_i \lambda_{ai} T_i$$

Such a tensor is in fact a vector. A tensor of zero rank implies that T' = T, or that such a tensor is a scalar.

Note that we can write Equation 11.59 as

$$\boxed{I'_{ij} = \sum_{k,l} \lambda_{ik} I_{kl} \lambda^t_{lj}} \tag{11.61}$$

Although matrices and tensors are distinct types of mathematical objects, the manipulation of tensors is in many respects the same as for matrices. Thus, Equation 11.61 can be expressed as a matrix equation:

$$\mathbf{I}' = \boldsymbol{\lambda}\mathbf{I}\boldsymbol{\lambda}^t \tag{11.62}$$

where we understand **I** to be the matrix consisting of the elements of the tensor {**I**}. Because we are considering only orthogonal transformation matrices, the transpose of $\boldsymbol{\lambda}$ is equal to its inverse, so we can express Equation 11.62 as

$$\boxed{\mathbf{I}' = \boldsymbol{\lambda}\mathbf{I}\boldsymbol{\lambda}^{-1}} \tag{11.63}$$

A transformation of this general type is called a **similarity transformation** (**I'** is similar to **I**).

EXAMPLE 11.7

Prove the assertion stated in Example 11.6 that the inertia tensor for a cube (with origin at the center of mass) is independent of the orientation of the axes.

Solution. The change in the inertia tensor under a rotation of the coordinate axes can be computed by making a similarity transformation. Thus, if the rotation is described by the matrix $\boldsymbol{\lambda}$, we have

$$\mathbf{I}' = \boldsymbol{\lambda}\mathbf{I}\boldsymbol{\lambda}^{-1} \tag{11.64}$$

But the matrix **I**, which is derived from the elements of the tensor {**I**} (Equation 11.52 of Example 11.4), is just the identity matrix **1** multiplied by a constant:

$$\mathbf{I} = \frac{1}{6}Mb^2\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} = \frac{1}{6}Mb^2\,\mathbf{1} \tag{11.65}$$

Therefore, the operations specified in Equation 11.64 are trivial:

$$\mathbf{I}' = \frac{1}{6}Mb^2\boldsymbol{\lambda}\mathbf{1}\boldsymbol{\lambda}^{-1} = \frac{1}{6}Mb^2\boldsymbol{\lambda}\boldsymbol{\lambda}^{-1} = \frac{1}{6}Mb^2\mathbf{1} = \mathbf{I} \tag{11.66}$$

Thus, the transformed inertia tensor is identical to the original tensor, independent of the details of the rotation.

Let us next determine what condition must be satisfied if we take an arbitrary inertia tensor and perform a coordinate rotation in such a way that the transformed inertia tensor is diagonal. Such an operation implies that the quantity I'_{ij} in Equation 11.59 must satisfy (see Equation 11.32) the relation

$$I'_{ij} = I_i \delta_{ij} \tag{11.67}$$

Thus,

$$I_i \delta_{ij} = \sum_{k,l} \lambda_{ik} \lambda_{jl} I_{kl} \tag{11.68}$$

If we multiply both sides of this equation by λ_{im} and sum over i, we obtain

$$\sum_i I_i \lambda_{im} \delta_{ij} = \sum_{k,l} \left(\sum_i \lambda_{im} \lambda_{ik} \right) \lambda_{jl} I_{kl} \tag{11.69}$$

The term in parentheses is just δ_{mk}, so the summation over i on the left-hand side of the equation and the summation over k on the right-hand side yield

$$I_j \lambda_{jm} = \sum_l \lambda_{jl} I_{ml} \tag{11.70}$$

Now the left-hand side of this equation can be written as

$$I_j \lambda_{jm} = \sum_l I_j \lambda_{jl} \delta_{ml} \tag{11.71}$$

so Equation 11.70 becomes

$$\sum_l I_j \lambda_{jl} \delta_{ml} = \sum_l \lambda_{jl} I_{ml} \tag{11.72a}$$

or

$$\sum_l (I_{ml} - I_j \delta_{ml}) \lambda_{jl} = 0 \tag{11.72b}$$

This is a set of simultaneous linear algebraic equations; for each value of j there are three such equations, one for each of the three possible values of m. For a nontrivial solution to exist, the determinant of the coefficients must vanish, so the principal moments of inertia, I_1, I_2, and I_3, are obtained as roots of the secular determinant for I:

$$\boxed{|I_{ml} - I\delta_{ml}| = 0} \tag{11.73}$$

This equation is just Equation 11.39; it is a cubic equation that yields the principal moments of inertia.

Thus, for any inertia tensor, the elements of which are computed for a given origin, it is possible to perform a rotation of the coordinate axes about that origin in such a way that the inertia tensor becomes diagonal. The new coordinate axes are then the principal axes of the body, and the new moments are the principal moments of inertia. Thus, for any body and for any choice of origin, there always exists a set of principal axes.

EXAMPLE 11.8

For the cube of Example 11.3, diagonalize the inertia tensor by rotating the coordinate axes.

Solution. We choose the origin to lie at one corner and perform the rotation in such a manner that the x_1-axis is rotated to the original diagonal of the cube. Such a rotation can conveniently be made in two steps: first, we rotate through an angle of 45° about the x_3-axis; second, we rotate through an angle of $\cos^{-1}(\sqrt{2/3})$ about the x_2'-axis. The first rotation matrix is

$$\boldsymbol{\lambda}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{11.74}$$

and the second rotation matrix is

$$\boldsymbol{\lambda}_2 = \begin{pmatrix} \sqrt{\frac{2}{3}} & 0 & \frac{1}{\sqrt{3}} \\ 0 & 1 & 0 \\ -\frac{1}{\sqrt{3}} & 0 & \sqrt{\frac{2}{3}} \end{pmatrix} \tag{11.75}$$

The complete rotation matrix is

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}_2\boldsymbol{\lambda}_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \sqrt{\frac{2}{3}} \end{pmatrix} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ -\sqrt{\frac{3}{2}} & \sqrt{\frac{3}{2}} & 0 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & \sqrt{2} \end{pmatrix} \tag{11.76}$$

The matrix form of the transformed inertia tensor (see Equation 11.62) is

$$\mathbf{I}' = \boldsymbol{\lambda}\mathbf{I}\boldsymbol{\lambda}^t \tag{11.77}$$

or, factoring β out of $\mathbf{I}$,

$$\mathbf{I}' = \frac{\beta}{3} \begin{pmatrix} 1 & 1 & 1 \\ -\sqrt{\frac{3}{2}} & \sqrt{\frac{3}{2}} & 0 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & \sqrt{2} \end{pmatrix} \begin{pmatrix} \frac{2}{3} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{2}{3} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 1 & -\sqrt{\frac{3}{2}} & -\frac{1}{\sqrt{2}} \\ 1 & \sqrt{\frac{3}{2}} & -\frac{1}{\sqrt{2}} \\ 1 & 0 & \sqrt{2} \end{pmatrix}$$

$$\mathbf{I}' = \frac{\beta}{3}\begin{pmatrix} 1 & 1 & 1 \\ -\sqrt{\frac{3}{2}} & \sqrt{\frac{3}{2}} & 0 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & \sqrt{2} \end{pmatrix}\begin{pmatrix} \frac{1}{6} & -\frac{11}{12}\sqrt{\frac{3}{2}} & -\frac{11}{12}\frac{\sqrt{2}}{2} \\ \frac{1}{6} & \frac{11}{12}\sqrt{\frac{3}{2}} & -\frac{11}{12}\frac{\sqrt{2}}{2} \\ \frac{1}{6} & 0 & \frac{11}{12}\sqrt{2} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{6}\beta & 0 & 0 \\ 0 & \frac{11}{12}\beta & 0 \\ 0 & 0 & \frac{11}{12}\beta \end{pmatrix} \tag{11.78}$$

Equation 11.78 is just the matrix form of the inertia tensor found by the diagonalization procedure using the secular determinant (Equation 11.41 of Example 11.5).

We have demonstrated two general procedures to diagonalize the inertia tensor. We previously pointed out that these methods are not limited to the inertia tensor but are generally valid. Either procedure can be very complicated. For example, if we wish to use the rotation procedure in the most general case, we must first construct a matrix that describes an arbitrary rotation. This entails three separate rotations, one about each of the coordinate axes. This rotation matrix must then be applied to the tensor in a similarity transformation. The off-diagonal elements of the resulting matrix must then be examined and values of the rotation angles determined so that these off-diagonal elements vanish. The actual use of such a procedure can tax the limits of human patience, but in some simple situations, this method of diagonalization can be used with profit. This is particularly true if the geometry of the problem indicates that only a simple rotation about one of the coordinate axes is necessary; the rotation angle can then be evaluated without difficulty (see, for example. Problems 11-16, 11-18, and 11-19).

In practice, there are systematic procedures for finding principal moments and principal axes of any inertia tensor. Standard computer programs and hand-calculator methods are available to find the n roots of an n^{th}-order polynomial and to diagonalize a matrix. When the principal moments are known, the principal axes are easily found.

The example of the cube illustrates the important point that the elements of the inertia tensor, the values of the principal moments of inertia, and the orientation of the principal axes for a rigid body all depend on the choice of origin for the system. Recall, however, that for the kinetic energy to be separable into translational and rotational portions, the origin of the body coordinate system must, in general, be taken to coincide with the center of mass of the body. However, for any choice of the origin for any body, there always exists an orientation of the axes that diagonalizes the inertia tensor. Hence, these axes become principal axes for that particular origin.

Next, we seek to prove that the principal axes actually form an orthogonal set. Let us assume that we have solved the secular equation and have determined the principal moments of inertia, all of which are distinct. We know that for each principal moment there exists a corresponding principal axis with the property that, if the angular velocity vector to lies along this axis, then the angular momentum vector **L** is similarly oriented; that is, to each I_j there corresponds an angular velocity $\boldsymbol{\omega}$ with components $\omega_{1j}, \omega_{2j}, \omega_{3j}$. (We use the subscript on the vector $\boldsymbol{\omega}$ and the second subscript on the components of to to designate the principal moment with which we are concerned.) For the m^{th} principal moment, we have

$$L_{im} = I_m \omega_{im} \tag{11.79}$$

In terms of the elements of the moment-of-inertia tensor, we also have

$$L_{im} = \sum_k I_{ik} \omega_{km} \tag{11.80}$$

Combining these two relations, we have

$$\sum_k I_{ik}\omega_{km} = I_m\omega_{im} \tag{11.81a}$$

Similarly, we can write for the nth principal moment:

$$\sum_i I_{ki}\omega_{in} = I_n\omega_{kn} \tag{11.81b}$$

If we multiply Equation 11.81a by ω_{in} and sum over i and then multiply Equation 11.81b by ω_{km} and sum over k, we have

$$\left.\begin{aligned}\sum_{i,k} I_{ik}\omega_{km}\omega_{in} &= \sum_i I_m\omega_{im}\omega_{in}\\ \sum_{i,k} I_{ki}\omega_{in}\omega_{km} &= \sum_k I_n\omega_{kn}\omega_{km}\end{aligned}\right\} \tag{11.82}$$

The left-hand sides of these equations are identical, because the inertia tensor is symmetrical ($I_{ik}=I_{ki}$). Therefore, on subtracting the second equation from the first, we have

$$I_m\sum_i \omega_{im}\omega_{in} - I_n\sum_k \omega_{km}\omega_{kn} = 0 \tag{11.83}$$

Because i and k are both dummy indices, we can replace them by l, say, and obtain

$$(I_m - I_n)\sum_l \omega_{lm}\omega_{ln} = 0 \tag{11.84}$$

By hypothesis, the principal moments are distinct, so that $I_m \neq I_n$. Therefore, Equation 11.84 can be satisfied only if

$$\sum_l \omega_{lm}\omega_{ln} = 0 \tag{11.85}$$

But this summation is just the definition of the scalar product of the vectors $\boldsymbol{\omega}_m$ and $\boldsymbol{\omega}_n$. Hence,

$$\boldsymbol{\omega}_m \cdot \boldsymbol{\omega}_n = 0 \tag{11.86}$$

Because the principal moments I_m and I_n were picked arbitrarily from the set of three moments, we conclude that each pair of principal axes is perpendicular; the three principal axes therefore constitute an orthogonal set.

If a double root of the secular equation exists, so that the principal moments are I_1, $I_2 = I_3$, then the preceding analysis shows that the angular velocity vectors satisfy the relations

$$\boldsymbol{\omega}_1 \perp \boldsymbol{\omega}_2, \qquad \boldsymbol{\omega}_1 \perp \boldsymbol{\omega}_3$$

but that nothing may be said regarding the angle between $\boldsymbol{\omega}_2$ and $\boldsymbol{\omega}_3$. But the fact that $I_2 = I_3$ implies that the body possesses an axis of symmetry. Therefore, $\boldsymbol{\omega}_1$ lies along the symmetry axis, and $\boldsymbol{\omega}_2$ and $\boldsymbol{\omega}_3$ are required only to lie in the plane perpendicular to $\boldsymbol{\omega}_1$. Consequently, there is no loss of generality if we also choose $\boldsymbol{\omega}_2 \perp \boldsymbol{\omega}_3$. Thus, the principal axes for a rigid body with an axis of symmetry can also be chosen to be an orthogonal set.

We have previously shown that the principal moments of inertia are obtained as the roots of the secular equation—a cubic equation. Mathematically, at least one of the roots of a cubic equation must be real, but there may be two imaginary roots. If the diagonalization procedures for the inertia tensor are to be physically meaningful, we must always obtain only real values for the principal moments. We can show in the following way that this is a general result. First, we assume the roots to be complex and use a procedure similar to that used in the preceding proof. But now we must also allow the quantities ω_{km} to become complex. There is no mathematical reason why we cannot do this, and we are not concerned with any physical interpretation of these quantities. We therefore write Equation 11.81a as before, but we take the complex conjugate of Equation 11.81b:

$$\left.\begin{aligned}\sum_k I_{ik}\omega_{km} &= I_m\omega_{im}\\ \sum_i I^*_{ki}\,\omega^*_{in} &= I^*_n\omega^*_{kn}\end{aligned}\right\} \tag{11.87}$$

Next, we multiply the first of these equations by ω^*_{in} and sum over i and multiply the second by ω_{km} and sum over k. The inertia tensor is symmetrical, and its elements are all real, so that $I_{ik} = I_{ki}^*$ Therefore, subtracting the second of these equations from the first, we find

$$(I_m - I^*_n)\sum_l \omega_{lm}\omega^*_{ln} = 0 \tag{11.88}$$

For the case m= n, we have

$$(I_m - I^*_m)\sum_l \omega_{lm}\omega^*_{lm} = 0 \tag{11.89}$$

The sum is just the definition of the scalar product of $\boldsymbol{\omega}_m$ and $\boldsymbol{\omega}^*_m$:

$$\boldsymbol{\omega}_m \cdot \boldsymbol{\omega}^*_m = |\boldsymbol{\omega}_m|^2 \geq 0 \tag{11.90}$$

Therefore, because the squared magnitude of $\boldsymbol{\omega}_m$ is in general positive, it must be true that I_m-I_m^* for Equation 11.89 to be satisfied. If a quantity and its complex conjugate are equal, then the imaginary parts must vanish identically. Thus, the principal moments of inertia are all real. Because {**I**} is real, the vectors $\boldsymbol{\omega}_m$ must also be real.

If m ≠ n in Equation 11.88 and if $I_m \neq I_n$, then the equation can be satisfied only if $\boldsymbol{\omega}_m \cdot \boldsymbol{\omega}_n = 0$; that is, these vectors are orthogonal, as before.

In all the proofs carried out in this section, we have referred to the inertia tensor. But examining these proofs reveals that the only properties of the inertia tensor that have actually been used are the facts that the tensor is symmetrical and that the elements are real. We may therefore conclude that any real, symmetric tensor has the following properties:

1. Diagonalization may be accomplished by an appropriate rotation of axes, that is, a similarity transformation.
2. The eigenvalues are obtained as roots of the secular determinant and are real.
3. The eigenvectors are real and orthogonal.

11.8 Eulerian Angles

The transformation from one coordinate system to another can be represented by a matrix equation of the form

$$\mathbf{x} = \boldsymbol{\lambda}\mathbf{x}'$$

If we identify the fixed system with $\mathbf{x}'$ and the body system with $\mathbf{x}$, then the rotation matrix $\boldsymbol{\lambda}$ completely describes the relative orientation of the two systems. The rotation matrix $\boldsymbol{\lambda}$ contains three independent angles. There are many possible choices for these angles; we find it convenient to use the Eulerian angles ϕ, θ, and ψ.

The Eulerian angles are generated in the following series of rotations, which takes the x_i' system into the x_i system.

1. The first rotation is counterclockwise through an angle ϕ about the x_3'-axis (Figure 11-9a) to transform the x_i' into the x_i''. Because the rotation takes place in the x_1'-x_2' plane, the transformation matrix is

$$\boldsymbol{\lambda}_\phi = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{11.91}$$

and

$$\mathbf{x}'' = \boldsymbol{\lambda}_\phi \mathbf{x}' \tag{11.92}$$

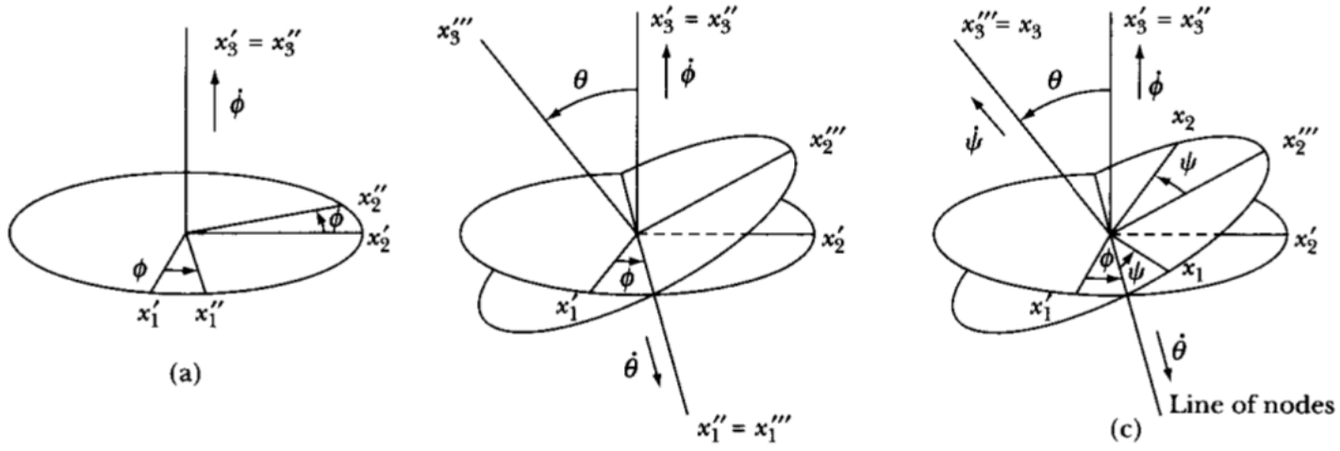


FIGURE 11-9 The Eulerian angles are used to rotate from the x_i' system to the x_i system. (a) First rotation is counterclockwise through an angle ϕ about the x_3'-axis. (b) Second rotation is counterclockwise through an angle θ about the x_1''-axis. (c) Third rotation is counterclockwise through an angle ψ about the x_3'''-axis.

2. The second rotation is counterclockwise through an angle θ about the x_1''-axis (Figure 11-9b) to transform the x_i'' into the x_1'''. Because the rotation is now in the x_2''-x_3'' plane, the transformation matrix is

$$\boldsymbol{\lambda}_\theta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \tag{11.93}$$

and

$$\mathbf{x}''' = \boldsymbol{\lambda}_\theta \mathbf{x}'' \tag{11.94}$$

3. The third rotation is counterclockwise through an angle ψ about the x'''_3-axis (Figure 11-9c) to transform the x'''_i into the x_i. The transformation matrix is

$$\boldsymbol{\lambda}_\psi = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{11.95}$$

and

$$\mathbf{x} = \boldsymbol{\lambda}_\psi \mathbf{x}''' \tag{11.96}$$

The line common to the planes containing the x_1- and x_2-axes and the x'_1- and x'_2-axes is called the **line of nodes**. The complete transformation from the x'_i system to the x_i system is given by

$$\begin{aligned} \mathbf{x} &= \boldsymbol{\lambda}_\psi \mathbf{x}''' = \boldsymbol{\lambda}_\psi \boldsymbol{\lambda}_\theta \mathbf{x}'' \\ &= \boldsymbol{\lambda}_\psi \boldsymbol{\lambda}_\theta \boldsymbol{\lambda}_\phi \mathbf{x}' \end{aligned} \tag{11.97}$$

and the rotation matrix $\boldsymbol{\lambda}$ is

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}_\psi \boldsymbol{\lambda}_\theta \boldsymbol{\lambda}_\phi \tag{11.98}$$

The components of this matrix are

$$\left.\begin{aligned} \lambda_{11} &= \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi \\ \lambda_{21} &= -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi \\ \lambda_{31} &= \sin\theta\sin\phi \\ \lambda_{12} &= \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi \\ \lambda_{22} &= -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi \\ \lambda_{32} &= -\sin\theta\cos\phi \\ \lambda_{13} &= \sin\psi\sin\theta \\ \lambda_{23} &= \cos\psi\sin\theta \\ \lambda_{33} &= \cos\theta \end{aligned}\right\} \tag{11.99}$$

(The components λ_{ij} are offset in the preceding equation to assist in the visualization of the complete $\boldsymbol{\lambda}$ matrix.)

Because we can associate a vector with an infinitesimal rotation, we can associate the time derivatives of these rotation angles with the components of the angular velocity vector $\boldsymbol{\omega}$. Thus,

$$\left.\begin{aligned} \omega_\phi &= \dot{\phi} \\ \omega_\theta &= \dot{\theta} \\ \omega_\psi &= \dot{\psi} \end{aligned}\right\} \tag{11.100}$$

The rigid-body equations of motion are most conveniently expressed in the body coordinate system (i.e., the x_i system), and therefore we must express the components of $\boldsymbol{\omega}$ in this system. We note that in Figure 11-9 the angular velocities $\dot{\phi}$, $\dot{\theta}$, and $\dot{\psi}$ are directed along the following axes:

$\dot{\boldsymbol{\phi}}$ along the x_3'- (fixed) axis

$\dot{\boldsymbol{\theta}}$ along the line of nodes

$\dot{\boldsymbol{\psi}}$ along the x_3- (body) axis

The components of these angular velocities along the body coordinate axes are

$$\left.\begin{aligned} \dot{\phi}_1 &= \dot{\phi} \sin\theta \sin\psi \\ \dot{\phi}_2 &= \dot{\phi} \sin\theta \cos\psi \\ \dot{\phi}_3 &= \dot{\phi} \cos\theta \end{aligned}\right\} \tag{11.101a}$$

$$\left.\begin{aligned} \dot{\theta}_1 &= \dot{\theta} \cos\psi \\ \dot{\theta}_2 &= -\dot{\theta} \sin\psi \\ \dot{\theta}_3 &= 0 \end{aligned}\right\} \tag{11.101b}$$

$$\left.\begin{aligned} \dot{\psi}_1 &= 0 \\ \dot{\psi}_2 &= 0 \\ \dot{\psi}_3 &= \dot{\psi} \end{aligned}\right\} \tag{11.101c}$$

Collecting the individual components of $\boldsymbol{\omega}$, we have, finally,

$$\begin{aligned}
\omega_1 &= \dot{\phi}_1 + \dot{\theta}_1 + \dot{\psi}_1 = \dot{\phi}\sin\theta\sin\psi + \dot{\theta}\cos\psi \\
\omega_2 &= \dot{\phi}_2 + \dot{\theta}_2 + \dot{\psi}_2 = \dot{\phi}\sin\theta\cos\psi - \dot{\theta}\sin\psi \\
\omega_3 &= \dot{\phi}_3 + \dot{\theta}_3 + \dot{\psi}_3 = \dot{\phi}\cos\theta + \dot{\psi}
\end{aligned} \tag{11.102}$$

These relations will be of use later in expressing the components of the angular momentum in the body coordinate system.

EXAMPLE 11.9

Using the Eulerian angles, find the transformation that moves the original x'_1-axis to the x'_2-x'_3 plane halfway between x'_2 and x'_3 and moves x'_2 perpendicular to the x'_2-x'_3 plane (Figure 11-10).

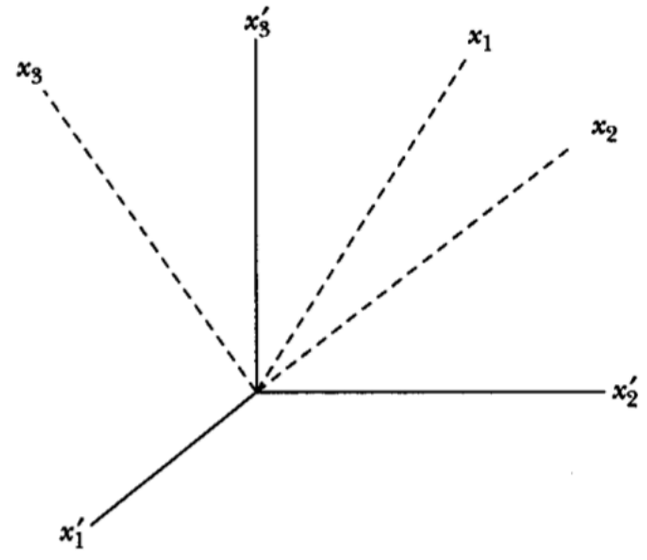


FIGURE 11-10 Example 11.9. We use Eulerian angles to rotate the x'_i system into the x_i system.

Solution. The key to transformations using Eulerian angles is the second rotation about the line of nodes, because this single rotation must move x'_3 to x_3. From the statement of the problem, x_3 must be in the x'_2-x'_3 plane, rotated 45° from x'_3. The first rotation must move x'_1 to x''_1 to have the correct position to rotate $x'_3 = x''_3$ to $x'''_3 = x_3$.

In this case, $x'_3 = x''_3$ is rotated $\theta = 45°$ about the original $x'_1 = x''_1$-axis so that $\phi = 0$ and

$$\boldsymbol{\lambda}_\phi = \mathbf{1} \tag{11.103}$$

$$\boldsymbol{\lambda}_\theta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \tag{11.104}$$

The last rotation, $\psi = 90°$, moves $x'_1 = x''_1 = x'''_1$ to x_1 to the position desired in the original x_2-x_3 plane.

$$\boldsymbol{\lambda}_\psi = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{11.105}$$

The transformation matrix $\boldsymbol{\lambda}$ is $\boldsymbol{\lambda} = \boldsymbol{\lambda}_\psi \boldsymbol{\lambda}_\theta \boldsymbol{\lambda}_\phi = \boldsymbol{\lambda}_\psi \boldsymbol{\lambda}_\theta$:

$$\boldsymbol{\lambda} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

$$\boldsymbol{\lambda} = \begin{pmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ -1 & 0 & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \tag{11.106}$$

Direction comparison between the x_i- and x'_i-axes shows that $\boldsymbol{\lambda}$ represents a single rotation describing the transformation.

11.9 Euler's Equations for a Rigid Body

Let us first consider the force-free motion of a rigid body. In such a case, the potential energy U vanishes and the Lagrangian L becomes identical with the rotational kinetic energy T. If we choose the x_i-axes to correspond to the principal axes of the body, then from Equation 11.35 we have

$$T = \frac{1}{2}\sum_i I_i \omega_i^2 \tag{11.107}$$

If we choose the Eulerian angles as the generalized coordinates, then Lagrange's equation for the coordinate ψ is

$$\frac{\partial T}{\partial \psi} - \frac{d}{dt}\frac{\partial T}{\partial \dot{\psi}} = 0 \tag{11.108}$$

which can be expressed as

$$\sum_i \frac{\partial T}{\partial \omega_i}\frac{\partial \omega_i}{\partial \psi} - \frac{d}{dt}\sum_i \frac{\partial T}{\partial \omega_i}\frac{\partial \omega_i}{\partial \dot{\psi}} = 0 \tag{11.109}$$

If we differentiate the components of $\boldsymbol{\omega}$ (Equation 11.102) with respect to ψ and $\dot{\psi}$, we have

$$\left.\begin{aligned} \frac{\partial \omega_1}{\partial \psi} &= \dot{\phi}\sin\theta\cos\psi - \dot{\theta}\sin\psi = \omega_2 \\ \frac{\partial \omega_2}{\partial \psi} &= -\dot{\phi}\sin\theta\sin\psi - \dot{\theta}\cos\psi = -\omega_1 \\ \frac{\partial \omega_3}{\partial \psi} &= 0 \end{aligned}\right\} \tag{11.110}$$

and

$$\left.\begin{aligned} \frac{\partial \omega_1}{\partial \dot{\psi}} &= \frac{\partial \omega_2}{\partial \dot{\psi}} = 0 \\ \frac{\partial \omega_3}{\partial \dot{\psi}} &= 1 \end{aligned}\right\} \tag{11.111}$$

From Equation 11.107, we also have

$$\frac{\partial T}{\partial \omega_i} = I_i \omega_i \tag{11.112}$$

Equation 11.109 therefore becomes

$$I_1\omega_1\omega_2 + I_2\omega_2(-\omega_1) - \frac{d}{dt} I_3\omega_3 = 0$$

or

$$(I_1 - I_2)\omega_1\omega_2 - I_3\dot{\omega}_3 = 0 \tag{11.113}$$

Because the designation of any particular principal axis as the x_3-axis is entirely arbitrary, Equation 11.113 can be permuted to obtain relations for $\dot{\omega}_1$ and $\dot{\omega}_2$:

$$\left.\begin{aligned}(I_2 - I_3)\omega_2\omega_3 - I_1\dot{\omega}_1 &= 0\\ (I_3 - I_1)\omega_3\omega_1 - I_2\dot{\omega}_2 &= 0\\ (I_1 - I_2)\omega_1\omega_2 - I_3\dot{\omega}_3 &= 0\end{aligned}\right\} \tag{11.114}$$

Equations 11.114 are called Euler's equations for force-free motion. It must be noted that, although Equation 11.113 for $\dot{\omega}_3$ is indeed the Lagrange equation for the coordinate ψ , the Euler equations for $\dot{\omega}_1$ and $\dot{\omega}_2$ are not the Lagrange equations for θ and ϕ .

To obtain Euler's equations for motion in a force field, we may start with the fundamental relation (see Equation 2.83) for the torque **N**:

$$\left(\frac{d\mathbf{L}}{dt}\right)_{\text{fixed}} = \mathbf{N} \tag{11.115}$$

where the designation "fixed" has been explicitly appended to $\dot{\mathbf{L}}$ because this relation is derived from Newton's equation and is therefore valid only in an inertial frame of reference. From Equation 10.12 we have

$$\left(\frac{d\mathbf{L}}{dt}\right)_{\text{fixed}} = \left(\frac{d\mathbf{L}}{dt}\right)_{\text{body}} + \boldsymbol{\omega} \times \mathbf{L} \tag{11.116}$$

or

$$\left(\frac{d\mathbf{L}}{dt}\right)_{\text{body}} + \boldsymbol{\omega} \times \mathbf{L} = \mathbf{N} \tag{11.117}$$

The component of this equation along the x_3-axis (note that this is a body axis) is

$$\dot{L}_3 + \omega_1 L_2 - \omega_2 L_1 = N_3 \tag{11.118}$$

But because we have chosen the x_i-axes to coincide with the principal axes of the body, we have, from Equation 11.34,

$$L_i = I_i \omega_i$$

so that

$$I_3\dot{\omega}_3 - (I_1 - I_2)\omega_1\omega_2 = N_3 \tag{11.119}$$

By permuting the subscripts, we can write all three components of **N**:

$$\left.\begin{aligned} I_1\dot{\omega}_1 - (I_2 - I_3)\omega_2\omega_3 &= N_1 \\ I_2\dot{\omega}_2 - (I_3 - I_1)\omega_3\omega_1 &= N_2 \\ I_3\dot{\omega}_3 - (I_1 - I_2)\omega_1\omega_2 &= N_3 \end{aligned}\right\} \tag{11.120}$$

Using the permutation symbol, we can write, in general

$$\boxed{(I_i - I_j)\omega_i\omega_j - \sum_k (I_k\dot{\omega}_k - N_k)\varepsilon_{ijk} = 0} \tag{11.121}$$

Equations 11.120 and 11.121 are the desired Euler equations for the motion of a rigid body in a force field.

The motion of a rigid body depends on the structure of the body only through the three numbers I_1, I_2, and I_3 - that is, the principal moments of inertia. Thus, any two bodies with the same principal moments move in exactly the same manner, regardless of the fact that they may have quite different shapes. (However, effects such as frictional retardation may depend on the shape of a body.)

EXAMPLE 11.10

Consider the dumbbell of Section 11.4. Find the angular momentum of the system and the torque required to maintain the motion shown in Figures 11-4 and 11-11.

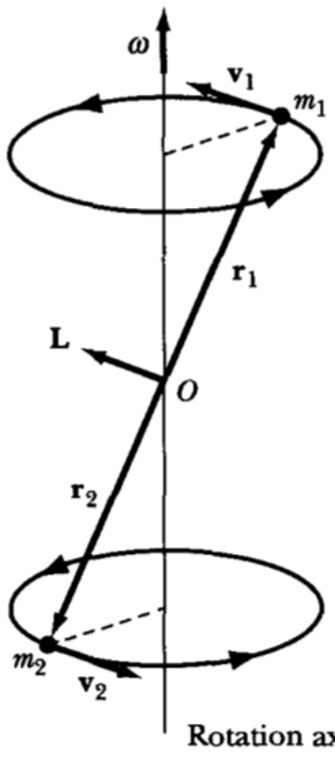


FIGURE 11-4 A dumbbell connected by masses m_1 and m_2 at the ends of its shaft. Note that $\boldsymbol{\omega}$ is not along the shaft, and that $\boldsymbol{\omega}$ and $\mathbf{L}$ are not collinear.

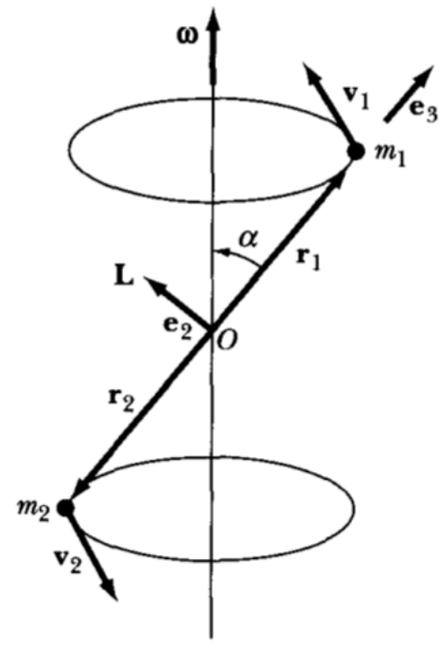


FIGURE 11-11 Example 11.10. The dumbbell with masses m_1 and m_2 at the ends of its shaft has its angular momentum $\mathbf{L}$ perpendicular to the shaft and $\mathbf{L}$ rotates around $\boldsymbol{\omega}$. The shaft maintains an angle α with $\boldsymbol{\omega}$.

Solution. Let $|\mathbf{r}_j| = |\mathbf{r}_2| = b$. Let the body fixed coordinate system have its origin at O and the symmetry axis x_3 be along the weightless shaft toward m_1.

$$\mathbf{L} = \sum_\alpha m_\alpha \mathbf{r}_\alpha \times \mathbf{v}_\alpha \tag{11.122}$$

Because $\mathbf{L}$ is perpendicular to the shaft and $\mathbf{L}$ rotates around $\boldsymbol{\omega}$ as the shaft rotates, let $\mathbf{e}_2$ be along $\mathbf{L}$:

$$\mathbf{L} = L_2 \mathbf{e}_2 \tag{11.123}$$

If α is the angle between $\boldsymbol{\omega}$ and the shaft, the components of $\boldsymbol{\omega}$ are

$$\left.\begin{aligned} \omega_1 &= 0 \\ \omega_2 &= \omega \sin\alpha \\ \omega_3 &= \omega \cos\alpha \end{aligned}\right\} \tag{11.124}$$

The principal axes are x_1, x_2, and x_3, and the principal moments of inertia are, from Equation 11.13a,

$$\left.\begin{aligned} I_1 &= (m_1 + m_2) b^2 \\ I_2 &= (m_1 + m_2) b^2 \\ I_3 &= 0 \end{aligned}\right\} \tag{11.125}$$

Combining Equations 11.124 and 11.125

$$\left.\begin{aligned} L_1 &= I_1 \omega_1 = 0 \\ L_2 &= I_2 \omega_2 = (m_1 + m_2) b^2 \omega \sin\alpha \\ L_3 &= I_3 \omega_3 = 0 \end{aligned}\right\} \tag{11.126}$$

which agrees with Equation 11.123.

Using Euler's equations (Equation 11.l20) and $\boldsymbol{\omega} = 0$, the torque components are

$$\left.\begin{aligned} N_1 &= -(m_1 + m_2) b^2 \omega^2 \sin\alpha \cos\alpha \\ N_2 &= 0 \\ N_3 &= 0 \end{aligned}\right\} \tag{11.127}$$

The torque required to maintain the motion if $\boldsymbol{\omega} = 0$ is directed along the x_1-axis.

11.10 Force-Free Motion of a Symmetric Top

If we consider a symmetric top, that is, a rigid body with $I_1 = I_2 \neq I_3$, then the force-free Euler equations (Equation 11.114) become

$$\left.\begin{aligned} (I_1 - I_3)\omega_2\omega_3 - I_1\dot{\omega}_1 &= 0 \\ (I_3 - I_1)\omega_3\omega_1 - I_1\dot{\omega}_2 &= 0 \\ I_3\dot{\omega}_3 &= 0 \end{aligned}\right\} \tag{11.128}$$

where I_1 has been substituted for I_2. Because for force-free motion the center of mass of the body is either at rest or in uniform motion with respect to the fixed or inertial frame of reference, we can, without loss of generality, specify that the body's center of mass is at rest and located at the origin of the fixed coordinate system. We consider the case in which the angular velocity vector $\boldsymbol{\omega}$ does not lie along a principal axis of the body, otherwise, the motion is trivial.

The first result for the motion follows from the third part of Equations 11.128, $\dot{\omega}_3 = 0$, or

$$\omega_3(t) = \text{const.} \tag{11.129}$$

The first two parts of Equation 11.128 can be written as

$$\left.\begin{aligned} \dot{\omega}_1 &= -\left(\frac{I_3 - I_1}{I_1}\omega_3\right)\omega_2 \\ \dot{\omega}_2 &= \left(\frac{I_3 - I_1}{I_1}\omega_3\right)\omega_1 \end{aligned}\right\} \tag{11.130}$$

Because the terms in the parentheses are identical and composed of constants, we may define

$$\Omega \equiv \frac{I_3 - I_1}{I_1}\omega_3 \tag{11.131}$$

so that

$$\left.\begin{aligned}\dot{\omega}_1 + \Omega\omega_2 &= 0\\ \dot{\omega}_2 - \Omega\omega_1 &= 0\end{aligned}\right\} \tag{11.132}$$

These are coupled equations of familiar form, and we can effect a solution by multiplying the second equation by i and adding to the first:

$$(\dot{\omega}_1 + i\dot{\omega}_2) - i\Omega(\omega_1 + i\omega_2) = 0 \tag{11.133}$$

If we define

$$\eta \equiv \omega_1 + i\omega_2 \tag{11.134}$$

then

$$\dot{\eta} - i\Omega\eta = 0 \tag{11.135}$$

with solution

$$\eta(t) = Ae^{i\Omega t} \tag{11.136}$$

Thus,

$$\omega_1 + i\omega_2 = A\cos\Omega t + iA\sin\Omega t \tag{11.137}$$

and therefore

$$\left.\begin{aligned}\omega_1(t) &= A\cos\Omega t\\ \omega_2(t) &= A\sin\Omega t\end{aligned}\right\} \tag{11.138}$$

Because ω_3 = constant, we note that the magnitude of to is also constant:

$$|\boldsymbol{\omega}| = \omega = \sqrt{\omega_1^2 + \omega_2^2 + \omega_3^2} = \sqrt{A^2 + \omega_3^2} = \text{constant} \tag{11.139}$$

Equations 11.138 are the parametric equations of a circle, so the projection of the vector $\boldsymbol{\omega}$ (which is of constant magnitude) onto the x_1-x_2 plane describes a circle with time (Figure 11-12).

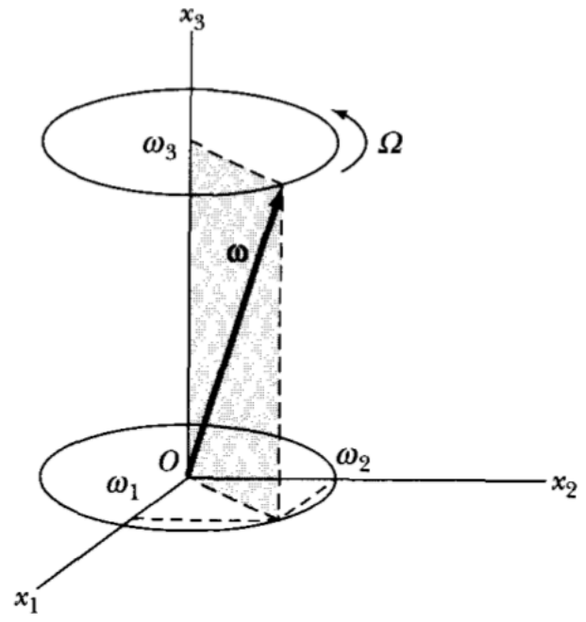


FIGURE 11-12 The angular velocity $\boldsymbol{\omega}$ of a force free symmetric top precesses with constant angular velocity Ω about the symmetric x_3-axis of the body. Thus $\boldsymbol{\omega}$ traces out a cone around the body symmetric axis.

The x_3-axis is the symmetry axis of the body, so we find that the angular velocity vector $\boldsymbol{\omega}$ revolves or *precesses* about the body x_3-axis with a constant angular frequency Ω. Thus, to an observer in the body coordinate system, $\boldsymbol{\omega}$ traces out a cone around the body symmetry axis, called the **body cone**.

Because we are considering force-free motion, the angular-momentum vector $\mathbf{L}$ is stationary in the fixed coordinate system and is constant in time. An additional constant of the motion for the force-free case is the kinetic energy, or in particular, because the body's center of mass is fixed, the *rotational* kinetic energy is constant:

$$T_{\text{rot}} = \frac{1}{2}\boldsymbol{\omega}\cdot\mathbf{L} = \text{constant} \tag{11.140}$$

But we have $\mathbf{L}$ = constant, so to must move such that its projection on the stationary angular-momentum vector is constant. Thus, $\boldsymbol{\omega}$ precesses around and makes a constant angle with the vector $\mathbf{L}$. In such a case, $\mathbf{L}$, $\boldsymbol{\omega}$, and the x_3- (body) axis (i.e., the unit vector $\mathbf{e}_3$) all lie in a *plane*. We can show this by proving that $\mathbf{L}\cdot(\boldsymbol{\omega}\times\mathbf{e}_3) = 0$.

First, $\boldsymbol{\omega} \times \mathbf{e}_3 = \omega_2 \mathbf{e}_1 - \omega_1 \mathbf{e}_2$. If we take the scalar product of this result with $\mathbf{L}$, we have $\mathbf{L} \cdot (\boldsymbol{\omega} \times \mathbf{e}_3) = I_1 \omega_2 \omega_3 - I_2 \omega_1 \omega_2$, because I_1=I_2 for the symmetric top. Therefore, if we designate the x_3-axis in the fixed coordinate system to coincide with $\mathbf{L}$, then to an observer in the fixed system, $\boldsymbol{\omega}$ traces out a cone around the fixed called the **space cone**. The situation is then described (Figure 11-13) by one cone rolling on another, such that $\boldsymbol{\omega}$ precesses around the x_3-axis in the body system and around the x_3-axis (or $\mathbf{L}$) in the space-fixed system.

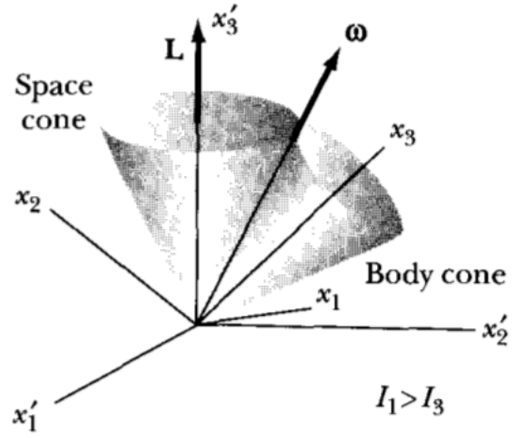


FIGURE 11-13 We let the angular momentum **L** be along the fixed x_3'-axis. The angular velocity **ω** traces out the **body cone** as it precesses about the x_3-axis in the body system, and it traces out the **space cone** as it precesses around the x_3'-axis in the space-fixed system. We can imagine the body cone rolling around the space cone.

The rate at which to precesses around the body symmetry axis is given by Equation 11.131:

$$\Omega = \frac{I_3 - I_1}{I_1} \omega_3$$

If $I_1 \simeq I_3$, then Ω becomes very small compared with ω_3 . Earth is slightly flattened near the poles, so its shape can be approximated by an oblate spheroid with $I_1 \simeq I_3$, but with $I_3 > I_1$. If Earth is considered to be a rigid body, then the moments I_1 and I_3 are such that $\Omega \simeq \omega_3/300$. Because the period of Earth's rotation is $2\pi/\omega$ = 1 day, and because $\omega_3 \simeq \omega$, the period predicted for the precession of the axis of rotation is $1/\Omega$ = 300 days.

The observed precession has an irregular period about 50 percent greater than that predicted on the basis of this simple theory; the deviation is ascribed to the facts that A) Earth is not a **rigid** body and B) the shape is not exactly that of an oblate spheroid, but rather has a higher-order deformation and actually resembles a flattened pear.

Earth's equatorial "bulge" together with the fact that Earth's rotational axis is inclined at an angle of approximately 23.5° to the plane of Earth's orbit around the sun (the plane of the ecliptic) produces a gravitational torque (caused by both the Sun and the Moon), which produces a slow precession of Earth's axis. The period of this precessional motion is approximately 26,000 years. Thus, in different epochs, different stars become the "pole star."

EXAMPLE 11.11

Show that the motion depicted in Figure 11-13 actually refers to the motion of a prolate object such as an elongated rod ($I_1 > I_3$), whereas for a flat disk ($I_3 > I_1$) the space cone would be inside the body cone rather than outside.

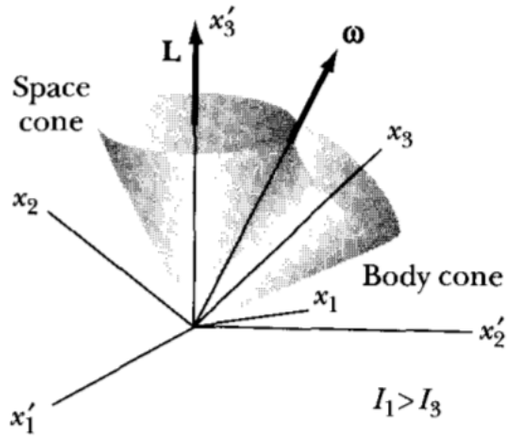


FIGURE 11-13 We let the angular momentum **L** be along the fixed x_3'-axis. The angular velocity **ω** traces out the **body cone** as it precesses about the x_3-axis in the body system, and it traces out the **space cone** as it precesses around the x_3'-axis in the space-fixed system. We can imagine the body cone rolling around the space cone.

Solution. If **L** is along x'_3, then the Euler angle θ (between the x_3- and x'_3-axes) is the angle between **L** and the x_3-axis. At a given instant, we align $\mathbf{e}_2$ to be in the plane defined by **L**, $\boldsymbol{\omega}$, and $\mathbf{e}_3$. Then, at this same instant.

$$\left.\begin{aligned} L_1 &= 0 \\ L_2 &= L\sin\theta \\ L_3 &= L\cos\theta \end{aligned}\right\} \tag{11.141}$$

Let α be the angle between to and the x_3-axis. Then, at this same instant, we have

$$\left.\begin{aligned} \omega_1 &= 0 \\ \omega_2 &= \omega\sin\alpha \\ \omega_3 &= \omega\cos\alpha \end{aligned}\right\} \tag{11.142}$$

We can also determine the components of $\mathbf{L}$ from Equation 11.34:

$$\left.\begin{aligned} L_1 &= I_1\omega_1 = 0 \\ L_2 &= I_1\omega_2 = I_1\omega\sin\alpha \\ L_3 &= I_3\omega_3 = I_3\omega\cos\alpha \end{aligned}\right\} \tag{11.143}$$

We can obtain the ratio L_2/L_3 from Equations 11.141 and 11.143,

$$\frac{L_2}{L_3} = \tan\theta = \frac{I_1}{I_3}\tan\alpha \tag{11.144}$$

so we have

Prolate spheroid

$$I_1 > I_3, \quad \theta > \alpha \tag{11.145a}$$

Oblate spheroid

$$I_3 > I_1, \quad \alpha > \theta \tag{11.145b}$$

The two cases are shown in Figure 11-14.

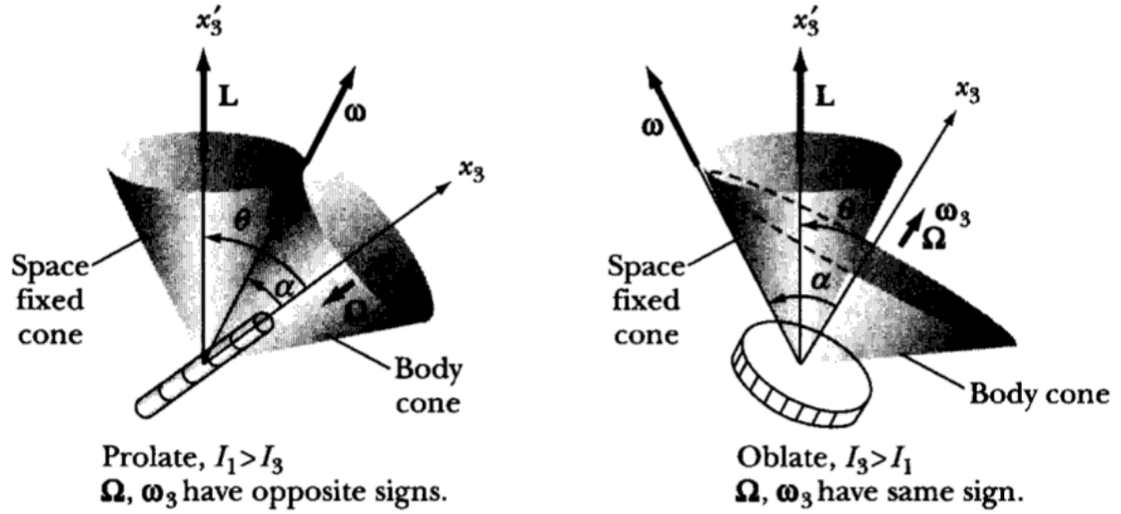


FIGURE 11-14 Example 11.11. (a) When the body is prolate ($I_1 > I_3$), we have the situation here and in Figure 11-13. (b) When the body is oblate ($I_3 > I_1$), the inside of the body cone rotates around the outside of the space cone. The space cone is at rest in either case.

From Equation 11.131, we determine that $\boldsymbol{\Omega}$ and ω_3 have the same sign if $I_3 > I_1$ but have opposite signs if $I_1 > I_3$. Thus, the sense of precession is opposite for the two cases. This fact and Equation 11.145 can be reconciled only if the space cone is outside the body cone for the prolate case but inside the body cone for the oblate case. The angular velocity to defines both cones as it rotates about **L** (space cone) and the symmetry axis x_3 (body cone). The line of contact between the space and body cones is the instantaneous axis of rotation (along $\boldsymbol{\omega}$). At any instant, this axis is at rest, so that the body cone rolls around the space cone without slipping. In both cases, the space cone is fixed, because **L** is constant.

EXAMPLE 11.12

With what angular velocity does the symmetry axis (x_3) and to rotate about the fixed angular momentum **L**?

Solution. Because $\mathbf{e}_3$, $\boldsymbol{\omega}$, and **L** are in the same plane, $\mathbf{e}_3$ and $\boldsymbol{\omega}$ precess about **L** with the same angular velocity.

In Section 11.8 we learned that $\dot{\phi}$ is the angular velocity along the x_3'-axis. If we use the same instant of time considered in the previous example (when $\mathbf{e}_2$ was in the plane of $\mathbf{e}_3$, $\boldsymbol{\omega}$, and $\mathbf{L}$), then the Euler angle $\psi = 0$, and from Equation 11.102

$$\omega_2 = \dot{\phi} \sin\theta$$

and

$$\dot{\phi} = \frac{\omega_2}{\sin\theta} \tag{11.146}$$

Substituting for ω_2 from Equation 11.142, we have

$$\dot{\phi} = \frac{\omega \sin\alpha}{\sin\theta} \tag{11.147}$$

We can rewrite $\dot{\phi}$ by substituting sin a from Equation 11.143 and $\sin\alpha$ from Equation 11.141:

$$\dot{\phi} = \omega \frac{L_2}{I_1\omega} \frac{L}{L_2} = \frac{L}{I_1} \tag{11.148}$$

11.11 Motion of a Symmetric Top with One Point Fixed

Consider a symmetric top with tip held fixed rotating in a gravitational field. In our previous development, we have been able to separate the kinetic energy into translational and rotational parts by taking the body's center of mass to be the origin of the rotating or body coordinate system. Alternatively, if we can choose the origins of the fixed and the body coordinate systems to coincide, then the translational kinetic energy vanishes, because $\mathbf{V} = \dot{\mathbf{R}} = 0$. Such a choice is quite convenient for discussing the top, because the stationary tip may then be taken as the origin for both coordinate systems. Figure 11-15 shows the Euler angles for this situation. The x_3'- (fixed) axis corresponds to the vertical, and we choose the x_3-(body) axis to be the symmetry axis of the top. The distance from the fixed tip to the center of mass is h, and the mass of the top is M.

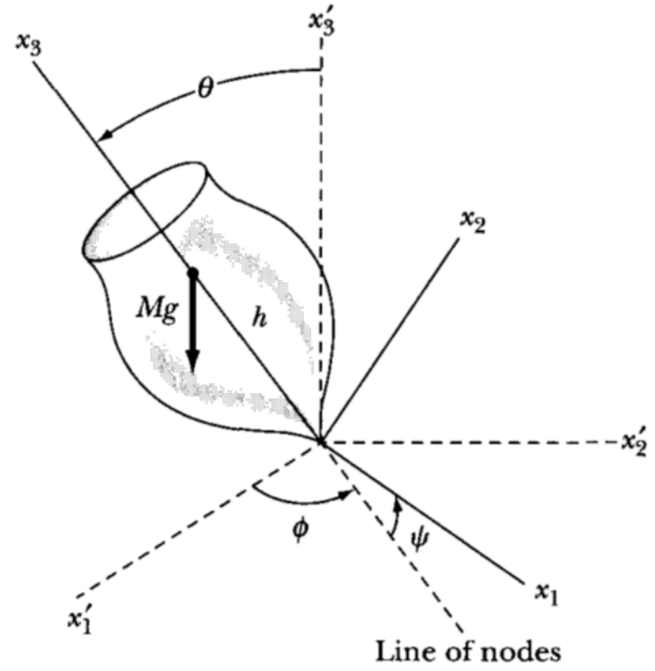


FIGURE 11-15 A symmetric top with its bottom tip fixed rotates in a gravitational field. The Euler angles relate the x_i'- (fixed) axes with the x_i- (body) axes. The angle ψ represents the rotation around the x_3 symmetry axis.

Because we have a symmetric top, the principal moments of inertia about the x_1- and x_2-axes are equal: $I_1 = I_2$. We assume $I_3 \neq I_1$ The kinetic energy is then given by

$$T = \frac{1}{2}\sum_i I_i \omega_i^2 = \frac{1}{2} I_1(\omega_1^2 + \omega_2^2) + \frac{1}{2} I_3 \omega_3^2 \tag{11.149}$$

According to Equation 11.102, we have

$$\begin{aligned}
\omega_1^2 &= (\dot{\phi} \sin\theta \sin\psi + \dot{\theta} \cos\psi)^2 \\
&= \dot{\phi}^2 \sin^2\theta \sin^2\psi + 2\dot{\phi}\dot{\theta} \sin\theta \sin\psi \cos\psi + \dot{\theta}^2 \cos^2\psi \\
\omega_2^2 &= (\dot{\phi} \sin\theta \cos\psi - \dot{\theta} \sin\psi)^2 \\
&= \dot{\phi}^2 \sin^2\theta \cos^2\psi - 2\dot{\phi}\dot{\theta} \sin\theta \sin\psi \cos\psi + \dot{\theta}^2 \sin^2\psi
\end{aligned}$$

so that

$$\omega_1^2 + \omega_2^2 = \dot{\phi}^2 \sin^2\theta + \dot{\theta}^2 \tag{11.150a}$$

and

$$\omega_3^2 = (\dot{\phi}\cos\theta + \dot{\psi})^2 \tag{11.150b}$$

Therefore,

$$T = \frac{1}{2} I_1(\dot{\phi}^2 \sin^2\theta + \dot{\theta}^2) + \frac{1}{2} I_3(\dot{\phi}\cos\theta + \dot{\psi})^2 \tag{11.151}$$

Because the potential energy is Mgh $\cos\theta$, the Lagrangian becomes

$$L = \frac{1}{2} I_1(\dot{\phi}^2 \sin^2\theta + \dot{\theta}^2) + \frac{1}{2} I_3(\dot{\phi}\cos\theta + \dot{\psi})^2 - Mgh\cos\theta \tag{11.152}$$

The Lagrangian is cyclic in both the ϕ- and ψ-coordinates. The momenta conjugate to these coordinates are therefore constants of the motion:

$$p_\phi = \frac{\partial L}{\partial \dot{\phi}} = (I_1 \sin^2\theta + I_3\cos^2\theta)\dot{\phi} + I_3\dot{\psi}\cos\theta = \text{constant} \tag{11.153}$$

$$p_\psi = \frac{\partial L}{\partial \dot{\psi}} = I_3(\dot{\psi} + \dot{\phi}\cos\theta) = \text{constant} \tag{11.154}$$

Because the cyclic coordinates are *angles*, the conjugate momenta are *angular momenta* - the angular momenta along the axes for which ϕ and ψ are the rotation angles, that is, the x_3'- (or vertical) axis and the x_3- (or body symmetry) axis, respectively. We note that this result is ensured by the construction shown in Figure 11-15, because the gravitational torque is directed along the line of nodes. Hence, the torque can have no component along either the x_3'- or the x_3-axis, both of which are perpendicular to the line of nodes. Thus, the angular momenta along these axes are constants of the motion.

Equations 11.153 and 11.154 can be solved for $\dot{\phi}$ and $\dot{\psi}$ in terms of θ . From Equation 11.154, we can write

$$\dot{\psi} = \frac{p_\psi - I_3 \dot{\phi} \cos\theta}{I_3} \tag{11.155}$$

and substituting this result into Equation 11.153, we find

$$(I_1 \sin^2\theta + I_3 \cos^2\theta)\dot{\phi} + (p_\psi - I_3\dot{\phi}\cos\theta)\cos\theta = p_\phi$$

or

$$(I_1 \sin^2\theta)\dot{\phi} + p_\psi \cos\theta = p_\phi$$

so that

$$\dot{\phi} = \frac{p_\phi - p_\psi \cos\theta}{I_1 \sin^2\theta} \tag{11.156}$$

Using this expression for $\dot{\phi}$ in Equation 11.155, we have

$$\dot{\psi} = \frac{p_\psi}{I_3} - \frac{(p_\phi - p_\psi \cos\theta)\cos\theta}{I_1 \sin^2\theta} \tag{11.157}$$

By hypothesis, the system we are considering is conservative; we therefore have the further property that the total energy is a constant of the motion:

$$E = \frac{1}{2} I_1(\dot{\phi}^2 \sin^2\theta + \dot{\theta}^2) + \frac{1}{2} I_3\omega_3^2 + Mgh\cos\theta = \text{constant} \tag{11.158}$$

Using the expression for ω_3 (e.g., see Equation 11.102), we note that Equation 11.154 can be written as

$$p_\psi = I_3\omega_3 = \text{constant} \tag{11.159a}$$

or

$$I_3\omega_3^2 = \frac{p_\psi^2}{I_3} = \text{constant} \tag{11.159b}$$

Therefore, not only is E a constant of the motion, but so is E - $\frac{1}{2}I_3\omega_3^2$; we let this quantity be E′:

$$E' \equiv E - \frac{1}{2} I_3\omega_3^2 = \frac{1}{2} I_1(\dot{\phi}^2 \sin^2\theta + \dot{\theta}^2) + Mgh\cos\theta = \text{constant} \tag{11.160}$$

Substituting into this equation the expression for $\dot{\phi}$ (Equation 11.156), we have

$$E' = \frac{1}{2} I_1\dot{\theta}^2 + \frac{(p_\phi - p_\psi\cos\theta)^2}{2I_1\sin^2\theta} + Mgh\cos\theta \tag{11.161}$$

which we can write as

$$E' = \frac{1}{2} I_1\dot{\theta}^2 + V(\theta) \tag{11.162}$$

where V(θ) is an "effective potential" given by

$$V(\theta) \equiv \frac{(p_\phi - p_\psi\cos\theta)^2}{2I_1\sin^2\theta} + Mgh\cos\theta \tag{11.163}$$

Equation 11.162 can be solved to yield t(θ):

$$t(\theta) = \int \frac{d\theta}{\sqrt{(2/I_1)[E' - V(\theta)]}} \tag{11.164}$$

This integral can (formally, at least) be inverted to obtain θ(t), which, in turn, can be substituted into Equations 11.156 and 11.157 to yield ϕ(t) and ψ(t). Because the Euler angles θ , ϕ , ψ completely specify the orientation of the top, the results for θ (t), ϕ(t)), and ψ(t) constitute a complete solution for the problem. It should be clear that such a procedure is complicated and not very illuminating. But we can obtain some qualitative features of the motion by examining the preceding equations in a manner analogous to that used for treating the motion of a particle in a central-force field (see Section 8.6).

Figure 11-16 shows the form of the effective potential V(θ) in the range $0 \leq \theta \leq \pi$, which clearly is the physically limited region for θ .

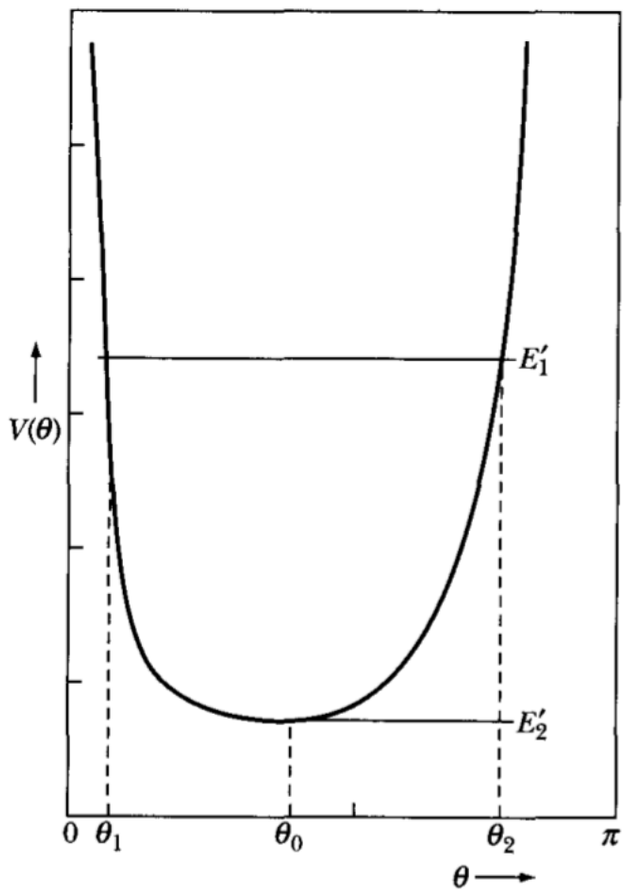


FIGURE 11-16 The effective potential $V(\theta)$ for the rotating top of Figure 11-15 is plotted versus the angle θ. We can study the angular limits of the inclination of the top by knowing the modified energy E'.

This energy diagram indicates that for any general values of E' (e.g., the value represented by E'_1) the motion is limited by two extreme values of θ - that is, θ_1 and θ_2 . which correspond to the turning points of the central-force problem and are roots of the denominator in Equation 11.164. Thus we find that the inclination of the rotating top is. in general, confined to the region $\theta_1 \leq \theta \leq \theta_2$. For the case that $E' = E'_2 = V_{min}$, θ is limited to the single value θ_0 , and the motion is a steady precession at a fixed angle of inclination. Such motion is similar to the occurrence of circular orbits in the central-force problem.

The value of θ_0 can be obtained by setting the derivative of V(θ) equal to zero. Thus,

$$\left.\frac{\partial V}{\partial \theta}\right|_{\theta=\theta_0} = \frac{-\cos\theta_0(p_\phi - p_\psi \cos\theta_0)^2 + p_\psi \sin^2\theta_0(p_\phi - p_\psi \cos\theta_0)}{I_1 \sin^3\theta_0} - Mgh \sin\theta_0 = 0 \tag{11.165}$$

If we define

$$\beta \equiv p_\phi - p_\psi \cos\theta_0 \tag{11.166}$$

then Equation 11.165 becomes

$$(\cos\theta_0)\beta^2 - (p_\psi \sin^2\theta_0)\beta + (MghI_1 \sin^4\theta_0) = 0 \tag{11.167}$$

This is a quadratic in β and can be solved with the result

$$\beta = \frac{p_\psi \sin^2\theta_0}{2\cos\theta_0}\left(1 \pm \sqrt{1 - \frac{4MghI_1\cos\theta_0}{p_\psi^2}}\right) \tag{11.168}$$

Because β must be a real quantity, the radicand in Equation 11.168 must be positive. If $\theta_0 < \pi/2$, we have

$$p_\psi^2 \geq 4MghI_1 \cos\theta_0 \tag{11.169}$$

But from Equation 11.159a, $p_\psi = I_3\omega_3$; thus,

$$\omega_3 \geq \frac{2}{I_3}\sqrt{MghI_1\cos\theta_0} \tag{11.170}$$

We therefore conclude that a steady precession can occur at the fixed angle of inclination θ_0 only if the angular velocity of spin is larger than the limiting value given by Equation 11.170

From Equation 11.156, we note that we can write (for $\theta = \theta_0$)

$$\dot{\phi}_0 = \frac{\beta}{I_1 \sin^2 \theta_0} \tag{11.171}$$

We therefore have two possible values of the precessional angular velocity $\dot{\phi}_0$, one for each of the values of β given by Equation 11.168:

$$\dot{\phi}_{0(+)} \rightarrow \text{Fast precession}$$

and

$$\dot{\phi}_{0(-)} \rightarrow \text{Slow precession}$$

If ω_3 (or p_ψ) is large (a fast top), then the second term in the radicand of Equation 11.168 is small, and we may expand the radical. Retaining only the first nonvanishing term in each case, we find

$$\left.\begin{aligned} \dot{\phi}_{0(+)} &\cong \frac{I_3\omega_3}{I_1 \cos \theta_0} \\ \dot{\phi}_{0(-)} &\cong \frac{Mgh}{I_3\omega_3} \end{aligned}\right\} \tag{11.172}$$

It is the slower of the two possible precessional angular velocities, $\dot{\phi}_{0(-)}$ that is usually observed.

The preceding results apply if $\theta_0 < \pi/2$; but if $\theta_0 > \pi/2$, the radicand in Equation 11.168 is always positive and there is no limiting condition on ω_3. Because the radical is greater than unity in such a case, the values of $\dot{\phi}_0$ for fast and slow precession have opposite signs; that is, for $\theta_0 < \pi/2$, the fast precession is in the same direction as that for $\theta_0 > \pi/2$, but the slow precession takes place in the opposite sense.

If $\theta_0 > \pi/2$, the fixed tip of the top is at a position above the center of mass. Such motion is possible, for example, with a gyroscopic top whose tip is actually a ball and rests in a cup that is fixed atop a pedestal.

For the general case, in which $\theta_1 < \theta \leq \theta_2$, Equation 11.156 indicates that $\dot{\phi}$ may or may not change sign as θ varies between its limits - depending on the values of p_ϕ and p_ψ . If $\dot{\phi}$ does not change sign, the top precesses monotonically around the x_3'-axis (see Figure 11-15), and the x_3- (or symmetry) axis oscillates between $\theta = \theta_1$ and $\theta = \theta_2$. This phenomenon is called **nutation**; the path described by the projection of the body symmetry axis on a unit sphere in the fixed system is shown in Figure 11-17a.

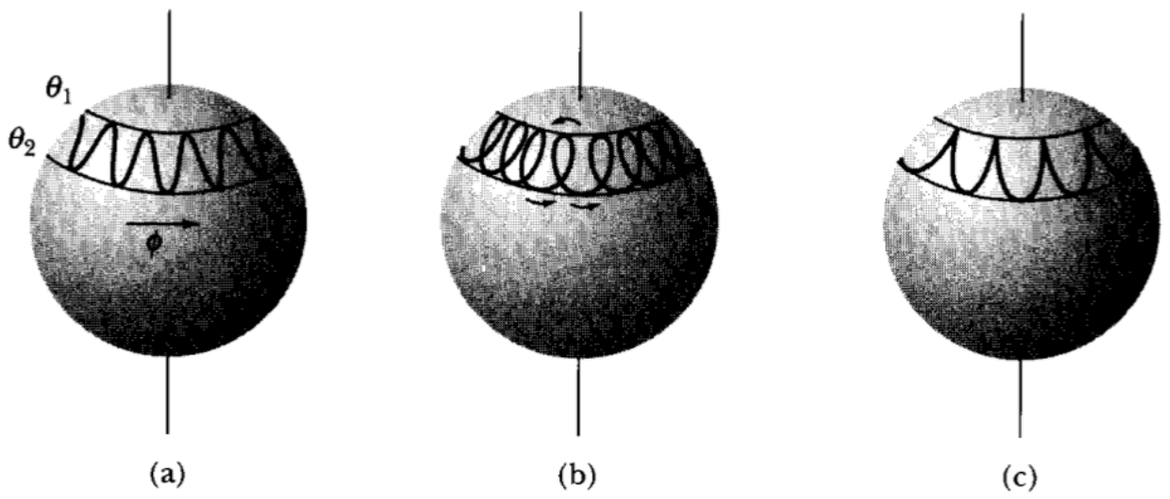


FIGURE 11-17 The rotating top also nutates between the limit angles θ_1 and θ_2. In (a) $\dot{\phi}$ does not change sign. In (b) $\dot{\phi}$ does change sign, and we see looping motion. In (c) the initial conditions include $\dot{\theta} = \dot{\phi} = 0$; this is the normal cusp-like motion when we spin a top and release it.

If $\dot{\phi}$ does change sign between the limiting values of θ , the precessional angular velocity must have opposite signs at $\theta = \theta_1$ and $\theta = \theta_2$. Thus, the nutational-precessional motion produces the looping motion of the symmetry axis depicted in Figure 11-17b.

Finally, if the values of p_ϕ and p_ψ are such that

$$(p_\phi - p_\psi \cos\theta)|_{\theta=\theta_1} = 0 \tag{11.173}$$

then

$$\dot{\phi}|_{\theta=\theta_1} = 0, \qquad \dot{\theta}|_{\theta=\theta_1} = 0 \tag{11.174}$$

Figure 11-17c shows the resulting cusplike motion. It is just this case that corresponds to the usual method of starting a top. First, the top is spun around its axis, then it is given a certain initial tilt and released. Thus, initial conditions are $\theta = \theta_1$ and $\dot{\theta} = 0 = \dot{\phi}$. Because the first motion of the top is to begin to fall in the gravitational field, the conditions are exactly those of Figure 11-17c, and the cusplike motion ensues. Figures 11-17a and 11-17b correspond to the motion in the event that there is an initial angular velocity $\dot{\phi}$ either in the direction of or opposite to the direction of precession.

11.12 Stability of Rigid-Body Rotations

We now consider a rigid body undergoing force-free rotation around one of its principal axes and inquire whether such motion is stable. "Stability" here means, as before (see Section 8.10), that if a small perturbation is applied to the system, the motion will either return to its former mode or will perform small oscillations about it.

We choose for our discussion a general rigid body for which all the principal moments of inertia are distinct, and we label them such that $I_3 > I_2 > I_1$. We let the body axes coincide with the principal axes, and we start with the body rotating around the x_1-axis - that is, around the principal axis associated with the moment of inertia I_1. Then,

$$\boldsymbol{\omega} = \omega_1 \mathbf{e}_1 \tag{11.175}$$

If we apply a small perturbation, the angular velocity vector assumes the form

$$\boldsymbol{\omega} = \omega_1 \mathbf{e}_1 + \lambda \mathbf{e}_2 + \mu \mathbf{e}_3 \tag{11.176}$$

where λ and μ are small quantities and correspond to the parameters used previously in other perturbation expansions. (λ and μ are sufficiently small so that their product can be neglected compared with all other quantities of interest to the discussion.)

The Euler equations (see Equation 11.114) become

$$\left.\begin{aligned}(I_2 - I_3)\lambda\mu - I_1\dot{\omega}_1 &= 0\\ (I_3 - I_1)\mu\omega_1 - I_2\dot{\lambda} &= 0\\ (I_1 - I_2)\lambda\omega_1 - I_3\dot{\mu} &= 0\end{aligned}\right\} \tag{11.177}$$

Because $\lambda\mu \approx 0$, the first of these equations requires $\dot{\omega}_1 = 0$, or ω_1 = constant.

Solving the other two equations for $\dot{\lambda}$ and $\dot{\mu}$, we find

$$\dot{\lambda} = \left(\frac{I_3 - I_1}{I_2}\,\omega_1\right)\mu \tag{11.178}$$

$$\dot{\mu} = \left(\frac{I_1 - I_2}{I_3}\,\omega_1\right)\lambda \tag{11.179}$$

where the terms in parentheses are both constants. These are coupled equations, but they cannot be solved by the method used in Section 11.10, because the constants in the two equations are different. The solution can be obtained by first differentiating the equation for $\dot{\lambda}$:

$$\ddot{\lambda} = \left(\frac{I_3 - I_1}{I_2}\,\omega_1\right)\dot{\mu} \tag{11.180}$$

The expression for $\dot{\mu}$ can now be substituted in this equation:

$$\ddot{\lambda} + \left(\frac{(I_1 - I_3)(I_1 - I_2)}{I_2 I_3}\,\omega_1^2\right)\lambda = 0 \tag{11.181}$$

The solution to this equation is

$$\lambda(t) = Ae^{i\Omega_{1\lambda}t} + Be^{-i\Omega_{1\lambda}t} \tag{11.182}$$

where

$$\Omega_{1\lambda} \equiv \omega_1 \sqrt{\frac{(I_1 - I_3)(I_1 - I_2)}{I_2 I_3}} \tag{11.183}$$

and where the subscripts 1 and λ indicate that we are considering the solution for λ when the rotation is around the x_1-axis.

By hypothesis, /i < 4 and /j < 4, so ^^^ is real. The solution for A(^) therefore represents oscillatory motion with a frequency /2l;^. We can similarly investigate /a@, with the result that /2i^ = CId, = fli. Thus, the small perturbations introduced by forcing small X2- and Xj-components on to do not increase with time but oscillate around the equilibrium values A = 0 and /a = 0. Consequently, the rotation around the Xj-axis is stable.

If we consider rotations around the Xg- and X3-axes, we can obtain expressions for fl2 ^nd fl^ from Equation 11.183 by permutation:

$$\Omega_1 = \omega_1 \sqrt{\frac{(I_1 - I_3)(I_1 - I_2)}{I_2 I_3}} \tag{11.184a}$$

$$\Omega_2 = \omega_2 \sqrt{\frac{(I_2 - I_1)(I_2 - I_3)}{I_1 I_3}} \tag{11.184b}$$

$$\Omega_3 = \omega_3 \sqrt{\frac{(I_3 - I_2)(I_3 - I_1)}{I_1 I_2}} \tag{11.184c}$$

But because $I_1 < I_2 < I_3$, we have

$$\Omega_1, \Omega_3 \text{ real}, \qquad \Omega_2 \text{ imaginary}$$

Thus, when the rotation takes place around either the x_1- or x_3-axes, the perturbation produces oscillatory motion and the rotation is stable.

When the rotation takes place around x_2, however, the fact that $\Omega_2$is imaginary results in the perturbation increasing with time without limit; such motion is unstable.

Because we have assumed a completely arbitrary rigid body for this discussion, we conclude that rotation around the principal axis corresponding to either the greatest or smallest moment of inertia is stable and that rotation around the principal axis corresponding to the intermediate moment is unstable. We can demonstrate this effect with, say, a book (kept closed by tape or a rubber band). If we toss the book into the air with an angular velocity around one of the principal axes, the motion is unstable for rotation around the intermediate axis and stable for the other two axes.

If two of the moments of inertia are equal ($I_1 = I_2$, say), then the coefficient of λ in Equation 11.179 vanishes, and we have $\dot{\mu} = 0$ or μ = constant. Equation 11.178 for λ can therefore be integrated to yield

$$\lambda(t) = C + Dt \tag{11.185}$$

and the perturbation increases linearly with the time; the motion around the x_1-axis is therefore unstable. We find a similar result for motion around the x_2-axis. Stability exists only for the x_3-axis, independent of whether I_3 is greater or less than $I_1 = I_2$.

A good example of the stability of rotating objects is seen by the satellites put into space by the space shuttle orbiter When the satellites are ejected from the payload bay, they are normally spinning in a stable configuration. In May 1992, when the astronauts attempted to grab in space the *Intelsat* satellite (which originally had failed to go into its designed orbit) to attach a rocket that would insert it into geosynchronous orbit, the spinning satellite was slowed down and stopped before the astronaut attempted to attach a grappling fixture to bring it into the payload bay. After each futile attempt, when the grappling fixture failed, the satellite tumbled even more. After spending two unsuccessful days trying to attach the grappling fixture, the astronauts had to abort their attempts because of the increased tumbling.

Ground controllers required a few hours to restabilize the satellite using jet thrusters. The satellite was left in a stable configuration of spinning slowly about its cyclindrical symmetry axis (a principal axis) until the next recovery attempt. Finally, on the third day, three astronauts went outside the orbiter, grabbed the slightly rotating satellite, stopped it, and put it into the payload bay where the rocket skirt was attached. The *Intelsat* satellite was finally successfully placed into orbit in time to broadcast the 1992 Barcelona Olympic summer games.

CHAPTER 12 - Coupled Oscillations

12.1 Introduction

In Chapter 3, we examined the motion of an oscillator subjected to an external driving force. The discussion was limited to the case in which the driving force is periodic; that is, the driver is itself a harmonic oscillator We considered the action of the driver on the oscillator, but we did not include the feedback effect of the oscillator on the driver. In many instances, ignoring this effect is unimportant, but if two (or many) oscillators are connected in such a way that energy can be transferred back and forth between (or among) them, the situation becomes the more complicated case of **coupled oscillations.** Motion of this type can be quite complex (the motion may not even be periodic), but we can always describe the motion of any oscillatory system in terms of normal coordinates, which have the property that each oscillates with a single, well-defined frequency; that is, the normal coordinates are constructed in such a way that no coupling occurs among them, even though there is coupling among the ordinary (rectangular) coordinates describing the positions of particles. Initial conditions can always be prescribed for the system so that in the subsequent motion only one normal coordinate varies with time. In this circumstance, we say that one of the **normal modes** of the system has been excited. If the system has n degrees of freedom (e.g., n-coupled one-dimensional oscillators or n/3-coupled three-dimensional oscillators), there are in general n normal modes, some of which may be identical. The general motion of the system is a complicated superposition of all the normal modes of oscillation, but we can always find initial conditions such that any given one of the normal modes is independently excited. Identifying each of a system's normal modes allows us to construct a revealing picture of the motion, even though the system's *general* motion is a complicated combination of all the normal modes.

It is relatively easy to demonstrate some of the coupled oscillator phenomena described in this chapter. For example, two pendula coupled by a spring between their mass bobs, two pendula hung from a rope, and masses connected by springs can all be experimentally examined in the classroom. Similarly, the tri-atomic molecule discussed here is a reasonable description of CO_2. Similar models can approximate other molecules.

In the following chapter, we shall continue the development begun here and discuss the motion of vibrating strings. This example by no means exhausts the usefulness of the normal-mode approach to the description of oscillatory systems; indeed, applications can be found in many areas of mathematical physics, such as the microscopic motions in crystalline solids and the oscillations of the electromagnetic field.

12.2 Two Coupled Harmonic Oscillators

A physical example of a coupled system is a solid in which the atoms interact by elastic forces between each other and oscillate about their equilibrium positions. Springs between the atoms represent the elastic forces. A molecule composed of a few such interacting atoms would be an even simpler model. We begin by considering a similar system of coupled motion in one dimension: two masses connected by a spring to each other and by springs to fixed positions (Figure 12-1).

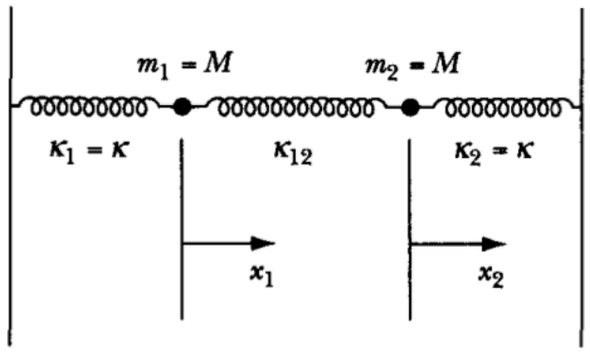


FIGURE 12-1 Two masses are connected by a spring to each other and by springs to fixed positions. This is a system of coupled motion in one dimension.

We return to this example throughout the chapter as we describe various instances of coupled motion.

We let each of the oscillator springs have a force constant κ : the force constant of the coupling spring is κ_{12} . We restrict the motion to the line connecting the masses, so the system has only two degrees of freedom, represented by the coordinates x_1 and x_2. Each coordinate is measured from the position of equilibrium.

Henceforth, we denote force constants by κ rather than by k as heretofore. The symbol k is reserved for (beginning in Chapter 13) an entirely different context.

If m_1 and m_2 are displaced from their equilibrium position by amounts x_1 and x_2, respectively, the force on m_1 is $-\kappa x_1 - \kappa_{12}(x_1 - x_2)$, and the force on m_2 is $-\kappa x_2 - \kappa_{12}(x_2 - x_1)$. Therefore the equations of motion are

$$\left.\begin{aligned} M\ddot{x}_1 + (\kappa + \kappa_{12})x_1 - \kappa_{12}x_2 &= 0 \\ M\ddot{x}_2 + (\kappa + \kappa_{12})x_2 - \kappa_{12}x_1 &= 0 \end{aligned}\right\} \tag{12.1}$$

Because we expect the motion to be oscillatory, we attempt a solution of the form

$$\left.\begin{aligned} x_1(t) &= B_1 e^{i\omega t} \\ x_2(t) &= B_2 e^{i\omega t} \end{aligned}\right\} \tag{12.2}$$

where the frequency ω is to be determined and where the amplitudes B_1 and B_2 may be complex. (Because a complex amplitude has a magnitude and a phase, we have the two arbitrary constants necessary in the solution of a second-order differential equation; that is, we could equally well write $x(t) = |B| \exp[i(\omega t - \delta)]$ or $x(t) = |B| \cos(\omega t - \delta)$, as in Equation 3.6b. Later (see Equation 12.9), we shall find it more convenient to use two distinct real amplitudes and the time-varying factors $\exp(i\omega t)$ and $\exp(-i\omega t)$. These various forms of solution are all entirely equivalent.) These trial solutions are complex functions. Thus, in the final step of the solution, the real parts of $x_1(t)$ and $x_2(t)$ will be taken, because the real part is all that is physically significant. We use this method of solution because of its great efficiency, and we use it again later, leaving out most of the details. Substituting these expressions for the displacements into the equations of motion, we find

$$\left.\begin{aligned} -M\omega^2 B_1 e^{i\omega t} + (\kappa + \kappa_{12})B_1 e^{i\omega t} - \kappa_{12}B_2 e^{i\omega t} &= 0 \\ -M\omega^2 B_2 e^{i\omega t} + (\kappa + \kappa_{12})B_2 e^{i\omega t} - \kappa_{12}B_1 e^{i\omega t} &= 0 \end{aligned}\right\} \tag{12.3}$$

Collecting terms and canceling the common exponential factor, we obtain

$$\left.\begin{aligned} (\kappa + \kappa_{12} - M\omega^2)B_1 - \kappa_{12}B_2 &= 0 \\ -\kappa_{12}B_1 + (\kappa + \kappa_{12} - M\omega^2)B_2 &= 0 \end{aligned}\right\} \tag{12.4}$$

For a nontrivial solution to exist for this pair of simultaneous equations, the determinant of the coefficients of B_1 and B_2 must vanish:

$$\begin{vmatrix} \kappa + \kappa_{12} - M\omega^2 & -\kappa_{12} \\ -\kappa_{12} & \kappa + \kappa_{12} - M\omega^2 \end{vmatrix} = 0 \tag{12.5}$$

The expansion of this secular determinant yields

$$(\kappa + \kappa_{12} - M\omega^2)^2 - \kappa_{12}^2 = 0 \tag{12.6}$$

Hence,

$$\kappa + \kappa_{12} - M\omega^2 = \pm \kappa_{12}$$

Solving for ω , we obtain

$$\omega = \sqrt{\frac{\kappa + \kappa_{12} \pm \kappa_{12}}{M}} \tag{12.7}$$

We therefore have two characteristic frequencies (or eigenfrequencies) for the system:

$$\omega_1 = \sqrt{\frac{\kappa + 2\kappa_{12}}{M}}, \qquad \omega_2 = \sqrt{\frac{\kappa}{M}} \tag{12.8}$$

Thus, the general solution to the problem is

$$\left.\begin{aligned} x_1(t) &= B_{11}^{+} e^{i\omega_1 t} + B_{11}^{-} e^{-i\omega_1 t} + B_{12}^{+} e^{i\omega_2 t} + B_{12}^{-} e^{-i\omega_2 t} \\ x_2(t) &= B_{21}^{+} e^{i\omega_1 t} + B_{21}^{-} e^{-i\omega_1 t} + B_{22}^{+} e^{i\omega_2 t} + B_{22}^{-} e^{-i\omega_2 t} \end{aligned}\right\} \tag{12.9}$$

where we have explicitly written both positive and negative frequencies, because the radicals in Equations 12.7 and 12.8 can carry either sign.

In Equation 12.9, the amplitudes are not all independent, as we may verify by substituting ω_1 and ω_2 into Equation 12.4. We find

$$\begin{aligned} &\text{for} \quad w = \omega_1: \quad B_{11} = -B_{21} \\ &\text{for} \quad w = \omega_2: \quad B_{12} = B_{22} \end{aligned}$$

The only subscripts necessary on the Bs are those indicating the particular eigenfrequency (i.e., the *second* subscripts). We can therefore write the general solution as

$$\left.\begin{aligned} x_1(t) &= B_1^+ e^{i\omega_1 t} + B_1^- e^{-i\omega_1 t} + B_2^+ e^{i\omega_2 t} + B_2^- e^{-i\omega_2 t} \\ x_2(t) &= -B_1^+ e^{i\omega_1 t} - B_1^- e^{-i\omega_1 t} + B_2^+ e^{i\omega_2 t} + B_2^- e^{-i\omega_2 t} \end{aligned}\right\} \tag{12.10}$$

Thus, we have *four* arbitrary constants in the general solution - just as we expect - because we have two equations of motion that are of *second* order.

We mentioned earlier that we could always define a set of coordinates that have a simple time dependence and that correspond to the excitation of the various oscillation modes of the system. Let us examine the pair of coordinates defined by

$$\left.\begin{aligned} \eta_1 &\equiv x_1 - x_2 \\ \eta_2 &\equiv x_1 + x_2 \end{aligned}\right\} \tag{12.11}$$

or

$$\left.\begin{aligned} x_1 &= \frac{1}{2}(\eta_2 + \eta_1) \\ x_2 &= \frac{1}{2}(\eta_2 - \eta_1) \end{aligned}\right\} \tag{12.12}$$

Substituting these expressions for x_1 and x_2 into Equation 12.1, we find

$$\left.\begin{aligned} M(\ddot{\eta}_1 + \ddot{\eta}_2) + (\kappa + 2\kappa_{12})\eta_1 + \kappa\eta_2 = 0 \\ M(\ddot{\eta}_1 - \ddot{\eta}_2) + (\kappa + 2\kappa_{12})\eta_1 - \kappa\eta_2 = 0 \end{aligned}\right\} \tag{12.13}$$

which can be solved (by adding and subtracting) to yield

$$\left.\begin{aligned} M\ddot{\eta}_1 + (\kappa + 2\kappa_{12})\eta_1 = 0 \\ M\ddot{\eta}_2 + \kappa\eta_2 = 0 \end{aligned}\right\} \tag{12.14}$$

The coordinates η_1 and η_2 now *uncoupled* and are therefore *independent*. The solutions are

$$\left.\begin{aligned} \eta_1(t) &= C_1^+ e^{i\omega_1 t} + C_1^- e^{-i\omega_1 t} \\ \eta_2(t) &= C_2^+ e^{i\omega_2 t} + C_2^- e^{-i\omega_2 t} \end{aligned}\right\} \tag{12.15}$$

where the frequencies ω_1 and ω_2 are given by Equations 12.8. Thus, η_1 and η_2 are the *normal coordinates* of the problem. In a later section, we establish a general method for obtaining the normal coordinates.

If we impose the special initial conditions $x_1(0) = -x_2(0)$ and $\dot{x}_1(0) = -\dot{x}_2(0)$, we find $\eta_2(0) = 0$ and $\dot{\eta}_2(0) = 0$, which leads to $C_2^+ = C_2^- = 0$; that is, $\eta_2(t) = 0$ for all values of t. Thus, the particles oscillate always *out of phase* and with frequency ω_1; this is the **antisymmetrical** mode of oscillation. However, if we begin with $x_1(0) = x_2(0)$ and $\dot{x}_1(0) = \dot{x}_2(0)$, we find $\eta_1(t) = 0$, and the particles oscillate in phase and with frequency ω_2; this is the **symmetrical** mode of oscillation. These results are illustrated schematically in Figure 12-2.

The general motion of the system is a linear combination of the symmetrical and antisymmetrical modes. The fact that the antisymmetrical mode has the higher frequency and the symmetrical mode has the lower frequency is actually a general result. In a complex system of linearly coupled oscillators, the mode possessing the highest degree of symmetry has the lowest frequency. If the symmetry is destroyed, then the springs must "work harder" in the antisymmetrical modes, and the frequency is raised.

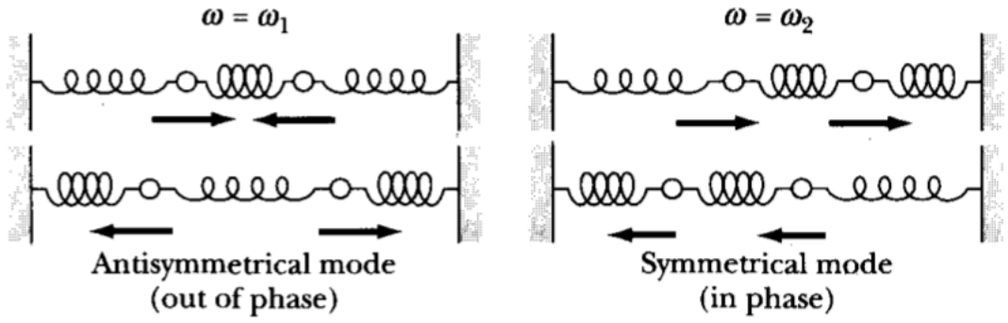


FIGURE 12-2 The two characteristic frequencies are indicated schematically. One is the antisymmetrical mode (masses are out of phase) and the other is the symmetrical mode (masses are in phase).

Notice that if we were to hold OT2 fixed and allow wj to oscillate, the frequency would be V (k -I- ki2)/ M. We would obtain the same result for the frequency of oscillation of ot2 if "*i were held fixed. The oscillators are identical and in the absence of coupling have the same oscillation frequency. The effect of coupling is to separate the common frequency, with one characteristic frequency becoming larger and one becoming smaller than the frequency for uncoupled motion. If we denote by (Oq the frequency for uncoupled motion, then Wi > (Oq > @2, and we may schematically indicate the effect of the coupling as in Figure 12-3a. The solution for the characteristic frequencies in the problem of three coupled identical masses is illustrated in Figure 12-3b.

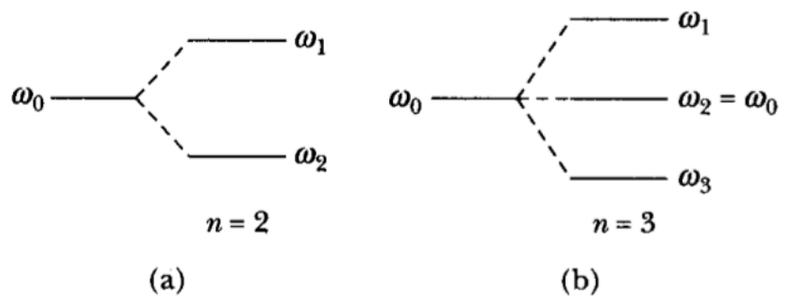


FIGURE 12-3 (a) Coupling separates the common frequency for two identical masses, with one characteristic frequency being higher and one being lower than the frequency ω_0 for uncoupled motion. (b) For three coupled identical masses, one characteristic frequency is smaller than ω_0 and one is larger. For n (number of oscillators) odd, one characteristic frequency is equal to ω_0. The separations are only schematic.

Again, we have a splitting of the characteristic frequencies, with one greater and one smaller than . This is a general result: For an even number n of identical nearest neighbor coupled oscillators, n/2 characteristic frequencies are greater than ω_0, and n/2 characteristic frequencies are smaller than ω_0. If n is odd, one characteristic frequency is equal to ω_0, and the remaining n-1 characteristic frequencies are symmetrically distributed above and below ω_0. The reader familiar with the phenomenon of the Zeeman effect in atomic spectra will appreciate the similarity with this result: In each case, there is a symmetrical splitting of the frequency caused by the introduction of an interaction (in one case by the application of a magnetic field and in the other by the coupling of particles through the intermediary of the springs).

12.3 Weak Coupling

Some of the more interesting cases of coupled oscillations occur when the coupling is *weak* - that is, when the force constant of the coupling spring is small compared with that of the oscillator springs: $\kappa_{12} \ll \kappa$ According to Equations 12.8, the frequencies ω_1 and ω_2 are

$$\omega_1 = \sqrt{\frac{\kappa + 2\kappa_{12}}{M}}, \qquad \omega_2 = \sqrt{\frac{\kappa}{M}} \tag{12.16}$$

If the coupling is weak, we may expand the expression for ω_1

$$\omega_1 = \sqrt{\frac{\kappa}{M}}\sqrt{1 + \frac{2\kappa_{12}}{\kappa}} = \sqrt{\frac{\kappa}{M}}\sqrt{1 + 4\varepsilon}$$

where

$$\varepsilon \equiv \frac{\kappa_{12}}{2\kappa} \ll 1 \tag{12.17}$$

The frequency ω_1 now reduces to

$$\omega_1 \cong \sqrt{\frac{\kappa}{M}}(1 + 2\varepsilon) \tag{12.18}$$

The natural frequency of either oscillator, when the other is held fixed, is

$$\omega_0 = \sqrt{\frac{\kappa + \kappa_{12}}{M}} \cong \sqrt{\frac{\kappa}{M}}(1 + \varepsilon) \tag{12.19}$$

or

$$\sqrt{\frac{\kappa}{M}} \cong \omega_0(1 - \varepsilon) \tag{12.20}$$

Therefore, the two characteristic frequencies are given approximately by

$$\left.\begin{aligned} \omega_1 &\cong \sqrt{\frac{\kappa}{M}}(1 + 2\varepsilon), & \omega_2 &= \sqrt{\frac{\kappa}{M}} \\ &\cong \omega_0(1 - \varepsilon)(1 + 2\varepsilon) & &\cong \omega_0(1 - \varepsilon) \\ &\cong \omega_0(1 + \varepsilon) & & \end{aligned}\right\} \tag{12.21}$$

We can now examine the way a weakly coupled system behaves. If we displace Oscillator 1 a distance D and release it from rest, the initial conditions for the system are

$$x_1(0) = D, \quad x_2(0) = 0, \quad \dot{x}_1(0) = 0, \quad \dot{x}_2(0) = 0 \tag{12.22}$$

If we substitute these initial conditions into Equation 12.10 for $x_1(t)$ and $x_2(t)$, we find the amplitudes to be

$$B_1^+ = B_1^- = B_2^+ = B_2^- = \frac{D}{4} \tag{12.23}$$

Then, $x_1(t)$ becomes

$$\begin{aligned} x_1(t) &= \frac{D}{4}[(e^{i\omega_1 t} + e^{-i\omega_1 t}) + (e^{i\omega_2 t} + e^{-i\omega_2 t})] \\ &= \frac{D}{2}(\cos \omega_1 t + \cos \omega_2 t) \\ &= D \cos\left(\frac{\omega_1 + \omega_2}{2} t\right) \cos\left(\frac{\omega_1 - \omega_2}{2} t\right) \end{aligned} \tag{12.24}$$

But, according to Equation 12.21,

$$\frac{\omega_1 + \omega_2}{2} = \omega_0; \quad \frac{\omega_1 - \omega_2}{2} = \varepsilon\omega_0 \tag{12.25}$$

Therefore (Note that in this fortuitous case, Xj and x^ were always real, so the real part did not have to be expressly taken in the final step as outlined after Equation 12.2.),

$$x_1(t) = (D \cos \varepsilon\omega_0 t)\cos \omega_0 t \tag{12.26a}$$

Similarly,

$$x_2(t) = (D \sin \varepsilon\omega_0 t)\sin \omega_0 t \tag{12.26b}$$

Because ε is small, the quantities $D\cos\varepsilon\omega_0 t$ and $D\sin\varepsilon\omega_0 t$ vary slowly with time. Therefore, $x_1(t)$ and $x_2(t)$ are essentially sinusoidal functions with slowly varying amplitudes. Although only x_1 is initially different from zero, as time increases the amplitude of x_1 decreases slowly with time, and the amplitude of x_2 increases slowly from zero.

Hence, energy is transferred from the first oscillator to the second. When $t = \pi/2\varepsilon\omega_0$, then $Dcos\varepsilon\omega_0 t = 0$, and all the energy has been transferred. As time increases further, energy is transferred back to the first oscillator. This is the familiar phenomenon of *beats* and is illustrated in Figure 12-4. (In the case illustrated, e = 0.08.)

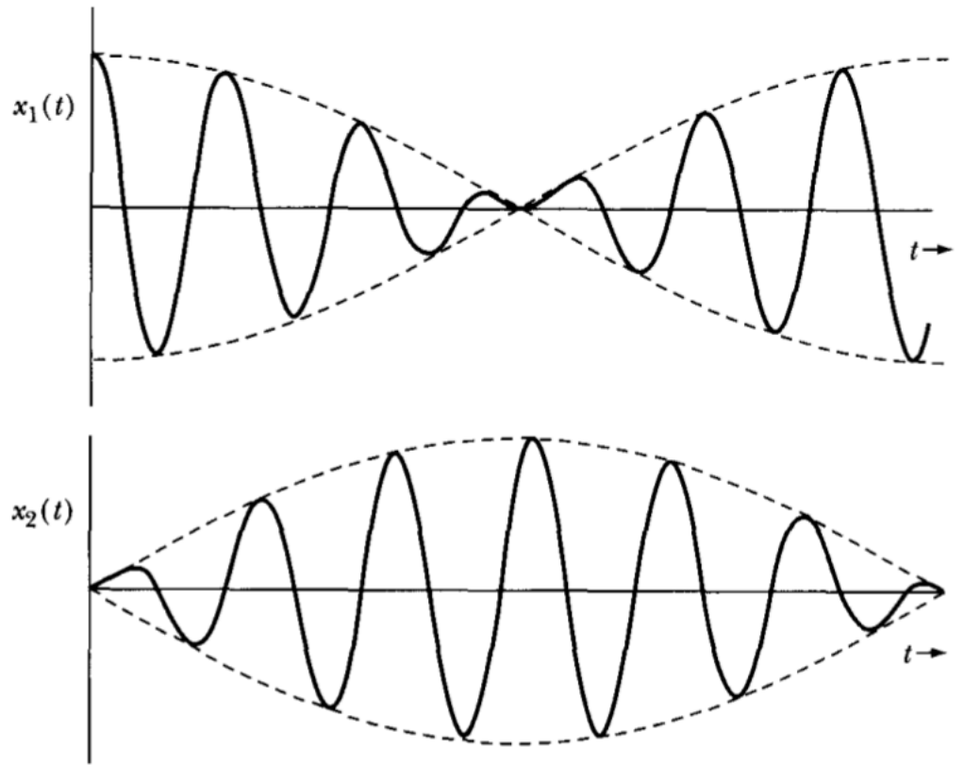


FIGURE 12-4 The solutions for $x_1(t)$ and $x_2(t)$ have a high frequency component (ω_0) that oscillates inside a slowly varying component ($\varepsilon\omega_0$). Energy is transferred back and forth between the two oscillators.

12.4 General Problem of Coupled Oscillations

In the preceding sections, we found that the effect of coupling in a simple system with two degrees of freedom produced two characteristic frequencies and two modes of oscillation. We now turn our attention to the general problem of coupled oscillations. Let us consider a conservative system described in terms of a set of generalized coordinates q_k and the time t. If the system has n degrees of freedom, then k = 1, 2,..., n. We specify that a configuration of stable equilibrium exists for the system and that at equilibrium the generalized coordinates have values q_{k0}. In such a configuration, Lagrange's equations are satisfied by

$$q_k = q_{k0}, \qquad \dot{q}_k = 0, \qquad \ddot{q}_k = 0, \qquad k = 1, 2, \ldots, n$$

Every nonzero term of the form $(d/dt)(\partial L/\partial \dot{q}_k)$ must contain at least either $\dot{q}_k$ or $\ddot{q}_k$, so all such terms vanish at equilibrium. From Lagrange's equation, we therefore have

$$\left.\frac{\partial L}{\partial q_k}\right|_0 = \left.\frac{\partial T}{\partial q_k}\right|_0 - \left.\frac{\partial U}{\partial q_k}\right|_0 = 0 \tag{12.27}$$

where the subscript 0 designates that the quantity is evaluated at equilibrium.

We assume that the equations connecting the generalized coordinates and the rectangular coordinates do not explicitly contain the time; that is, we have

$$x_{\alpha,i} = x_{\alpha,i}(q_j) \quad \text{or} \quad q_j = q_j(x_{\alpha,i})$$

The kinetic energy is thus a homogeneous quadratic function of the generalized velocities (see Equation 7.121):

$$T = \frac{1}{2}\sum_{j,k} m_{jk}\dot{q}_j\dot{q}_k \tag{12.28}$$

Therefore, in general,

$$\left.\frac{\partial T}{\partial q_k}\right|_0 = 0, \quad k = 1, 2, \ldots, n \tag{12.29}$$

and hence, from Equation 12.27, we have

$$\left.\frac{\partial U}{\partial q_k}\right|_0 = 0, \quad k = 1, 2, \ldots, n \tag{12.30}$$

We may further specify that the generalized coordinates q_k be measured from the equilibrium positions; that is, we choose $q_{k0} = 0$. (If we originally had chosen a set of coordinates q'_k such that $q'_{k0} \neq 0$, we could always effect a simple linear transformation of the form $q_k = q'_k + \alpha_k$ such that $q_{k0} = 0$,)

The expansion of the potential energy in a Taylor series about the equilibrium configuration yields

$$U(q_1, q_2, \ldots, q_n) = U_0 + \sum_k \left.\frac{\partial U}{\partial q_k}\right|_0 q_k + \frac{1}{2}\sum_{j,k} \left.\frac{\partial^2 U}{\partial q_j \partial q_k}\right|_0 q_j q_k + \cdots \tag{12.31}$$

The second term in the expansion vanishes in view of Equation 12.30, and - without loss of generality - we may choose to measure U in such a way that $U_0 = 0$. Then, if we restrict the motion of the generalized coordinates to be small, we may neglect all terms in the expansion containing products of the q_k of degree higher than second. This is equivalent to restricting our attention to simple harmonic oscillations, in which case only terms quadratic in the coordinates appear. Thus,

$$U = \frac{1}{2}\sum_{j,k} A_{jk} q_j q_k \tag{12.32}$$

where we define

$$A_{jk} \equiv \left.\frac{\partial^2 U}{\partial q_j \partial q_k}\right|_0 \tag{12.33}$$

Because the order of differentiation is immaterial (if U has continuous second partial derivatives), the quantity A_{jk} is symmetrical; that is, $A_{jk} = A_{kj}$.

We have specified that the motion of the system is to take place in the vicinity of the equilibrium configuration, and we have shown (Equation 12.30) that U must have a minimum value when the system is in this configuration. Because we have chosen $U = 0$ at equilibrium, we must have, in general, $U \geq 0$. It should be clear that we must also have $T \geq 0$. That is, both U and Tare positive definite quantities, in that they are always positive unless the coordinates (in the case of U) or the velocities (in the case of T) are zero, in which case they vanish.

Equations 12.28 and 12.32 are of a similar form:

$$T = \frac{1}{2}\sum_{j,k} m_{jk}\dot{q}_j\dot{q}_k$$

$$U = \frac{1}{2}\sum_{j,k} A_{jk}q_jq_k \tag{12.34}$$

The quantities A_{jk} are just numbers (see Equation 12.33); but the m_{jk} may be functions of the coordinates (see Equation 7.119):

$$m_{jk} = \sum_\alpha m_\alpha \sum_i \frac{\partial x_{\alpha,i}}{\partial q_j}\frac{\partial x_{\alpha,i}}{\partial q_k}$$

We can expand the m_{jk} about the equilibrium position with the result

$$m_{jk}(q_1, q_2, \ldots, q_n) = m_{jk}(q_{l0}) + \sum_l \left.\frac{\partial m_{jk}}{\partial q_l}\right|_0 q_l + \cdots \tag{12.35}$$

We wish to retain only the first nonvanishing term in this expansion; but, unlike the expansion of the potential energy (Equation 12.31), we cannot choose the constant term $m_{j0}(q_{l0})$ to be zero, so this leading term becomes the constant value of m_{jk} in this approximation. This is the same order of approximation as that used for U, because the next higher order term in T would involve the cubic quantity $q_jq_kq_l$ and the next higher order term in U would contain $\dot{q}_j\dot{q}_kq_l$ - In the small oscillation approximation, T should be treated similarly to U, and just like U is normally expanded to order q^2, one needs to expand T to order $\dot{q}^2$, and m_{jk} is evaluated at equilibrium.

Thus, in Equation 12.34, the m_{jk} and the A_{jk} are n x n arrays of numbers specifying the way the motions of the various coordinates are coupled. For example, if $m_{rs} \neq 0$ for $r \neq s$, then the kinetic energy contains a term proportioned to $\dot{q}_r \dot{q}_s$, and a coupling exists between the r^{th} and s^{th} coordinate. If, however, m_{jk} is *diagonal*, so that $m_{jk} \neq 0$ for j = k but vanishes otherwise, then the kinetic energy is of the form

$$T = \frac{1}{2} \sum_r m_r \dot{q}_r^2$$

where m_{rr} has been abbreviated to m_r.

We note that if a diagonal element of m_{jk} (say, m_{rr}) vanishes, then the problem can be reduced to one of n - 1 degrees of freedom.

Thus, the kinetic energy is a simple sum of the kinetic energies associated with the various coordinates. As we see below, if, in addition, A_{jk} is diagonal so that U is also a simple sum of individual potential energies, then each coordinate behaves in an uncomplicated manner, undergoing oscillations with a single, well-defined frequency. The problem is therefore to find a coordinate transformation that simultaneously diagonalizes both m_{jk} and A_{jk} and thereby renders the system describable in the simplest possible terms. Such coordinates are the *normal coordinates*.

The equations of motion of the system with kinetic and potential energies given by Equation 12.34 are obtained from Lagrange's equation

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

But because T is a function only of the generalized velocities and U is a function only of the generalized coordinates, Lagrange's equation for the k^{th} coordinate becomes

$$\frac{\partial U}{\partial q_k} + \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_k} = 0 \qquad (12.36)$$

$$\left.\begin{aligned}\frac{\partial U}{\partial q_k} &= \sum_j A_{jk} q_j \\ \frac{\partial T}{\partial \dot{q}_k} &= \sum_j m_{jk} \dot{q}_j\end{aligned}\right\} \tag{12.37}$$

The equations of motion then become

$$\boxed{\sum_j (A_{jk} q_j + m_{jk} \ddot{q}_j) = 0} \tag{12.38}$$

This is a set of n second-order linear homogeneous differential equations with constant coefficients. Because we are dealing with an oscillatory system, we expect a solution of the form

$$q_j(t) = a_j e^{i(\omega t - \delta)} \tag{12.39}$$

where the a_j are real amplitudes and where the phase δ has been included to give the two arbitrary constants (a_j and δ) required by the second-order nature of each of the differential equations. (Only the real part of the right-hand side is to be considered.) The frequency ω and the phase δ are to be determined by the equations of motion. If ω is a real quantity, then the solution (Equation 12.39) represents oscillatory motion. That ω is indeed real may be seen by the following physical argument. Suppose that ω contains an imaginary part $i\omega_i$ (in which ω_i is real). This produces terms of the form $e^{\omega_i t}$ and $e^{-\omega_i t}$ in the expression of q_j. Thus, when the total energy of the system is computed, T + U contains factors that increase or decrease monotonically with the time. But this violates the assumption that we are dealing with a conservative system: therefore, the frequency ω must be a real quantity.

With a solution of the form given by Equation 12.39, the equations of motion become

$$\boxed{\sum_j (A_{jk} - \omega^2 m_{jk})\, a_j = 0} \tag{12.40}$$

where the common factor exp[i(ωt - δ)] has been canceled. This is a set of n linear, homogeneous, algebraic equations that the a_j must satisfy. For a nontrivial solution to exist, the determinant of the coefficients must vanish:

$$|A_{jk} - \omega^2 m_{jk}| = 0 \tag{12.41}$$

To be more explicit, this is an n x n determinant of the form

$$\begin{vmatrix} A_{11} - \omega^2 m_{11} & A_{12} - \omega^2 m_{12} & A_{13} - \omega^2 m_{13} \cdots \\ A_{12} - \omega^2 m_{12} & A_{22} - \omega^2 m_{22} & A_{23} - \omega^2 m_{23} \cdots \\ A_{13} - \omega^2 m_{13} & A_{23} - \omega^2 m_{23} & A_{33} - \omega^2 m_{33} \cdots \\ \vdots & \vdots & \vdots \end{vmatrix} = 0 \tag{12.42}$$

where the symmetry of the A_{jk} and m_{jk} has been explicitly included.

The equation represented by this determinant is called the **characteristic equation** or **secular equation** of the system and is an equation of degree n in ω^2. There are, in general, n roots we may label ω_r^2 . The ω_r are called the **characteristic frequencies** or **eigenfrequencies** of the system. (In some situations, two or more of the ω_r can be equal; this is the phenomenon of **degeneracy** and is discussed later.) Just as in the procedure for determining the directions of the principal axes for a rigid body, each of the roots of the characteristic equation may be substituted into Equation 12.40 to determine the ratios $a_1 : a_2 : a_3 : \cdots : a_n$ for each value of ω_r . Because there are n values of ω_r , we can construct n sets of ratios of the a_j. Each of the sets defines the components of the n-dimensional vector $\mathbf{a}_r$, called an **eigenvector** of the system. Thus $\mathbf{a}_r$ is the eigenvector associated with the eigenfrequency ω_r . We designate by a_{jr} the j^{th} component of the r^{th} eigenvector.

Because the principle of superposition applies for the differential equation Equation 12.38), we must write the general solution for q_j as a linear combination of the solutions for each of the n values of r:

$$q_j(t) = \sum_r a_{jr} e^{i(\omega_r t - \delta_r)} \tag{12.43}$$

Because it is only the real part of $q_j\{t)$ that is physically meaningful, we actually have

$$q_j(t) = \mathrm{Re} \sum_r a_{jr} e^{i(\omega_r t - \delta_r)} = \sum_r a_{jr} \cos(\omega_r t - \delta_r) \tag{12.44}$$

The motion of the coordinate q_j is therefore compounded of motions with each of the n values of the frequencies ω_r . The q_j evidently are not the normal coordinates that simplify the problem. We continue the search for normal coordinates in Section 12.6.

EXAMPLE 12.1

Find the characteristic frequencies for the case of the two masses connected by springs of Section 12.2 by means of the general formalism just developed.

Solution. The situation is that shown in Figure 12-1.

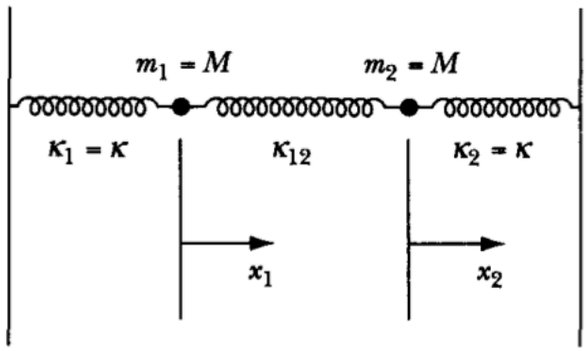


FIGURE 12-1 Two masses are connected by a spring to each other and by springs to fixed positions. This is a system of coupled motion in one dimension.

The potential energy of the system is

$$\begin{aligned} U &= \frac{1}{2}\kappa x_1^2 + \frac{1}{2}\kappa_{12}(x_2 - x_1)^2 + \frac{1}{2}\kappa x_2^2 \\ &= \frac{1}{2}(\kappa + \kappa_{12})x_1^2 + \frac{1}{2}(\kappa + \kappa_{12})x_2^2 - \kappa_{12}x_1x_2 \end{aligned} \tag{12.45}$$

The term proportional to x_1x_2 is the factor that expresses the coupling in the system. Calculating the A_{jk} we find

$$\left.\begin{aligned} A_{11} &= \left.\frac{\partial^2 U}{\partial x_1^2}\right|_0 = \kappa + \kappa_{12} \\ A_{12} &= \left.\frac{\partial^2 U}{\partial x_1 \partial x_2}\right|_0 = -\kappa_{12} = A_{21} \\ A_{22} &= \left.\frac{\partial^2 U}{\partial x_2^2}\right|_0 = \kappa + \kappa_{12} \end{aligned}\right\} \tag{12.46}$$

The kinetic energy of the system is

$$T = \frac{1}{2} M\dot{x}_1^2 + \frac{1}{2} M\dot{x}_2^2 \tag{12.47}$$

According to Equation 12.28,

$$T = \frac{1}{2} \sum_{j,k} m_{jk}\dot{x}_j\dot{x}_k \tag{12.48}$$

Identifying terms between these two expressions for T, we find

$$\left.\begin{aligned} m_{11} &= m_{22} = M \\ m_{12} &= m_{21} = 0 \end{aligned}\right\} \tag{12.49}$$

Thus, the secular determinant (Equation 12.42) becomes

$$\begin{vmatrix} \kappa + \kappa_{12} - M\omega^2 & -\kappa_{12} \\ -\kappa_{12} & \kappa + \kappa_{12} - M\omega^2 \end{vmatrix} = 0 \tag{12.50}$$

This is exactly Equation 12.5, so the solutions are the same (see Equations 12.7 and 12.8) as before:

$$\omega = \sqrt{\frac{\kappa + \kappa_{12} \pm \kappa_{12}}{M}}$$

The eigenfrequencies are

$$\omega_1 = \sqrt{\frac{\kappa + 2\kappa_{12}}{M}}, \qquad \omega_2 = \sqrt{\frac{\kappa}{M}}$$

The results of the two procedures are identical.

12.5 Orthogonality of the Eigenvectors

We now wish to show that the eigenvectors $\mathbf{a}_r$ form an orthonormal set. Rewriting Equation 12.40 for the sth root of the secular equation, we have

$$\omega_s^2 \sum_k m_{jk} a_{ks} = \sum_k A_{jk} a_{ks} \tag{12.51}$$

Next, we write a comparable equation for the r^{th} root by substituting r for s and interchanging J and k:

$$\omega_r^2 \sum_j m_{jk} a_{jr} = \sum_j A_{jk} a_{jr} \tag{12.52}$$

where we have used the symmetry of the m_{jk} and A_{jk}. We now multiply Equation 12.51 by a_{jr} and sum over j and also multiply Equation 12.52 by a_{ks} and sum over k:

$$\left.\begin{aligned} \omega_s^2 \sum_{j,k} m_{jk} a_{jr} a_{ks} &= \sum_{j,k} A_{jk} a_{jr} a_{ks} \\ \omega_r^2 \sum_{j,k} m_{jk} a_{jr} a_{ks} &= \sum_{j,k} A_{jk} a_{jr} a_{ks} \end{aligned}\right\} \tag{12.53}$$

The right-hand sides of Equations 12.53 are now equal, so subtracting the first of these equations from the second, we have

$$(\omega_r^2 - \omega_s^2) \sum_{j,k} m_{jk} a_{jr} a_{ks} = 0 \tag{12.54}$$

We now examine the two possibilities r = s and r ≠ s. For r ≠ s, the term $(\omega_r^2 - \omega_s^2)$ is, in general, different from zero. (The case of degeneracy, or multiple roots, is discussed later.) Therefore the sum must vanish identically:

$$\sum_{j,k} m_{jk} a_{jr} a_{ks} = 0, \quad r \neq s \tag{12.55}$$

For the case r= s, the term $(\omega_r^2 - \omega_s^2)$ vanishes and the sum is indeterminate. The sum, however, cannot vanish identically. To show this, we write the kinetic energy for the system and substitute the expressions for $\dot{q}_j$ and $\dot{q}_k$ from Equation 12.44:

$$T = \frac{1}{2} \sum_{j,k} m_{jk} \dot{q}_j \dot{q}_k$$

$$= \frac{1}{2} \sum_{j,k} m_{jk} \left[\sum_r \omega_r a_{jr} \sin(\omega_r t - \delta_r) \right] \left[\sum_s \omega_s a_{ks} \sin(\omega_s t - \delta_s) \right]$$

$$= \frac{1}{2} \sum_{r,s} \omega_r \omega_s \sin(\omega_r t - \delta_r) \sin(\omega_s t - \delta_s) \sum_{j,k} m_{jk} a_{jr} a_{ks}$$

Thus, for r = s, the kinetic energy becomes

$$T = \frac{1}{2} \sum_r \omega_r^2 \sin^2(\omega_r t - \delta_r) \sum_{j,k} m_{jk} a_{jr} a_{kr} \tag{12.56}$$

We note first that

$$\omega_r^2 \sin^2(\omega_r t - \delta_r) \geq 0$$

We also know that T is positive and can become zero only if all the velocities vanish identically. Therefore,

$$\sum_{j,k} m_{jk} a_{jr} a_{kr} \geq 0$$

Thus, the sum is, in general, positive and can vanish only in the trivial instance that the system is not in motion - that is, that the velocities vanish identically and T = 0.

We previously remarked that only the ratios of the a_{jr} are determined when the ω_r are substituted into Equation 12.40. We now remove this indeterminacy by imposing an additional condition on the a_{jr}. We require that

$$\sum_{j,k} m_{jk} a_{jr} a_{kr} = 1 \tag{12.57}$$

The a_{jr} are then said to be *normalized*. Combining Equations 12.55 and 12.57, we may write

$$\boxed{\sum_{j,k} m_{kj} a_{jr} a_{ks} = \delta_{rs}} \tag{12.58}$$

Because a_{jr} is the j^{th} component of the r^{th} eigenvector, we represent $\mathbf{a}_r$ by

$$\mathbf{a}_r = \sum_j a_{jr} \mathbf{e}_j \tag{12.59}$$

The vectors $\mathbf{a}_r$ defined in this way constitute an **orthonormal** set; that is, they are orthogonal according to the result given by Equation 12.55, and they have been normalized by setting the sum in Equation 12.57 equal to unity.

All the preceding discussion bears a striking resemblance to the procedure given in Chapter 11 for determining the principal moments of inertia and the principal axes for a rigid body. Indeed, the problems are mathematically identical, except that we are now dealing with a system with n degrees of freedom.

The quantities m_{jk} and A_{jk} are actually tensor elements, because m and A are two-dimensional arrays that relate different physical quantities, and as such, we write them as $\{\mathbf{m}\}$ and $\{\mathbf{A}\}$. The secular equation for determining the eigenfrequencies is the same as that for obtaining the principal moments of inertia, and the eigenvectors $\mathbf{a}_r$ correspond to the principal axes.

Indeed, the proof of the orthogonality of the eigenvectors is merely a generalization of the proof given in Section 11.6 of the orthogonality of the principal axes. Although we have made a physical argument regarding the reality of the eigenfrequencies, we could carry out a mathematical proof using the same procedure used to show that the principal moments of inertia are real.

12.6 Normal Coordinates

As we have seen (Equation 12.43), the general solution for the motion of the coordinate q_j must be a sum over terms, each of which depends on an individual eigenfrequency. In the previous section, we showed that the vectors $\mathbf{a}_r$ are orthogonal (Equation 12.55) and, as a matter of convenience, we even normalized their components a_{jr} (Equation 12.57) to arrive at Equation 12.58; that is, we have removed all ambiguity in the solution for the q_j, so it is no longer possible to specify an arbitrary displacement for a particle. Because such a restriction is not physically meaningful, we must introduce a constant scale factor α_r (which depends on the initial conditions of the problem) to account for the loss of generality introduced by the arbitrary normalization. Thus,

$$q_j(t) = \sum_r \alpha_r a_{jr} e^{i(\omega_r t - \delta_r)} \tag{12.60}$$

To simplify the notation, we write

$$q_j(t) = \sum_r \beta_r a_{jr} e^{i\omega_r t} \tag{12.61}$$

where the quantities β_r are new scale factors (now complex) that incorporate the phases of δ_r.

We now define a quantity η_r ,

$$\boxed{\eta_r(t) \equiv \beta_r e^{i\omega_r t}} \tag{12.62}$$

so that

$$\boxed{q_j(t) = \sum_r a_{jr}\eta_r(t)} \tag{12.63}$$

The η_r, by definition, are quantities that undergo oscillation at only one frequency. They may be considered as new coordinates, called *normal coordinates* for the system. The η_r satisfy equations of the form

$$\ddot{\eta}_r + \omega_r^2 \eta_r = 0 \tag{12.64}$$

There are n independent such equations, so the equations of motion expressed in normal coordinates become completely separable.

EXAMPLE 12.2

Derive Equation 12.64 directly by using Lagrange's equations of motion.

Solution. We note that from Equation 12.63

$$\dot{q}_j = \sum_j a_{jr} \dot{\eta}_r$$

and from Equation 12.34 we have, for the kinetic energy.

$$\begin{aligned} T &= \frac{1}{2} \sum_{j,k} m_{jk} \dot{q}_j \dot{q}_k \\ &= \frac{1}{2} \sum_{j,k} m_{jk} \left(\sum_r a_{jr} \dot{\eta}_r \right) \left(\sum_s a_{ks} \dot{\eta}_s \right) \\ &= \frac{1}{2} \sum_{r,s} \left(\sum_{j,k} m_{jk} a_{jr} a_{ks} \right) \dot{\eta}_r \dot{\eta}_s \end{aligned}$$

The sum in the parentheses is just δ_{rs} according to the orthonormality condition (Equation 12.58). Therefore,

$$T = \frac{1}{2} \sum_{r,s} \dot{\eta}_r \dot{\eta}_s \delta_{rs} = \frac{1}{2} \sum_r \dot{\eta}_r^2 \tag{12.65}$$

Similarly, from Equations 12.34 we have for the potential energy,

$$U = \frac{1}{2}\sum_{j,k} A_{jk} q_j q_k$$

$$= \frac{1}{2}\sum_{r,s}\left(\sum_{j,k} A_{jk} a_{jr} a_{ks}\right)\eta_r \eta_s$$

The first equation in Equation 12.53 is

$$\sum_{j,k} A_{jk} a_{jr} a_{ks} = \omega_s^2 \sum_{j,k} m_{jk} a_{jr} a_{ks}$$

$$= \omega_s^2 \delta_{rs}$$

so the potential energy becomes

$$U = \frac{1}{2}\sum_{r,s} \omega_s^2 \eta_r \eta_s \delta_{rs} = \frac{1}{2}\sum_r \omega_r^2 \eta_r^2 \tag{12.66}$$

Using Equations 12.65 and 12.66, the Lagrangian is

$$L = \frac{1}{2}\sum_r (\dot{\eta}_r^2 - \omega_r^2 \eta_r^2) \tag{12.67}$$

and Lagrange's equations are

$$\frac{\partial L}{\partial \eta_r} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\eta}_r} = 0$$

or

$$\ddot{\eta}_r + \omega_r^2 \eta_r = 0$$

as found in Equation 12.64.

Thus, when the configuration of a system is expressed in normal coordinates, both the potential and kinetic energies become simultaneously diagonal. Because it is the off-diagonal elements of $\{\mathbf{m}\}$ and $\{\mathbf{A}\}$ that give rise to the coupling of the particles' motions, it should be evident that a choice of coordinates that renders these tensors diagonal uncouples the coordinates and makes the problem completely separable into the independent motions of the normal coordinates, each with its particular normal frequency.

The foregoing has been a mathematical description of the methods used to determine the characteristic frequencies ω_r and to describe the coordinates η_r of the normal mode motion. The actual application of the method can be summarized by several statements:

1. **Choose generalized coordinates and find T and U in the normal Lagrangian method. This corresponds to using Equations 12.34.**
2. **Represent A_{jk} and m_{jk} as tensors in $n \times n$ arrays, and use Equation 12.42 to determine the n values of eigenfrequencies ω_r.**
3. **For each value of ω_r, determine the ratios $a_{1r} : a_{2r} : a_{3r} : \cdots : a_{nr}$ by substituting into Equation 12.40:**

$$\sum_j (A_{jk} - \omega_r^2 m_{jk})\, a_{jr} = 0 \tag{12.68}$$

4. **If needed, determine the scale factors β_r (Equation 12.60) from the initial conditions.**
5. **Determine the normal coordinates η_r by appropriate linear combinations of the q_j coordinates that display oscillations at the single eigenfrequency ω_r. The description of motion for this single normal coordinate η_r is called a normal mode. The general motion (Equation 12.63) of the system is a complicated superposition of the normal modes.**

We now apply these steps in several examples.

EXAMPLE 12.3

Determine the eigenfrequencies, eigenvectors, and normal coordinates of the mass-spring example in Section 12.2 by using the procedure just described. Assume $\kappa_{12} \approx \kappa$.

Solution. The eigenfrequencies were determined in Example 12.1, where we found T and U (step 1). We can find the components for A_{jk}, directly from Equation 12.46 or by inspection from Equation 12.45, making sure A_{jk}, is symmetrical

$$\{\mathbf{A}\} = \begin{Bmatrix} \kappa + \kappa_{12} & -\kappa_{12} \\ -\kappa_{12} & \kappa + \kappa_{12} \end{Bmatrix} \tag{12.69}$$

The array m_{jk} can easily be determined from Equation 12.47:

$$\{\mathbf{m}\} = \begin{Bmatrix} M & 0 \\ 0 & M \end{Bmatrix} \tag{12.70}$$

We use Equation 12.42 to determine the eigenfrequencies ω_r .

$$\begin{vmatrix} \kappa + \kappa_{12} - M\omega^2 & -\kappa_{12} \\ -\kappa_{12} & \kappa + \kappa_{12} - M\omega^2 \end{vmatrix} = 0$$

which is identical to Equation 12.50 with the results of Equation 12.8 for ω_1 and ω_2 .

We use Equation 12.68 to determine the eigenvector components a_r. We have two equations for each value of r, but because we can determine only the ratios a_1/a_2, one equation for each r is sufficient. For r = 1, k = 1, we have

$$(A_{11} - \omega_1^2 m_{11})\, a_{11} + (A_{21} - \omega_1^2 m_{21})\, a_{21} = 0 \tag{12.71}$$

or, inserting the values for A_{11}, A_{21}, ω_1^2 , and m_{11}, and using the simplification that $\kappa_{12} \approx \kappa$,

$$\left(2\kappa - \frac{3\kappa}{M} \cdot M\right) a_{11} - \kappa a_{21} = 0$$

with the result

$$a_{11} = -a_{21} \tag{12.72}$$

For r = 2, k = 1, we have

$$\left(2\kappa - \frac{\kappa}{M} \cdot M\right) a_{12} - \kappa a_{22} = 0$$

with the result

$$a_{12} = a_{22} \tag{12.73}$$

The general motion (Equation 12.63) becomes

$$\left.\begin{aligned} x_1 &= a_{11}\eta_1 + a_{12}\eta_2 \\ x_2 &= a_{21}\eta_1 + a_{22}\eta_2 \end{aligned}\right\} \tag{12.74}$$

Using Equations 12.72 and 12.73, this becomes

$$\left.\begin{aligned} x_1 &= a_{11}\eta_1 + a_{22}\eta_2 \\ x_2 &= -a_{11}\eta_1 + a_{22}\eta_2 \end{aligned}\right\} \tag{12.75}$$

Adding x_1 and x_2 gives

$$\eta_2 = \frac{1}{2a_{22}}(x_1 + x_2) \tag{12.76}$$

Subtracting x_2 from x_1 gives

$$\eta_1 = \frac{1}{2a_{11}}(x_1 - x_2) \tag{12.77}$$

The normal coordinate η_2 can be determined by finding the conditions when the other normal coordinate η_1 remains equal to zero. From Equation 12.77, $\eta_1 = 0$ when $x_1 = x_2$. Thus, for normal mode 2(η_2), the two masses oscillate in phase (the symmetrical mode). The distance between the particles is always the same, and they oscillate as if the spring connecting them were a rigid, weightless rod.

Similarly, we can find the conditions for the normal coordinate η_1 by determining when $\eta_2 = 0 (x_2 = -x_1)$. In normal mode 1(η_1), the particles oscillate out of phase (the antisymmetrical mode).

This analysis (summarized in Table 12-1) confirms our previous results (Section 12.2), and the particle motion is as shown in Figure 12-2. Such motions for atoms in molecules are common. Remember that we let $\kappa = \kappa_{12}$ during this example.

TABLE 12-1 Normal Mode Motions

Normal mode	Eigenfrequency	Particle oscillation	Particle velocities
1	$\omega_1 = \sqrt{\frac{3\kappa}{M}}$	Out of phase	Equal but opposite
2	$\omega_2 = \sqrt{\frac{\kappa}{M}}$	In phase	Equal

We may determine the components of the eigenvectors (Equation 12.59),

$$\left.\begin{aligned} \omega_1\text{:} \quad \mathbf{a}_1 &= a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2 \\ \omega_2\text{:} \quad \mathbf{a}_2 &= a_{12}\mathbf{e}_1 + a_{22}\mathbf{e}_2 \end{aligned}\right\} \tag{12.78}$$

by using Equations 12.72 and 12.73.

$$\left.\begin{aligned} \mathbf{a}_1 &= a_{11}(\mathbf{e}_1 - \mathbf{e}_2) \\ \mathbf{a}_2 &= a_{22}(\mathbf{e}_1 + \mathbf{e}_2) \end{aligned}\right\} \tag{12.79}$$

Although normally not required, we may determine the values of a_{11} and a_{22} from the orthonormality condition of Equation 12.58 with the result

$$\left.\begin{aligned} a_{11} &= -a_{21} = \frac{1}{\sqrt{2M}} \\ a_{12} &= a_{22} = \frac{1}{\sqrt{2M}} \end{aligned}\right\} \tag{12.80}$$

In this example, it was not necessary to determine the scale factors β, nor to write down the complete solution, because the initial conditions were not given.

EXAMPLE 12.4

Determine the eigenfrequencies and describe the normal mode motion for two pendula of equal lengths b and equal masses m connected by a spring of force constant κ as shown in Figure 12-5. The spring is unstretched in the equilibrium position.

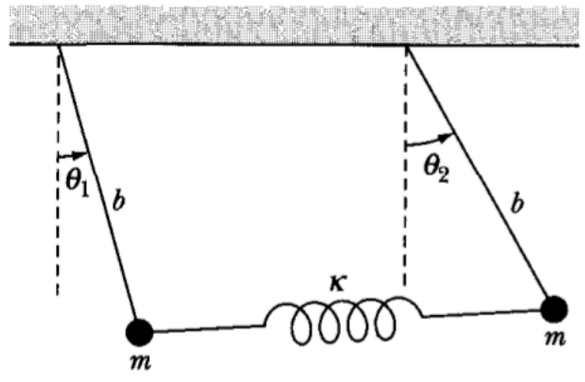


FIGURE 12-5 Example 12.4. Two pendula of equal lengths having equal masses are connected by a spring.

Solution. We choose θ_1 and θ_2 (Figure 12-5) as the generalized coordinates. The potential energy is chosen to be zero in the equilibrium position. The kinetic and potential energies of the system are, for small angles,

$$T = \frac{1}{2}m(b\dot{\theta}_1)^2 + \frac{1}{2}m(b\dot{\theta}_2)^2 \tag{12.81}$$

$$U = mgb(1 - \cos\theta_1) + mgb(1 - \cos\theta_2) + \frac{1}{2}\kappa(b\sin\theta_1 - b\sin\theta_2)^2 \tag{12.82}$$

Using the small oscillation assumption $\sin\theta \approx \theta$ and $\cos\theta \approx 1 - \theta^2/2$, we can write

$$U = \frac{mgb}{2}(\theta_1^2 + \theta_2^2) + \frac{\kappa b^2}{2}(\theta_1 - \theta_2)^2 \tag{12.83}$$

The components of $\{\mathbf{A}\}$ and $\{\mathbf{m}\}$ are

$$\{\mathbf{m}\} = \begin{Bmatrix} mb^2 & 0 \\ 0 & mb^2 \end{Bmatrix} \tag{12.84}$$

$$\{\mathbf{A}\} = \begin{Bmatrix} mgb + \kappa b^2 & -\kappa b^2 \\ -\kappa b^2 & mgb + \kappa b^2 \end{Bmatrix} \tag{12.85}$$

The determinant needed to find the eigenfrequencies ω is

$$\begin{vmatrix} mgb + \kappa b^2 - \omega^2 mb^2 & -\kappa b^2 \\ -\kappa b^2 & mgb + \kappa b^2 - \omega^2 mb^2 \end{vmatrix} = 0 \tag{12.86}$$

which gives the characteristic equation

$$b^2(mg + \kappa b - \omega^2 mb)^2 - (\kappa b^2)^2 = 0$$

$$(mg + \kappa b - \omega^2 mb)^2 = (\kappa b)^2$$

or

$$mg + \kappa b - \omega^2 mb = \pm \kappa b \tag{12.87}$$

Taking the plus sign, $\omega = \omega_1$.

$$mg + \kappa b - \omega_1^2 mb = \kappa b$$

$$\omega_1^2 = \frac{g}{b} \tag{12.88}$$

Taking the minus sign in Equation 12.87, $\omega = \omega_2$,

$$mg + \kappa b - \omega_2^2 mb = -\kappa b$$

$$\omega_2^2 = \frac{g}{b} + \frac{2\kappa}{m} \tag{12.89}$$

Putting the values of ω_1 and ω_2 into Equation 12.40 gives, for k = 1,

$$(mgb + \kappa b^2 - \omega_r^2 mb^2)\, a_{1r} - \kappa b^2 a_{2r} = 0 \tag{12.90}$$

If r = 1, then

$$\left(mgb + \kappa b^2 - \frac{g}{b} mb^2\right) a_{11} - \kappa b^2 a_{21} = 0$$

and

$$a_{11} = a_{21} \tag{12.91}$$

If r = 2, then

$$\left(mgb + \kappa b^2 - \frac{g}{b} mb^2 - \frac{2\kappa}{m} mb^2\right) a_{12} - \kappa b^2 a_{22} = 0$$

and

$$a_{12} = -a_{22} \tag{12.92}$$

We write the coordinates θ_1 and θ_2 in terms of the normal coordinates by

$$\left.\begin{aligned} \theta_1 &= a_{11}\eta_1 + a_{12}\eta_2 \\ \theta_2 &= a_{21}\eta_1 + a_{22}\eta_2 \end{aligned}\right\} \tag{12.93}$$

Using Equations 12.91 and 12.92, Equations 12.93 become

$$\left.\begin{aligned} \theta_1 &= a_{11}\eta_1 - a_{22}\eta_2 \\ \theta_2 &= a_{11}\eta_1 + a_{22}\eta_2 \end{aligned}\right\} \tag{12.94}$$

The normal modes are easily determined, by adding and subtracting θ_1 and θ_2 , to be

$$\left.\begin{aligned} \eta_1 &= \frac{1}{2a_{11}}(\theta_1 + \theta_2) \\ \eta_2 &= \frac{1}{2a_{22}}(\theta_2 - \theta_1) \end{aligned}\right\} \tag{12.95}$$

Because normal coordinate η_1 occurs when $\eta_2 = 0$, then $\theta_2 = \theta_1$ for normal mode 1 (symmetrical). Similarly, normal coordinate η_2 occurs when $\eta_1 = 0$ ($\theta_2 = -\theta_1$), and normal mode 2 is antisymmetrical. The normal mode motions are shown in Figure 12-6. Notice that for mode 1, the spring is neither compressed nor extended. The two pendula merely oscillate in unison with their natural frequencies ($\omega_1 = \omega_0 = \sqrt{g/b}$). These motions can be easily demonstrated in the laboratory or classroom. The higher frequency of normal mode 2 is easily displayed for a stiff spring.

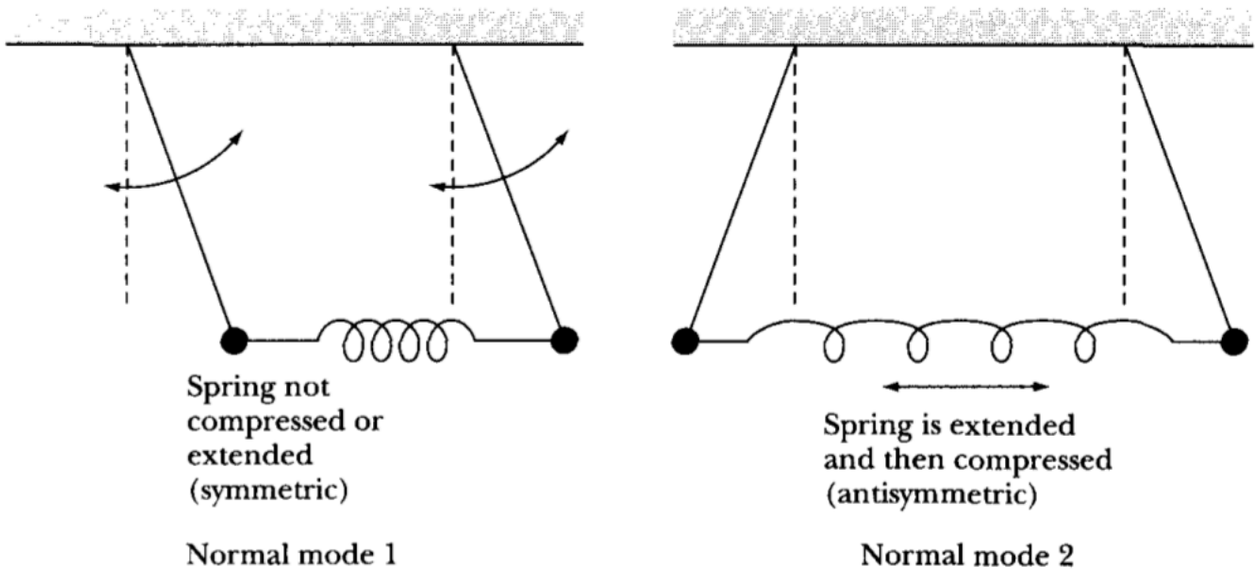


FIGURE 12-6 Example 12.4. The two normal mode motions are shown.

12.7 Molecular Vibrations

We mentioned previously that molecular vibrations are good examples of the applications of the small oscillations discussed in this chapter. A molecule containing n atoms generally has 3 n degrees of freedom. Three of these degrees of freedom are needed to describe the translational motion, and, generally, three are needed to describe rotations. Thus, there are 3n - 6 vibrational degrees of freedom. For molecules with collinear atoms, only two possible rotational degrees of freedom exist, because rotation about the axis through the atoms is insignificant. In this case, there are 3n - 5 degrees of freedom for vibrations.

We want to consider here only the vibrations occurring in a plane. We eliminate the translational and rotational degrees of freedom by appropriate transformations and choice of coordinate systems. For motion in a plane, there are 2n degrees of freedom. Because two are translational and one is rotational, generally 2n - 3 normal vibrations occur in the plane [leaving (3n - 6) - (2n - 3) = n - 3 degrees of freedom for vibrations of the atoms out of the plane].

Linear molecules may have both longitudinal and transverse vibrations. The longitudinal vibrations occur along the line of the atoms. For n atoms, there are n degrees of freedom along the line, but one of them corresponds to translation. Thus, there are n - 1 possible vibrations in the longitudinal direction for n atoms in a linear molecule. If a total of 3n - 5 vibrational degrees of freedom exist for a linear molecule, there must be (3n - 5) - (n - 1) = 2n - 4 transverse vibrations causing the atoms to vibrate perpendicular to the line of atoms. But from symmetry, any two mutually perpendicular directions suffice - so there are really only half the number of transverse frequencies, or n - 2.

EXAMPLE 12.5

Determine the eigenfrequencies and describe the normal mode motion of a symmetrical linear triatomic molecule (Figure 12-7) similar to CO_2. The central atom (carbon) has mass M, and the symmetrical atoms (oxygen) have masses m. Both longitudinal and transverse vibrations are possible.

Solution. For three atoms, the preceding analysis indicates that we have two longitudinal and one transverse vibrational degrees of freedom if we eliminate the translational and rotational degrees of freedom.

We can solve the longitudinal and transverse motions separately, because they are independent. In Figure 12-7b, we represent the atomic displacements from equilibrium by x_1, x_2, x_3. The elastic forces between atoms are represented by springs of force constant κ_j.

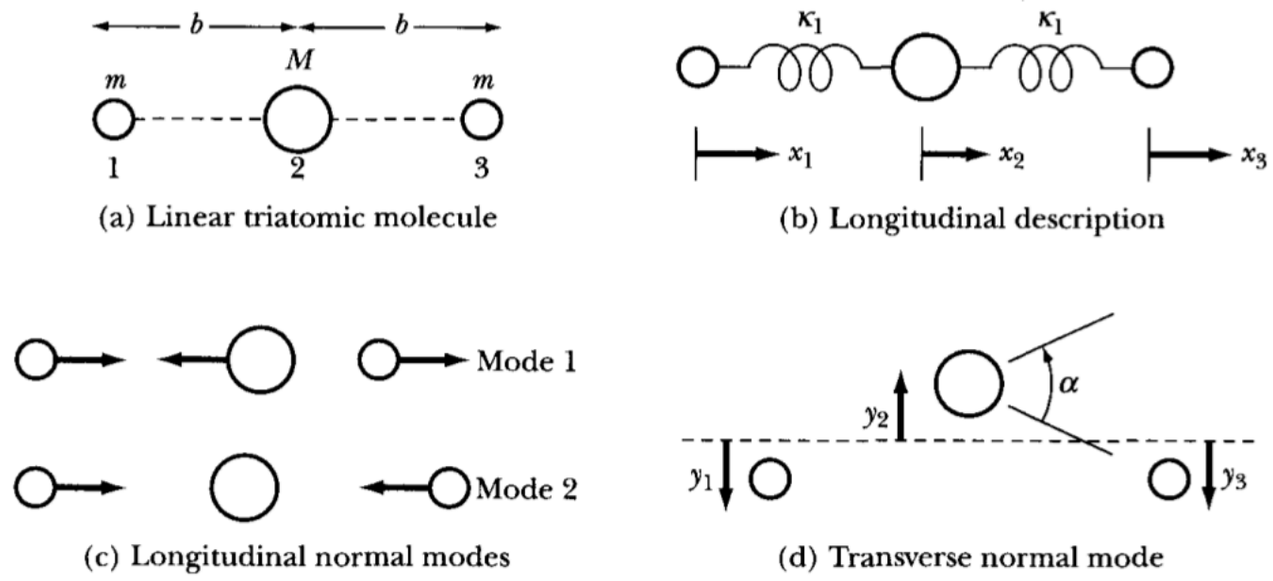


FIGURE 12-7 Example 12.5. (a) A linear triatomic molecule (for example, CO_2 with central mass M and symmetrical masses m. (b) The elastic forces between atoms are represented by springs; the atomic displacements from equilibrium are x_1, x_2, and x_3. (c) The two longitudinal normal modes. (d) The transverse normal mode.

We have three longitudinal variables but only two degrees of freedom. We must eliminate the translational possibility by requiring the center of mass to be constant during the vibrations. This is satisfied if

$$m(x_1 + x_3) + M(x_2) = 0 \tag{12.96}$$

Therefore, we can eliminate the variable x_2:

$$x_2 = -\frac{m}{M}(x_1 + x_3) \tag{12.97}$$

The kinetic energy becomes

$$\begin{aligned} T &= \frac{1}{2}m\dot{x}_1^2 + \frac{1}{2}m\dot{x}_3^2 + \frac{1}{2}M\dot{x}_2^2 \\ &= \frac{1}{2}m\dot{x}_1^2 + \frac{1}{2}m\dot{x}_3^2 + \frac{1}{2}\frac{m^2}{M}(\dot{x}_3^2 + \dot{x}_1^2 + 2\dot{x}_3\dot{x}_1) \end{aligned} \tag{12.98}$$

Having the $\dot{x}_3\dot{x}_1$ coupling term in the kinetic energy (called "dynamic coupling") can be inconvenient when solving Equation 12.42 for the eigenfrequencies.

We use a transformation to eliminate the dynamic coupling. Let

$$\left.\begin{aligned} q_1 &= x_3 + x_1 \\ q_2 &= x_3 - x_1 \end{aligned}\right\} \tag{12.99a}$$

Then

$$\left.\begin{aligned} x_3 &= \frac{1}{2}(q_1 + q_2) \\ x_1 &= \frac{1}{2}(q_1 - q_2) \end{aligned}\right\} \tag{12.99b}$$

and

$$x_2 = -\frac{m}{M} q_1$$

and the kinetic energy (Equation 12.98) becomes

$$T = \frac{m}{4}\dot{q}_2^2 + \frac{(mM + 2m^2)}{4M}\dot{q}_1^2 \tag{12.100}$$

The potential energy is

$$U = \frac{1}{2}\kappa_1(x_2 - x_1)^2 + \frac{1}{2}\kappa_1(x_3 - x_2)^2 \tag{12.101}$$

and with the transformations. Equations 12.99, the potential energy becomes (after considerable reduction)

$$U = \left(\frac{2m + M}{2M}\right)^2 \kappa_1 q_1^2 + \frac{1}{4}\kappa_1 q_2^2 \tag{12.102}$$

The eigenfrequencies are determined by inspection, using Equation 12.42

$$\begin{vmatrix} \frac{1}{2}\left(\frac{2m+M}{M}\right)^2\kappa_1 - \omega^2\left(\frac{mM+2m^2}{2M}\right) & 0 \\ 0 & \frac{\kappa_1}{2} - \omega^2\frac{m}{2} \end{vmatrix} = 0 \tag{12.103}$$

to be

$$\left.\begin{aligned} \omega_1^2 &= \frac{(2m+M)}{mM}\kappa_1 \\ \omega_2^2 &= \frac{\kappa_1}{m} \end{aligned}\right\} \tag{12.104}$$

Because the tensor formed by the coefficients of Equation 12.40 is already diagonal, the variables q_1 and q_2 represent the normal coordinates (unnormalized).

$$\left.\begin{aligned} q_1 &= a_{11}\eta_1 + a_{12}\eta_2 \\ q_2 &= a_{21}\eta_1 + a_{22}\eta_2 \end{aligned}\right\} \tag{12.105}$$

But

$$a_{12} = 0 \quad \text{and} \quad a_{21} = 0$$

$$q_1 = a_{11}\eta_1$$

$$q_2 = a_{22}\eta_2$$

As usual, we determine the motion of one normal mode when the other is zero. The descriptions of the longitudinal normal mode motion are given in Table 12-2. Normal mode 1 has the end atoms in symmetrical motion, but the central atom (from Equation 12.97) moves opposite to x_1 and x_3. Normal mode 2 has the end atoms vibrating antisymmetrically, but the central atom is at rest. This motion is displayed in Figure 12-7c.

TABLE 12-2 Longitudinal Normal Mode Motions

Mode	Eigenfrequencies	Variable	Motion
1	$\sqrt{\frac{(2m + M)}{mM}\kappa_1}$	$q_1 = x_3 + x_1$	$x_3 = x_1 \quad (q_2 = 0)$ $x_2 = -\frac{2m}{M}x_1$
2	$\sqrt{\frac{\kappa_1}{m}}$	$q_2 = x_3 - x_1$	$x_3 = -x_1 \quad (q_1 = 0)$ $x_2 = 0$

Because we have eliminated rotations in our system, the transverse vibrations must be as shown in Figure 12-7d, with the end atoms moving in phase (y_1=y_3) opposite to that of y_2. An equation similar to Equation 12.97 relates y_2 to y_1 and y_3 to keep the center of mass constant.

$$m(y_1 + y_3) + M(y_2) = 0 \tag{12.106}$$

$$y_2 = -\frac{m}{M}(y_1 + y_3) \tag{12.107}$$

We represent the single degree of freedom for the transverse vibration by the angle α representing the bending of the line of atoms. We assume α is small.

$$\alpha = \frac{(y_1 - y_2) + (y_3 - y_2)}{b}$$

The kinetic energy for the transverse mode is

$$T = \frac{1}{2}m(\dot{y}_1^2 + \dot{y}_3^2) + \frac{1}{2}M\dot{y}_2^2$$

Because y_1=y_3 and using Equation 12.107, α and T become

$$\alpha = \frac{2y_1}{bM}(2m + M) \tag{12.108}$$

$$T = \frac{m}{M}(M + 2m)\dot{y}_1^2$$

$$T = \frac{mMb^2}{4(2m + M)}\dot{\alpha}^2 \tag{12.109}$$

The potential energy represents the binding of the line of atoms. We assume the restoring force to be proportional to the total deviation from a straight line ($b\alpha$), so the potential energy is

$$U = \frac{1}{2}\kappa_2(b\alpha)^2 \tag{12.110}$$

Equations 12.109 and 12.110 are similar to those for the mass-spring, with the vibrational frequency determined to be

$$\omega_3^2 = \frac{2(M + 2m)}{mM}\kappa_2 \tag{12.111}$$

The transverse normal mode is represented by

$$y_1 = y_3 \tag{12.112}$$

$$y_2 = -\frac{m}{M}(y_1 + y_3) \tag{12.113}$$

as already discussed and shown in Figure 12-7d.

The CO_2 molecule is an example of the symmetrical linear molecule just discussed. Electromagnetic radiation resulting from the first and third normal modes is observed, because the electrical center of the molecule deviates from the center of mass (m: O^-; M: C^{++}). But no radiation emanates from normal mode 2, because the electrical center is coincident with the center of mass and thus the system has no dipole moment.

12.8 Three Linearly Coupled Plane Pendula - an Example of Degeneracy

EXAMPLE 12.6

Consider three identical pendula suspended from a slightly yielding support. Because the support is not rigid, a coupling occurs between the pendula, and energy can be transferred from one pendulum to the other Find the eigenfrequencies and eigenvectors and describe the normal mode motion. Figure 12-8 shows the geometry of the problem.

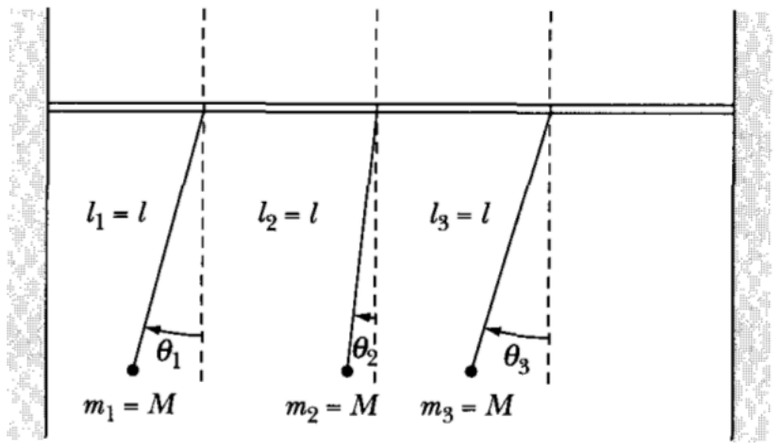


FIGURE 12-8 Example 12.6. Three identical pendula are suspended from a slightly yielding support that allows energy to be transferred between pendula. Such an experiment is easy to set up and demonstrate.

Solution. To simplify the notation, we adopt a system of units (sometimes called *natural units*) in which all lengths are measured in units of the length of the pendula l , all masses in units of the pendula masses M, and accelerations in units of g. Therefore, in our equations the values of the quantities M, l , and g are numerically equal to unity. If the coupling between each pair of the pendula is the same, we have

$$\left.\begin{aligned} T &= \frac{1}{2}(\dot{\theta}_1^2 + \dot{\theta}_2^2 + \dot{\theta}_3^2) \\ U &= \frac{1}{2}(\theta_1^2 + \theta_2^2 + \theta_3^2 - 2\varepsilon\theta_1\theta_2 - 2\varepsilon\theta_1\theta_3 - 2\varepsilon\theta_2\theta_3) \end{aligned}\right\} \tag{12.114}$$

Thus, the tensor $\{\mathbf{m}\}$ is diagonal.

$$\{\mathbf{m}\} = \begin{Bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Bmatrix} \tag{12.115}$$

but $\{\mathbf{A}\}$ has the form

$$\{\mathbf{A}\} = \begin{Bmatrix} 1 & -\varepsilon & -\varepsilon \\ -\varepsilon & 1 & -\varepsilon \\ -\varepsilon & -\varepsilon & 1 \end{Bmatrix} \tag{12.116}$$

The secular determinant is

$$\begin{vmatrix} 1-\omega^2 & -\varepsilon & -\varepsilon \\ -\varepsilon & 1-\omega^2 & -\varepsilon \\ -\varepsilon & -\varepsilon & 1-\omega^2 \end{vmatrix} = 0 \tag{12.117}$$

Expanding, we have

$$(1-\omega^2)^3 - 2\varepsilon^3 - 3\varepsilon^2(1-\omega^2) = 0$$

which can be factored to

$$(\omega^2 - 1 - \varepsilon)^2(\omega^2 - 1 + 2\varepsilon) = 0$$

and hence the roots are

$$\left.\begin{aligned} \omega_1 &= \sqrt{1+\varepsilon} \\ \omega_2 &= \sqrt{1+\varepsilon} \\ \omega_3 &= \sqrt{1-2\varepsilon} \end{aligned}\right\} \tag{12.118}$$

Notice that we have a double root: $\omega_1 = \omega_2 = \sqrt{1+\varepsilon}$. The normal modes corresponding to these frequencies are therefore *degenerate* - that is, these two modes are indistinguishable.

We now evaluate the quantities a_{jr} beginning with a_{j3}. Again we note that, because the equations of motion determine only the ratios, we need consider only two of the three available equations; the third equation is automatically satisfied. Using the equation

$$\sum_j (A_{jk} - \omega_3^2 m_{jk})\, a_{j3} = 0$$

we find

$$\left.\begin{aligned} 2\varepsilon a_{13} - \varepsilon a_{23} - \varepsilon a_{33} &= 0 \\ -\varepsilon a_{13} + 2\varepsilon a_{23} - \varepsilon a_{33} &= 0 \end{aligned}\right\} \tag{12.119}$$

Equations 12.119 yield

$$a_{13} = a_{23} = a_{33} \tag{12.120}$$

and from the normalization condition we have

$$a_{13}^2 + a_{23}^2 + a_{33}^2 = 1$$

or

$$a_{13} = a_{23} = a_{33} = \frac{1}{\sqrt{3}} \tag{12.121}$$

Thus, we find that for r = 3 there is no problem in evaluating the components of the eigenvector as. (This is a general rule: There is no indefiniteness in evaluating the eigenvector components for a nondegenerate mode.) Because all the components of $\mathbf{a}_3$ are equal, this corresponds to the mode in which all three pendula oscillate in phase.

Let us now attempt to evaluate the a_{j1} and a_{j2}. From the six possible equations of motion (three values of j and two values of r), we obtain only two different relations:

$$\varepsilon(a_{11} + a_{21} + a_{31}) = 0 \qquad *(12.122)$$

$$\varepsilon(a_{12} + a_{22} + a_{32}) = 0 \qquad *(12.123)$$

The orthogonality equation is

$$\sum_{j,k} m_{jk} a_{jr} a_{ks} = 0, \qquad r \neq s$$

but, because $m_{jk} = \delta_{jk}$ this becomes

$$\sum_j a_{jr} a_{js} = 0, \qquad r \neq s$$

which leads to only one new equation:

$$a_{11}a_{12} + a_{21}a_{22} + a_{31}a_{32} = 0 \qquad *(12.125)$$

(The other two possible equations are identical with Equations 12.122 and 12.123 above.) Finally, the normalization conditions yield

$$a_{11}^2 + a_{21}^2 + a_{31}^2 = 1 \qquad *(12.126)$$

$$a_{12}^2 + a_{22}^2 + a_{32}^2 = 1 \qquad *(12.127)$$

Thus, we have a total of only *five* (starred,*) equations for the six unknowns a_{j1} and a_{j2}. This indeterminacy in the eigenvectors corresponding to a double root is exactly the same as that encountered in constructing the principal axes for a rigid body with an axis of symmetry; the two equivalent principal axes may be placed in any direction as long as the set of three axes is orthogonal. Therefore, we are at liberty to arbitrarily specify the eigenvectors $\mathbf{a}_1$ and $\mathbf{a}_2$, as long as the orthogonality and normalizing relations are satisfied. For a simple system such as we are discussing, it should not be difficult to construct these vectors, so we do not give any general rules here.

If we arbitrarily choose $a_{31} = 0$, the indeterminacy is removed. We then find

$$\mathbf{a}_1 = \frac{1}{\sqrt{2}}(1, -1, 0), \quad \mathbf{a}_2 = \frac{1}{\sqrt{6}}(1, 1, -2) \qquad (12.128)$$

from which we can verify that the starred relations are all satisfied.

Recall that the nondegenerate mode corresponds to the in-phase oscillation of all three pendula:

$$\mathbf{a}_3 = \frac{1}{\sqrt{3}}(1, 1, 1) \tag{12.129}$$

We now see that the degenerate modes each correspond to out-of-phase oscillation. For example, $\mathbf{a}_2$ in Equation 12.128 represents two pendula oscillating together with a certain amplitude, whereas the third is out of phase and has twice the amplitude. Similarly, $\mathbf{a}_1$ in Equation 12.128 represents one pendulum stationary and the other two in out-of-phase oscillation. The eigenvectors $\mathbf{a}_1$ and $\mathbf{a}_2$ already given are only one set of an infinity of sets satisfying the conditions of the problem. But all such eigenvectors represent some sort of out-of-phase oscillation.

12.9 The Loaded String

We now consider a more complex system consisting of an elastic string (or a spring) on which a number of identical particles are placed at regular intervals. The ends of the string are constrained to remain stationary. Let the mass of each of the n particles be m, and let the spacing between particles at equilibrium be d. Thus, the length of the string is L = (n + l)d. The equilibrium situation'is shown in Figure 12-9.

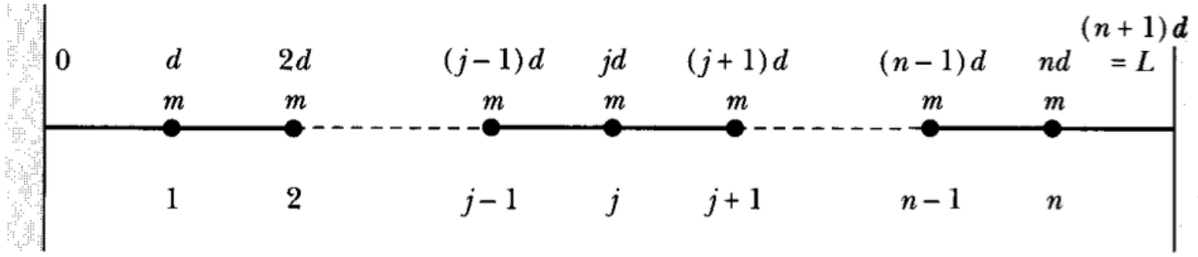


FIGURE 12-9 A schematic of the loaded string. In equilibrium, identical masses are spaced equidistantly. The ends of the string are fixed.

We wish to treat the case of small transverse oscillations of the particles about their equilibrium positions. First, we consider the vertical displacements of the masses numbered j - 1, j, and j + I (Figure 12-10).

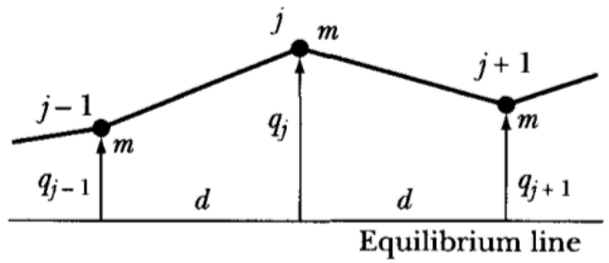


FIGURE 12-10 Vertical displacements (q_{j-1}, q_j, and q_{j+1}) of masses on the loaded string.

If the vertical displacements q_{j-1}, q_j and q_{j+1} are small, then the tension t in the string is approximately constant and equal to its value at equilibrium. For small displacements, the string section between any pair of particles makes only small angles with the equilibrium line. Approximating the sines of these angles by the tangents, the expression for the force that tends to restore the j^{th} particle to its equilibrium position is

$$F_j = -\frac{\tau}{d}(q_j - q_{j-1}) - \frac{\tau}{d}(q_j - q_{j+1}) \tag{12.130}$$

The force F_j is, according to Newton's law, equal to $m\ddot{q}_j$, Equation 12.130 can therefore be written as

$$\ddot{q}_j = \frac{\tau}{md}(q_{j-1} - 2q_j + q_{j+1}) \tag{12.131}$$

which is the equation of motion for the j^{th} particle. The system is coupled, because the force on the j^{th} particle depends on the positions of the $(j - 1)^{th}$ and $(j + 1)^{th}$ particles; this is therefore an example of nearest neighbor interaction, in which the coupling is only to the adjacent particles. It is not necessary that the interaction be confined to nearest neighbors. If the force between pairs of particles were electrostatic, for example, then each particle would be coupled to *all* the other particles. The problem can then become quite complicated. But even if the force is electrostatic, the $1/r^2$ dependence on distance frequently permits us to neglect interactions at distances greater than one interparticle spacing, so that the simple expression for the force given in Equation 12.130 is approximately correct.

We have considered only the motion perpendicular to the line of the string: transverse oscillations. It is easy to show that exactly the same form for the equations of motion results if we consider longitudinal vibrations - that is, motions along the line of the string. In this case, the factor τ/d is replaced by κ, the force constant of the string (see Problem 12-24).

Although we used Newton's equation to obtain the equations of motion (Equation 12.131), we may equally well use the Lagrangian method. The potential energy arises from the work done to stretch the n + 1 string segments (we consider the potential energy to be only the elastic energy in the string; that is, we do not consider the individual masses to have any gravitational (or any other) potential energy.):

$$U = \frac{\tau}{2d} \sum_{j=1}^{n+1} (q_{j-1} - q_j)^2 \tag{12.132}$$

where q_0 and q_{n+1} are identically zero, because these positions correspond to the fixed ends of the string. We note that Equation 12.132 yields an expression for the force on the j^{th} particle that is the same as the previous result (Equation 12.130):

$$F_j = -\frac{\partial U}{\partial q_j} = -\frac{\tau}{2d} \frac{\partial}{\partial q_j} [(q_{j-1} - q_j)^2 + (q_j - q_{j+1})^2] \tag{12.133}$$

$$= \frac{\tau}{d}(q_{j-1} - 2q_j + q_{j+1})$$

The kinetic energy is given by the sum of the kinetic energies of the n individual particles:

$$T = \frac{1}{2} m \sum_{j=1}^{n} \dot{q}_j^2 \tag{12.134}$$

Because $\dot{q}_{n+1} = 0$, we may extend the sum in Equation 12.134 to j = n + 1 so that the range of j is the same as that in the expression for the potential energy. Then, the Lagrangian becomes

$$L = \frac{1}{2}\sum_{j=1}^{n+1}\left[m\dot{q}_j^2 - \frac{\tau}{d}(q_{j-1} - q_j)^2\right] \tag{12.135}$$

It should be obvious that the equation of motion for the j[th] particle must arise from only those terms in the Lagrangian containing q_j or $\dot{q}_j$. If we expand the sum in L, we find

$$L = \cdots + \frac{1}{2}m\dot{q}_j^2 - \frac{1}{2}\frac{\tau}{d}(q_{j-1} - q_j)^2 - \frac{1}{2}\frac{\tau}{d}(q_j - q_{j+1})^2 - \cdots \tag{12.136}$$

where we have written only those terms that contain either q_j or $\dot{q}_j$. Applying Lagrange's equation for the coordinate q_j we have

$$m\ddot{q}_j - \frac{\tau}{d}(q_{j-1} - 2q_j + q_{j+1}) = 0 \tag{12.137}$$

Thus, the result is the same as that obtained by using the Newtonian method.

To solve the equations of motion, we substitute, as usual,

$$q_j(t) = a_j e^{i\omega t} \tag{12.138}$$

where a_j can be *complex*. Substituting this expression for $q_j(t)$ into Equation 12.137, we find

$$-\frac{\tau}{d}a_{j-1} + \left(2\frac{\tau}{d} - m\omega^2\right)a_j - \frac{\tau}{d}a_{j+1} = 0 \tag{12.139}$$

where j = 1, 2,...,n, but because the ends of the string are fixed, we must have $a_0 = a_{n+1} = 0$.

Equation 12.139 represents a **linear difference equation** that can be solved for the eigenfrequencies ω_r by setting the determinant of the coefficients equal to zero. We therefore have the following secular determinant:

$$\begin{vmatrix} \lambda & -\frac{\tau}{d} & 0 & 0 & 0 & \cdots \\ -\frac{\tau}{d} & \lambda & -\frac{\tau}{d} & 0 & 0 & \cdots \\ 0 & -\frac{\tau}{d} & \lambda & -\frac{\tau}{d} & 0 & \cdots \\ 0 & 0 & -\frac{\tau}{d} & \lambda & -\frac{\tau}{d} & \cdots \\ 0 & 0 & 0 & \cdot & \cdot & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{vmatrix} = 0 \tag{12.140}$$

where we have used

$$\lambda \equiv 2\frac{\tau}{d} - m\omega^2 \tag{12.141}$$

This secular determinant is a special case of the general determinant (Equation 12.42) that results if the tensor **m** is diagonal and the tensor **A** involves a coupling only between adjacent particles. Thus, Equation 12.140 consists only of diagonal elements plus elements once-removed from the diagonal

For the case n = 1 (i.e., a single mass suspended between two identical springs), we have $\lambda = 0$, or

$$\omega = \sqrt{\frac{2\tau}{md}}$$

We may adapt this result to the case of longitudinal motion by replacing τ/d by κ ; we then obtain the familiar expression.

$$\omega = \sqrt{\frac{2\kappa}{m}}$$

For the case n = 2, and with τ/d replaced by κ , we have $\lambda^2 = \kappa^2$, or

$$\omega = \sqrt{\frac{2\kappa \pm \kappa}{m}}$$

which are the same frequencies as those found in Section 12.2 for two coupled masses (Equation 12.8).

The secular equation should be relatively easy to solve directly for small values of n, but the solution becomes quite complicated for large n. In such cases, it is simpler to use the following method. We try a solution of the form

$$a_j = ae^{i(j\gamma - \delta)} \tag{12.142}$$

where a is *real* The use of this device is justified if we can find a quantity γ and a phase δ such that the conditions of the problem are all satisfied. Substituting a_j in this form into Equation 12.139 and canceling the phase factor, we find

$$-\frac{\tau}{d}e^{-i\gamma} + \left(2\frac{\tau}{d} - m\omega^2\right) - \frac{\tau}{d}e^{i\gamma} = 0$$

Solving for ω^2 , we obtain

$$\left.\begin{aligned} \omega^2 &= \frac{2\tau}{md} - \frac{\tau}{md}(e^{i\gamma} + e^{-i\gamma}) \\ &= \frac{2\tau}{md}(1 - \cos\gamma) \\ &= \frac{4\tau}{md}\sin^2\frac{\gamma}{2} \end{aligned}\right\} \tag{12.143}$$

Because we know that the secular determinant is of order n and therefore yields exactly n values for ω^2 , we can write

$$\omega_r = 2\sqrt{\frac{\tau}{md}} \sin\frac{\gamma_r}{2}, \qquad r = 1, 2, \ldots, n \tag{12.144}$$

We now evaluate the quantity γ_r and the phase δ_r by applying the boundary condition that the ends of the string remain fixed. Thus, we have

$$a_{jr} = a_r e^{i(j\gamma_r - \delta_r)} \tag{12.145}$$

or, because it is only the real part that is physically meaningful,

$$a_{jr} = a_r \cos(j\gamma_r - \delta_r) \tag{12.146}$$

The boundary condition is

$$a_{0r} = a_{(n+1)r} \equiv 0 \tag{12.147}$$

For Equation 12.146 to yield $a_{jr} = 0$ for $j = 0$, it should be clear that δ_r must be $\pi/2$ (or some odd integer multiple thereof). Hence,

$$\begin{aligned} a_{jr} &= a_r \cos\left(j\gamma_r - \frac{\pi}{2}\right) \\ &= a_r \sin j\gamma_r \end{aligned} \tag{12.148}$$

For $j = n + 1$, we have

$$a_{(n+1)r} = 0 = a_r \sin(n + 1)\gamma_r$$

Therefore,

$$(n + 1)\gamma_r = s\pi, \qquad s = 1, 2, \ldots$$

or

$$\gamma_r = \frac{s\pi}{n + 1}, \qquad s = 1, 2, \ldots$$

But there are just n distinct values of γ_r because Equation 12.144 requires n distinct values of ω_r . Therefore, the index s runs from 1 to n. Because there is a one-to-one correspondence between the values of s and the values of r, we can simply replace s in this last expression by the index r:

$$\gamma_r = \frac{r\pi}{n+1}, \qquad r = 1, 2, \ldots, n \tag{12.149}$$

The a_{jr} (Equation 12.148) then becomes

$$\boxed{a_{jr} = a_r \sin\left(j\frac{r\pi}{n+1}\right)} \tag{12.150}$$

The general solution for q_j (see Equation 12.61) is

$$\begin{aligned} q_j &= \sum_r \beta'_r a_{jr} e^{i\omega_r t} \\ &= \sum_r \beta'_r a_r \sin\left(j\frac{r\pi}{n+1}\right) e^{i\omega_r t} \\ &= \sum_r \beta_r \sin\left(j\frac{r\pi}{n+1}\right) e^{i\omega_r t} \end{aligned} \tag{12.151}$$

where we have written $\beta_r = \beta'_r a_r$. Furthermore, for the frequency we have

$$\boxed{\omega_r = 2\sqrt{\frac{\tau}{md}} \sin\left(\frac{r\pi}{2(n+1)}\right)} \tag{12.152}$$

We note that this expression yields the same results found for the case of two coupled oscillators (Equations 12.8) when we insert n = 2, r = 1, 2 and replace τ/d by $\kappa(=\kappa_{12})$.

Notice also that if either r=Oorr= n+ lis substituted into Equation 12.150, then all the Jimplitude factors Uj^ vanish identically. These values of r therefore refer to null modes. Moreover, if r takes on the values w + 2, w + 3,..., 2w + 1, then the a- are the same (except for a trivial sign change and in reverse order) as for r = 1, 2,..., n; also, r = 2w + 2 yields the next null mode. We conclude, therefore, that there are indeed only n distinct modes and that increasing r beyond n merely duplicates the modes for smaller n. (A similar argument applies for r < 0.) These conclusions are illustrated in Figure 12-11 for the case n = 3.

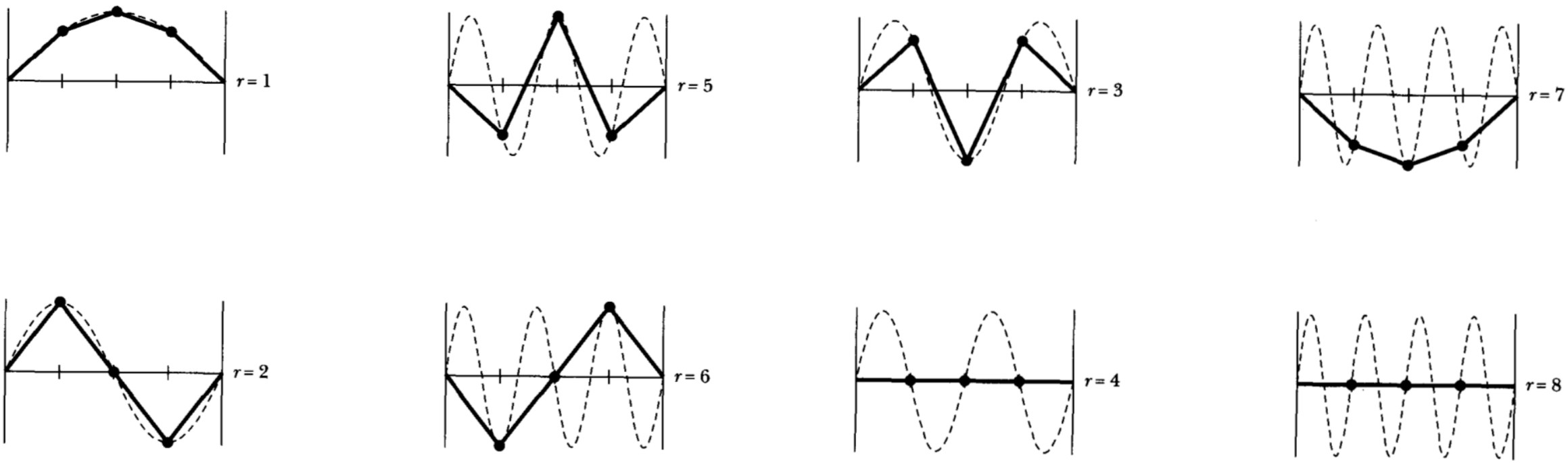


FIGURE 12-11 The normal mode motion for the case of $n = 3$ masses. Only $r = 1, 2, 3$ are distinct modes, because $r = 4$ is a null mode, and $r = 7, 6, 5, 8$ are duplicates of 1, 2, 3, and 4, respectively, with a change in sign. The dashed curves represent the sinusoidal behavior of the amplitude factors a_{jr} and are not physically meaningful features of the motion.

The distinct modes are specified by r = 1, 2, 3; r = 4 is a null mode. The displacement patterns are duplicated for r = 7, 6, 5, 8, but with a change in sign. In Figure 12-11, the dashed curves merely represent the sinusoidal behavior of the amplitude factors a_{jr} for various values of r; the only physically meaningful features of these curves are the values at the positions occupied by the particles (j= 1,2,3). The "high frequency" of the sine curves for r = 5, 6, 7, 8 is thus not at all related to the frequency of the particles' motions; these latter frequencies are the same as for r = 1,2, 3, 4.

The normal coordinates of the system (Equation 12.62) are

$$\eta_r(t) \equiv \beta_r e^{i\omega_r t} \tag{12.153}$$

so that

$$q_j(t) = \sum_r \eta_r \sin\left(j\frac{r\pi}{n+1}\right) \tag{12.154}$$

This equation for q_j is similar to the pre\ious expression (Equation 12.63) except that the quantities a_{jr} are now replaced by $\sin[j(r\pi)/(n+1)]$.

Because β_r may be complex, we write for the real part of q_j,

$$real: \quad q_j(t) = \sum_r \sin\left(j\frac{r\pi}{n+1}\right)(\mu_r \cos\omega_r t - \nu_r \sin\omega_r t) \tag{12.155}$$

where

$$\beta_r = \mu_r + i\nu_r \tag{12.156}$$

The initial value of $q_j(t)$ can be obtained from Equation 12.155:

$$q_j(0) = \sum_r \mu_r \sin\left(j\frac{r\pi}{n+1}\right) \tag{12.157}$$

$$\dot{q}_j(0) = -\sum_r \omega_r \nu_r \sin\left(j\frac{r\pi}{n+1}\right) \tag{12.158}$$

If we multiply Equation 12.157 by $\sin[j(s\pi)/(n+1)]$ and sum over j, we find

$$\sum_j q_j(0) \sin\left(j\frac{s\pi}{n+1}\right) = \sum_{j,r} \mu_r \sin\left(j\frac{r\pi}{n+1}\right)\sin\left(j\frac{s\pi}{n+1}\right) \tag{12.159}$$

A relationship in the form of a trigonometric identity is available for the sine terms:

$$\sum_{j=1}^{n} \sin\left(j\frac{r\pi}{n+1}\right)\sin\left(j\frac{s\pi}{n+1}\right) = \frac{n+1}{2}\delta_{rs}, \qquad r, s = 1, 2, \ldots, n$$

so that Equation 12.159 becomes

$$\begin{aligned}\sum_j q_j(0)\sin\left(j\frac{s\pi}{n+1}\right) &= \sum_r \mu_r \frac{n+1}{2}\delta_{rs} \\ &= \frac{n+1}{2}\mu_s\end{aligned} \tag{12.160}$$

or

$$\mu_s = \frac{2}{n+1}\sum_j q_j(0)\sin\left(j\frac{s\pi}{n+1}\right) \tag{12.161a}$$

A similar procedure for ν_s yields

$$\nu_s = -\frac{2}{\omega_s(n+1)}\sum_j \dot{q}_j(0)\sin\left(j\frac{s\pi}{n+1}\right) \tag{12.161b}$$

Thus, we have evaluated all the necessary quantities, and the description of the vibrations of a loaded string is therefore complete.

We should note the following point regjirding the normalization procedures used here. First, in Equation 12.57 we arbitrarily normalized the a_{jr} to unity. Thus, the a_{jr} are required to be independent of the initial conditions imposed on the system. The scale factors α_r and β_r then allowed the magnitude of the oscillations to be varied by the selection of the initial conditions. Next, in the problem of the loaded string, we found that instead of the quantities a_{jr}, there arose the sine functions $\sin[j(r\pi)/(n+1)]$, and these functions possess a normalization property (Equation 12.160) that is specified by trigonometric identities. Therefore, in this case it is not possible arbitrarily to impose a normalization condition; we are automatically presented with the condition. But this is no restriction; it means only that the scale factors β for this case have a slightly different form. Thus, there are certain constants that occur in the two problems that, for convenience, are separated in different ways in the two cases.

EXAMPLE 12.7

Consider a loaded string consisting of three particles regularly spaced on the string. At t= 0 the center particle (only) is displaced a distance a and released from rest. Describe the subsequent motion.

Solution. The initial conditions are

$$\left.\begin{aligned} q_2(0) = a, \quad q_1(0) = q_3(0) = 0 \\ \dot{q}_1(0) = \dot{q}_2(0) = \dot{q}_3(0) = 0 \end{aligned}\right\} \tag{12.162}$$

Because the initial velocities are zero, the v_r vanish. The μ_r are given by (Equation 12.161a):

$$\mu_r = \frac{2}{n+1}\sum_j q_j(0)\sin\left(j\frac{r\pi}{n+1}\right)$$

$$= \frac{1}{2}a\sin\left(\frac{r\pi}{2}\right) \tag{12.163}$$

because only the term j = 2 contributes to the sum. Thus,

$$\mu_1 = \frac{1}{2}a, \qquad \mu_2 = 0, \qquad \mu_3 = -\frac{1}{2}a \tag{12.164}$$

The quantities $\sin[j(r\pi)/(n+1)]$ that appear in the expression for $q_j(t)$ (Equation 12.155) are

$j =$ \ $r = 1$	2	3
1	$\frac{\sqrt{2}}{2}$	1	$\frac{\sqrt{2}}{2}$
2	1	0	−1
3	$\frac{\sqrt{2}}{2}$	−1	$\frac{\sqrt{2}}{2}$

(12.165)

The displacements of the three particles therefore are

$$\left.\begin{aligned} q_1(t) &= \frac{\sqrt{2}}{4}a(\cos\omega_1 t - \cos\omega_3 t) \\ q_2(t) &= \frac{1}{2}a(\cos\omega_1 t + \cos\omega_3 t) \\ q_3(t) &= \frac{\sqrt{2}}{4}a(\cos\omega_1 t - \cos\omega_3 t) = q_1(t) \end{aligned}\right\} \tag{12.166}$$

where the characteristic frequencies are given by Equation 12.152:

$$\omega_r = 2\sqrt{\frac{\tau}{md}}\sin\left(\frac{r\pi}{8}\right), \qquad r = 1, 2, 3 \tag{12.167}$$

Notice that because the *middle* particle was initially displaced, no vibration mode occurs in which this particle is at rest; that is, mode 2 with frequency ω_2 (see Figure 12-11) is absent.

CHAPTER 13 - Continuous Systems; Waves

13.1 Introduction

We have so far been considering particles, systems of particles, or rigid bodies. Now, we want to consider bodies (gases, liquids, or solids) that are not rigid, that is, bodies whose particles move (however slightly) with respect to one another. The general study of such bodies is quite complex. However, one aspect of these continuous bodies is very important throughout physics—the ability to transmit wave motion. A disturbance on one part of the body can be transmitted by wave propagation throughout the body.

The simplest example of such phenomena is a vibrating suing stretched under uniform tension between two fixed supports. As usual, the simple example represents many of the important results needed to understand other physical examples, such as stretched membranes and waves in solids. Waves may be either transverse or longitudinal. An example of a longitudinal wave is the vibration of molecules along the direction of propagation of a wave moving in a solid rod. Longitudinal waves occur in fluids and solids and are of great importance in acoustics.

Whereas both transverse and longitudinal waves may occur in solids, only longitudinal waves occur inside fluids, in which shearing forces are not possible. We have already considered (Chapter 12) both kinds of vibrations for a system of particles. A detailed study of the transverse vibrating string is important for several reasons. A study of a one-dimensional model of such string vibrations allows a mathematical solution with results that are applicable to more complex two-and three-dimensional problems. The modes of oscillation are similar. In particular, the application of boundary conditions (fixed ends), which are of extreme importance in many areas of physics, is easiest in one-dimensional problems. Boundary conditions play a role in the use of partial differential equations similar to the role initial conditions play in ordinary differential equations using Newtonian or Lagrangian techniques.

In this chapter, we extend the discussion of the vibrations of a loaded string presented in Chapter 12 by examining the consequences of allowing the number of particles on the string to become infinite (while maintaining a constant linear mass density). In this way, we pass to the case of a **continuous string**. All the results of interest for such a string can be obtained by this limiting process - including the derivation of the important **wave equation**, one of the truly fundamental equations of mathematical physics.

The solutions of the wave equation are in general subject to limitations imposed by certain physical restrictions peculiar to a given problem. These limitations frequently take the form of conditions on the solution that must be met at the extremes of the intervals of space and time that are of physical interest. We must therefore deal with a **boundary-value problem** involving a partial differential equation. Indeed, such a description characterizes essentially the whole of what we call *mathematical physics*.

We confine ourselves here to the solution to a one-dimensional wave equation. Such waves can describe a two-dimensional wave in two dimensions and can describe, for example, the motion of a vibrating string. The compression (or sound) waves that may be transmitted through an elastic medium, such as a gas, can also be approximated as one-dimensional waves if the medium is large enough that the edge effects are unimportant. In such a case, the condition of the medium is approximately the same at every point on a plane, and the properties of the wave motion are then functions only of the distance along a line normal to the plane. Such a wave in an extended medium, called a *plane wave*, is mathematically identical to the one-dimensional waves treated here.

13.2 Continuous String as a Limiting Case of the Loaded String

In the preceding chapter, we considered a set of equally spaced point masses suspended by a string. We now wish to allow the number of masses to become infinite so that we have a continuous string. To do this, we must require that as $n \to \infty$ we simultaneously let the mass of each particle and the distance between each particle approach zero ($m \to 0, d \to 0$) in such a manner that the ratio m/d remains constant. We note that m/d = ρ is just the linear mass density of the string.

Thus, we have

$$\left.\begin{aligned} &n \to \infty, \quad d \to 0, \quad \text{such that } (n+1)\,d = L \\ &m \to 0, \quad d \to 0, \quad \text{such that } \frac{m}{d} = \rho = \text{constant} \end{aligned}\right\} \tag{13.1}$$

From Equation 12.154, we have

$$q_j(t) = \sum_r \eta_r(t) \sin\left(j \frac{r\pi}{n+1}\right) \tag{13.2}$$

We can now write

$$j \frac{r\pi}{n+1} = r\pi \frac{jd}{(n+1)\,d} = r\pi \frac{x}{L} \tag{13.3}$$

where jd = x now specifies the distance along the continuous string. Thus, $q_j(t)$ becomes a continuous function of the variables x and t:

$$q(x, t) = \sum_r \eta_r(t) \sin\left(\frac{r\pi x}{L}\right) \tag{13.4}$$

or

$$q(x, t) = \sum_r \beta_r e^{i\omega_r t} \sin\left(\frac{r\pi x}{L}\right) \tag{13.5}$$

In the case of a loaded string containing n particles, there are n degrees of freedom of motion and therefore n normal modes and n characteristic frequencies. Thus, in Equation 12.154 (or Equation 13.2) the sum is over the range r = 1 to r = n. But now the number of particles is infinite, so there is an infinite set of normal modes and the sum in Equations 13.4 and 13.5 runs from r = 1 to r = ∞ . There are, then, infinitely many constants (the real and imaginary parts of the β_r) that must be evaluated to completely specify the motion of the continuous string.

This is exactly the situation encountered in representing some function as a Fourier series - the infinitely many constants are specified by certain integrals involving the original function (see Equations 3.91). We may view the situation in another way: There are infinitely many arbitrary constants in the solution of the equation of motion, but there are also infinitely many initial conditions available for their evaluation, namely, the continuous functions q(x, 0) and $\dot{q}$(x, 0). The real and imaginary parts of the β_r, can thus be obtained in terms of the initial conditions by a procedure analogous to that used in Section 12.9. Using $\beta_r = \mu_r + i\nu_r$, we have from Equation 13.5,

$$q(x,0) = \sum_r \mu_r \sin\left(\frac{r\pi x}{L}\right) \tag{13.6a}$$

$$\dot{q}(x,0) = -\sum_r \omega_r \nu_r \sin\left(\frac{r\pi x}{L}\right) \tag{13.6b}$$

Next, we multiply each of these equations by $\sin(s\pi x/L)$ and integrate from x= 0 to x = L. We can make use of the trigonometric relation

$$\int_0^L \sin\left(\frac{r\pi x}{L}\right)\sin\left(\frac{s\pi x}{L}\right)dx = \frac{L}{2}\delta_{rs} \tag{13.7}$$

from which we obtain

$$\mu_r = \frac{2}{L}\int_0^L q(x,0)\sin\left(\frac{r\pi x}{L}\right)dx \tag{13.8a}$$

$$\nu_r = -\frac{2}{\omega_r L}\int_0^L \dot{q}(x,0)\sin\left(\frac{r\pi x}{L}\right)dx \tag{13.8b}$$

The characteristic frequency ω_r may also be obtained as the limiting value of the result for the loaded string. From Equation 12.152, we have

$$\omega_r = 2\sqrt{\frac{\tau}{md}}\,\sin\left[\frac{r\pi}{2(n+1)}\right] \tag{13.9}$$

which can be written as

$$\omega_r = \frac{2}{d}\sqrt{\frac{\tau}{\rho}}\,\sin\left(\frac{r\pi d}{2L}\right) \tag{13.10}$$

When $d \to 0$, we can approximate the sine term by its argument, with the result

$$\omega_r = \frac{r\pi}{L}\sqrt{\frac{\tau}{\rho}} \tag{13.11}$$

EXAMPLE 13.1

Find the displacement q(x, t) for a "plucked string," where one point of the string is displaced (such that the string assumes a triangular shape) and then released from rest. Consider the case shown in Figure 13-1, in which the center of the string is displaced a distance h.

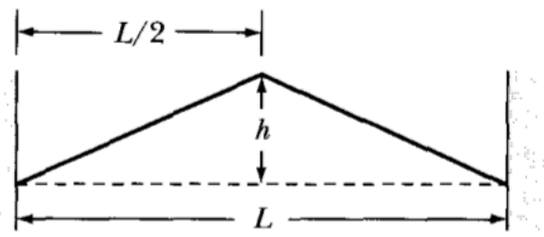


FIGURE 13-1 Example 13.1. A string is "plucked" by pulling the center of the string a distance h from equilibrium so that the string has a triangular shape. The string is released from rest in this position.

Solution. The initial conditions are

$$q(x,0) = \begin{cases} \dfrac{2h}{L}x, & 0 \le x \le L/2 \\ \dfrac{2h}{L}(L-x), & L/2 \le x \le L \end{cases} \tag{13.12}$$

$$\dot{q}(x,0) = 0$$

Because the string is released from rest, all the ν_r vanish. The μ_r are given by

$$\mu_r = \frac{4h}{L^2}\int_0^{L/2} x \sin\left(\frac{r\pi x}{L}\right)dx + \frac{4h}{L^2}\int_{L/2}^{L} (L - x)\sin\left(\frac{r\pi x}{L}\right)dx$$

Integrating,

$$\mu_r = \frac{8h}{r^2\pi^2}\sin\frac{r\pi}{2}$$

so that

$$\mu_r = \begin{cases} 0, & r \text{ even} \\ \dfrac{8h}{r^2\pi^2}(-1)^{\frac{1}{2}(r-1)}, & r \text{ odd} \end{cases}$$

Therefore,

$$q(x, t) = \frac{8h}{\pi^2}\left[\sin\left(\frac{\pi x}{L}\right)\cos\omega_1 t - \frac{1}{9}\sin\left(\frac{3\pi x}{L}\right)\cos\omega_3 t + \cdots\right] \tag{13.13}$$

where the ω_r are proportional to r and are given by Equation 13.11.

From Equation 13.13, we see that the fundamental mode (with frequency ω_1) and all the *odd* harmonics (with frequencies ω_3 , ω_5 , etc.) are excited but that none of the even harmonics are involved in the motion. Because the initial displacement was symmetrical, the subsequent motion must also be symmetrical, so none of the even modes (for which the center position of the string is a node) are excited. In general, if the string is plucked at some arbitrary point, none of the harmonics with nodes at that point will be excited.

As we prove in the next section, the energy in each of the excited modes is proportional to the square of the coefficient of the corresponding term in Equation 13.13. Thus, the energy ratios for the fundamental, third harmonic, fifth harmonic, and so on are $1 : \frac{1}{81} : \frac{1}{625} : \cdots$. Therefore the energy in the system (or the intensity of the emitted sound) is dominated by the fundamental. The third harmonic is 19 dB down from the fundamental and the fifth harmonic is down by 28 dB.

The decibel (dB) is a unit of relative sound intensity (or acoustic power). The intensity ratio of a sound with intensity I to a sound with intensity I_0 is given by $10 \log(I/I_0)$ dB. Thus, for the fundamental (I_0) and third harmonic (I), we have $10 \log (1/81) = -19.1$ dB or "19 dB down" in intensity. A ratio of 3 dB corresponds approximately to a factor of two in relative intensity.

13.3 Energy of a Vibrating String

Because we have made the assumption that frictional forces are not present, the total energy of a vibrating string must remain constant. We now show this explicitly; moreover, we show that the energy of the string is expressed simply as the sum of contributions from each of the normal modes. According to Equation 13.4, the displacement of the string is given by

$$q(x, t) = \sum_r \eta_r(t) \sin\left(\frac{r\pi x}{L}\right) \tag{13.14}$$

where the normal coordinates are

$$\eta_r(t) = \beta_r e^{i\omega_r t} \tag{13.15}$$

As always, the β_r are complex quantities and the physically meaningful normal coordinates are obtained by taking the *real part* of Equation 13.15.

The kinetic energy of the string is obtained by calculating the kinetic energy for an element of the string, $\frac{1}{2}(\rho dx)\dot{q}^2$, and then integrating over the length. Thus,

$$T = \frac{1}{2}\rho \int_0^L \left(\frac{\partial q}{\partial t}\right)^2 dx \tag{13.16}$$

or, using Equation 13.14,

$$T = \frac{1}{2}\rho \int_0^L \left[\sum_r \dot{\eta}_r \sin\left(\frac{r\pi x}{L}\right)\right]^2 dx \tag{13.17}$$

The square of the series can be expressed as a double sum, this technique ensuring that all cross terms are properly included:

$$T = \frac{1}{2}\rho \sum_{r,s} \dot{\eta}_r \dot{\eta}_s \int_0^L \sin\left(\frac{r\pi x}{L}\right) \sin\left(\frac{s\pi x}{L}\right) dx \tag{13.18}$$

The integral is now the same as that in Equation 13.7, so

$$T = \frac{\rho L}{4} \sum_{r,s} \dot{\eta}_r \dot{\eta}_s \delta_{rs}$$

$$= \frac{\rho L}{4} \sum_{r} \dot{\eta}_r^2 \tag{13.19}$$

In the evaluation of the kinetic energy, we must be careful to take the product of real quantities. We must therefore compute the square of the real part of $\dot{\eta}_r$:

$$(\mathrm{Re}\ \dot{\eta}_r)^2 = \left(\mathrm{Re}\ \frac{d}{dt}\left[(\mu_r + i\nu_r)(\cos \omega_r t + i \sin \omega_r t)\right]\right)^2$$

$$= (-\omega_r \mu_r \sin \omega_r t - \omega_r \nu_r \cos \omega_r t)^2$$

The kinetic energy of the string is therefore

$$T = \frac{\rho L}{4} \sum_{r} \omega_r^2 (\mu_r \sin \omega_r t + \nu_r \cos \omega_r t)^2 \tag{13.20}$$

The potential energy of the string can be calculated easily by writing down the expression for the loaded string and then passing to the limit of a continuous string. (Recall that we consider the potential energy to be only the elastic energy in the string.) For the loaded string.

$$U = \frac{1}{2} \frac{\tau}{d} \sum_{j} (q_{j-1} - q_j)^2$$

Multiplying and dividing by d,

$$U = \frac{1}{2}\tau\sum_j\left(\frac{q_{j-1} - q_j}{d}\right)^2 d$$

In passing to the limit, $d \to 0$, the term in parentheses becomes just the partial derivative of q(x, t) with respect to x, and the sum (including the factor d) becomes an integral:

$$U = \frac{1}{2}\tau\int_0^L\left(\frac{\partial q}{\partial x}\right)^2 dx \tag{13.21}$$

Using Equation 13.14, we have

$$\frac{\partial q}{\partial x} = \sum_r \frac{r\pi}{L}\eta_r \cos\left(\frac{r\pi x}{L}\right) \tag{13.22}$$

so that

$$U = \frac{1}{2}\tau\int_0^L\left[\sum_r \frac{r\pi}{L}\eta_r \cos\left(\frac{r\pi x}{L}\right)\right]^2 dx \tag{13.23}$$

Again, the squared term can be written as a double sum, and because the trigonometric relation (Equation 13.7) applies for cosines as well as sines, we have

$$\begin{aligned} U &= \frac{\tau}{2}\sum_{r,s}\frac{r\pi}{L}\frac{s\pi}{L}\eta_r\eta_s\int_0^L \cos\left(\frac{r\pi x}{L}\right)\cos\left(\frac{s\pi x}{L}\right)dx \\ &= \frac{\tau}{2}\sum_{r,s}\frac{r\pi}{L}\frac{s\pi}{L}\eta_r\eta_s\cdot\frac{L}{2}\delta_{rs} \\ &= \frac{\tau}{2}\sum_r\frac{r^2\pi^2}{L^2}\cdot\frac{L}{2}\eta_r^2 \\ &= \frac{\rho L}{4}\sum_r \omega_r^2\eta_r^2 \end{aligned} \tag{13.24}$$

where Equation 13.11 has been used in the last line to express the result in terms of ω_r^2 . Evaluating the square of the real part of η_r , we have, finally.

$$U = \frac{\rho L}{4} \sum_r \omega_r^2 (\mu_r \cos \omega_r t - \nu_r \sin \omega_r t)^2 \tag{13.25}$$

The total energy is now obtained by adding Equations 13.20 and 13.25, in which the cross terms cancel and the squared terms add to unity:

$$\begin{aligned} E &= T + U \\ &= \frac{\rho L}{4} \sum_r \omega_r^2 (\mu_r^2 + \nu_r^2) \end{aligned} \tag{13.26a}$$

or

$$E = \frac{\rho L}{4} \sum_r \omega_r^2 |\beta_r|^2 \tag{13.26b}$$

The total energy is therefore constant in time and, furthermore, is given by a sum of contributions from each of the normal modes.

The kinetic and potential energies each vary with time, so it is sometimes useful to calculate the time-averaged kinetic and potential energies - that is, the averages over one complete period of the fundamental vibration r = 1:

$$\langle T \rangle = \frac{\rho L}{4} \sum_r \omega_r^2 \langle (\mu_r \sin \omega_r t + \nu_r \cos \omega_r t)^2 \rangle \tag{13.27}$$

where the slanted brackets denote an average over the time interval $2\pi/\omega_1$. The averages of $\sin^2 \omega_1 t$ or $\cos^2 \omega_1 t$ over this interval are equal to $\frac{1}{2}$. Similarly, the averages of $\sin^2 \omega_r t$ and $\cos^2 \omega_r t$ for r ≥ 2 are also $\frac{1}{2}$, because the period of the fundamental vibration is always some integer times the period of a higher harmonic vibration. The averages of the cross terms, $\cos \omega_r t \sin \omega_r t$, all vanish. Therefore,

$$\langle T\rangle = \frac{\rho L}{8}\sum_r \omega_r^2(\mu_r^2 + \nu_r^2)$$

$$= \frac{\rho L}{8}\sum_r \omega_r^2|\beta_r|^2 \tag{13.28}$$

For the time-averaged potential energy, we have a similar result:

$$\langle U\rangle = \frac{\rho L}{4}\sum_r \omega_r^2\langle(\mu_r \cos\,\omega_r t - \nu_r \sin\,\omega_r t)^2\rangle$$

$$= \frac{\rho L}{8}\sum_r \omega_r^2(\mu_r^2 + \nu_r^2)$$

$$= \frac{\rho L}{8}\sum_r \omega_r^2|\beta_r|^2 \tag{13.29}$$

We therefore have the important result that the *average kinetic energy of a vibrating string is equal to the average potential energy* (which also follows from the virial theorem.

$$\langle T\rangle = \langle U\rangle \tag{13.30}$$

Notice also the simplification that results from the use of normal coordinates: Both $\langle T\rangle$ and $\langle U\rangle$ are simple sums of contributions from each of the normal modes.

13.4 Wave Equation

Our procedure thus far has been to describe the motion of a continuous string as the limiting case of the loaded string for which we have a complete solution; we have not yet written down the fundamental equation of motion for the continuous case. We may accomplish this by returning to the loaded string and again using the limit technique - but now on the equation of motion rather than on the solution. Equation 12.131 can be expressed as

$$\frac{m}{d}\ddot{q}_j = \frac{\tau}{d}\left(\frac{q_{j-1} - q_j}{d}\right) - \frac{\tau}{d}\left(\frac{q_j - q_{j+1}}{d}\right) \tag{13.31}$$

As d approaches zero, we have

$$\frac{q_j - q_{j+1}}{d} \rightarrow \frac{q(x) - q(x+d)}{d} \rightarrow -\left.\frac{\partial q}{\partial x}\right|_{x+d/2}$$

which is the derivative at x + d/2. For the other term in Equation 13.31, we have

$$\frac{q_{j-1} - q_j}{d} \rightarrow \frac{q(x-d) - q(x)}{d} \rightarrow -\left.\frac{\partial q}{\partial x}\right|_{x-d/2}$$

which is the derivative at x - d/2. The limiting value of the right-hand side of Equation 13.31 is therefore

$$\lim_{d\to 0} \tau\left(\frac{\left.\frac{\partial q}{\partial x}\right|_{x+d/2} - \left.\frac{\partial q}{\partial x}\right|_{x-d/2}}{d}\right) = \tau\left.\frac{\partial^2 q}{\partial x^2}\right|_x = \tau\frac{\partial^2 q}{\partial x^2}$$

Also in the limit, m/d becomes ρ, so the equation of motion is

$$\rho\ddot{q} = \tau\frac{\partial^2 q}{\partial x^2} \tag{13.32}$$

or

$$\boxed{\frac{\partial^2 q}{\partial x^2} = \frac{\rho}{\tau}\frac{\partial^2 q}{\partial t^2}} \tag{13.33}$$

This is the wave equation in one dimension. In Section 13.6, we shall discuss the solutions to this equation.

We now want to show that Equation 13.33 can also be easily obtained by considering the forces on a continuous string. Only transverse waves are considered. A portion of the string fixed at both ends, as discussed so far in this chapter, is shown in Figure 13-2.

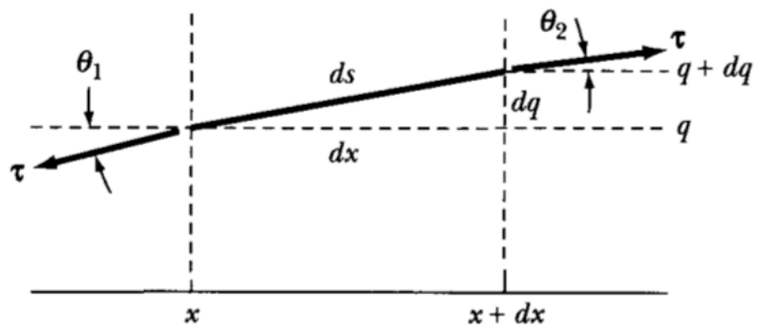


FIGURE 13-2 A portion, length *ds*, of a string fixed at both ends is shown. The displacement from equilibrium is q on the left and $q + dq$ on the right. The tensions τ at each end are equal in magnitude, but not in direction. Only transverse waves are considered.

We assume that the string has a constant mass density ρ (mass/length). We consider a length ds of the string described by s(x, t). The tensions τ on each end of the string are equal in magnitude but not in direction. This imbalance leads to a force and thus an acceleration of the system. We assume that the displacement q (perpendicular to x) is small. The mass dm of the length of string ds is ρ ds. The horizontal components of the tension are approximately equal and opposite, so we neglect the movement of the string in the x-direction. The force in the q-direction is

$$\Delta F = \rho \, ds \frac{\partial^2 q}{\partial t^2} \tag{13.34}$$

where ΔF represents the difference in tension at x and x + dx. We use partial derivatives to describe the acceleration, $\partial^2 q / \partial t^2$, because we are not considering the x-dependence of the displacement q(x, t).

The force can be found from the difference in the y-components of the tension.

$$
\begin{aligned}
(\Delta F)_y &= -\tau \sin\theta_1 + \tau \sin\theta_2 \\
&= -\tau \tan\theta_1 + \tau \tan\theta_2 \\
&= -\tau \left(\frac{\partial q}{\partial x}\right)\Bigg|_x + \tau \left(\frac{\partial q}{\partial x}\right)\Bigg|_{x+dx} \\
&= \tau \frac{\partial^2 q}{\partial x^2} dx
\end{aligned}
\tag{13.35}
$$

where we let $\sin\theta \approx \tan\theta$ because the angles θ are small for small displacements.

We now set Equations 13.34 and 13.35 equal, letting $ds \approx dx$:

$$
\tau \frac{\partial^2 q}{\partial x^2} dx = \rho dx \frac{\partial^2 q}{\partial t^2}
\tag{13.36}
$$

$$
\frac{\partial^2 q}{\partial x^2} = \frac{\rho}{\tau} \frac{\partial^2 q}{\partial t^2}
$$

Equation 13.36 is identical to Equation 13.33 but does not provide the useful information obtained earlier from the normal coordinate method.

13.5 Forced and Damped Motion

We can easily determine Lagrange's equations of motion for the vibrating string by using the kinetic energy from Equation 13.19 and the potential energy from Equation 13.24:

$$
\begin{aligned}
L &= T - U \\
&= \frac{\rho b}{4} \sum_r \dot{\eta}_r^2 - \frac{\rho b}{4} \sum_r \omega_r^2 \eta_r^2 \\
&= \frac{\rho b}{4} \sum_r (\dot{\eta}_r^2 - \omega_r^2 \eta_r^2)
\end{aligned}
\tag{13.37}
$$

where the length of the string has been set equal to b to avoid confusion between the Ls. The ease of the normal coordinate description is apparent. The equations of motion follow from Equation 13.37:

$$
\ddot{\eta}_r + \omega_r^2 \eta_r = 0 \tag{13.38}
$$

Next, we add a force per unit length F(x, t) acting along the string. We also add a damping force proportional to the velocity. The wave equation (Equation 13.33) now becomes

$$
\rho \frac{\partial^2 q}{\partial t^2} + D \frac{\partial q}{\partial t} - \tau \frac{\partial^2 q}{\partial x^2} = F(x, t) \tag{13.39}
$$

where each term represents a force per unit length, and D is the damping (resistive) term. Equation 13.39 is solved using normal coordinates. As we did in Section 13.2, we use a solution

$$
q(x, t) = \sum_r \eta_r(t) \sin\left(\frac{r\pi x}{b}\right) \tag{13.40}
$$

Substitution of Equation 13.40 into Equation 13.39 gives Lagrange's equations of motion - similar to Equation 13.38 but with the damping and forced terms added:

$$
\sum_{r=1}^{\infty} \left[\left(\rho \ddot{\eta}_r + D \dot{\eta}_r + \frac{r^2 \pi^2 \tau}{b^2} \eta_r \right) \sin\left(\frac{r\pi x}{b}\right) \right] = F(x, t) \tag{13.41}
$$

The sum over r is again from 1 to ∞ because we are considering a continuous string. The solution of Equation 13.41 parallels that of Section 13.2 (which we do not repeat here in detail) by comparing real and imaginary components. We multiply each side of Equation 13.41 by $\sin(s\pi x/b)$ and integrate over dx from 0 to b (remember that b = L = length of string). Using Equation 13.7, we have

$$\sum_{r=1}^{\infty}\left(\rho\ddot{\eta}_r + D\dot{\eta}_r + \frac{r^2\pi^2\tau}{b^2}\eta_r\right)\frac{b}{2}\delta_{rs} = \int_0^b F(x,t)\sin\left(\frac{s\pi x}{b}\right)dx \tag{13.42}$$

which becomes

$$\ddot{\eta}_s + \frac{D}{\rho}\dot{\eta}_s + \frac{s^2\pi^2\tau}{\rho b^2}\eta_s = \frac{2}{\rho b}\int_0^b F(x,t)\sin\left(\frac{s\pi x}{b}\right)dx \tag{13.43}$$

We now let $f_s(t)$ be the Fourier coefficient of the Fourier expansion of F(x, t), which is on the right side of Equation 13.43:

$$f_s(t) = \int_0^b F(x,t)\sin\left(\frac{s\pi x}{b}\right)dx \tag{13.44}$$

In normal coordinate terms, Equation 13.43 simply becomes

$$\ddot{\eta}_s + \frac{D}{\rho}\dot{\eta}_s + \frac{s^2\pi^2\tau}{\rho b^2}\eta_s = \frac{2}{\rho b}f_s(t) \tag{13.45}$$

It is now apparent that $f_s(t)$ is the component of F(x, t) effective in driving the normal coordinate s.

EXAMPLE 13.2

Reconsider Example 13.1. A sinusoidal driving force of angular frequency ω drives the string at x = b/2. Find the displacement.

Solution. The driving force per unit length is

$$\left.\begin{aligned} F(x,t) &= F_0 \cos \omega t, \quad && x = b/2 \\ &= 0. && x \neq b/2 \end{aligned}\right\} \tag{13.46}$$

The driving Fourier coefficient becomes

$$f_s(t) = F_0 \cos \omega t \sin \frac{s\pi}{2} \tag{13.47}$$

Notice that $f_s(t) = 0$ for even values of s. Only the odd terms are driven.

If we include a small damping term, Equation 13.45 becomes

$$\ddot{\eta}_s + \frac{D}{\rho}\dot{\eta}_s + \frac{s^2\pi^2\tau}{\rho b^2}\eta_s = \frac{2}{\rho b} F_0 \cos \omega t \sin \frac{s\pi}{2} \tag{13.48}$$

With the damping term effective, we need not determine a complementary solution, which will be damped out. We need only find a particular (steady-state) solution, as was done in Section 3.6. Equation 13.48 may be compared with Equation 3.53, where

$$\left.\begin{aligned} \frac{D}{\rho} &= 2\beta \\ \frac{s^2\pi^2\tau}{\rho b^2} &= \omega_0^2 \\ \frac{2F_0 \sin(s\pi/2)}{\rho b} &= A \end{aligned}\right\} \tag{13.49}$$

The solution (see Equation 3.60) for $\eta(t)$ becomes

$$\eta_s(t) = \frac{2F_0 \sin(s\pi/2) \cos(\omega t - \delta)}{\rho b \sqrt{\left(\frac{s^2\pi^2\tau}{\rho b^2} - \omega^2\right) + \frac{D^2}{\rho^2}\omega^2}} \tag{13.50}$$

where

$$\delta = \tan^{-1}\left[\frac{D\omega}{\rho\left(\frac{s^2\pi^2\tau}{\rho b^2} - \omega^2\right)}\right] \tag{13.51}$$

and the displacement of q(x, t) is

$$q(x, t) = \sum_r \frac{2F_0 \sin\frac{r\pi}{2} \cos(\omega t - \delta) \sin\left(\frac{r\pi x}{b}\right)}{\rho b \sqrt{\left(\frac{r^2\pi^2\tau}{\rho b^2} - \omega^2\right) + \frac{D^2}{\rho^2}\omega^2}} \tag{13.52}$$

where we have neglected the part of the solution that is damped out. Equation 13.52 represents many of the features discussed previously. Depending on the driving frequency, only a few of the normal coordinates may dominate because of the resonance effects inherent in the denominator. If the damping term is negligible, the dominant normal coordinate terms are

$$r^2 = \frac{\omega^2\rho b^2}{\pi^2\tau} \tag{13.53}$$

and because of the $\sin(r\pi/2)$ term of Equation 13.52, only odd values of r are effective.

13.6 General Solutions of the Wave Equation

The one-dimensional wave equation for a vibrating string (see Equation 13.33) is

$$\frac{\partial^2 \Psi}{\partial x^2} - \frac{\rho}{\tau}\frac{\partial^2 \Psi}{\partial t^2} = 0 \tag{13.54}$$

where ρ is the linear mass density of the string, τ is the tension, and Ψ is called the wave function. We use the notation $\Psi = \Psi(x,t)$ to denote a *time-dependent* wave function and $\psi = \psi(x)$ to denote a *time-independent* wave function.

The dimensions of ρ are $[ML^{-1}]$ and the dimensions of τ are those of a force, namely, $[MLT^{-2}]$. The dimensions of ρ/τ are therefore $[T^2L^{-2}]$ that is, the dimensions of the reciprocal of a squared velocity. If we write $\sqrt{\rho/\tau} = v$, the wave equation becomes

$$\boxed{\frac{\partial^2 \Psi}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 \Psi}{\partial t^2} = 0} \tag{13.55}$$

One of our tasks is to give a physical interpretation of the velocity v; it is not sufficient to say that v is the "velocity of propagation" of the wave.

To show that Equation 13.55 does indeed represent a general wave motion, we introduce two new variables,

$$\left.\begin{aligned} \xi &\equiv x + vt \\ \eta &\equiv x - vt \end{aligned}\right\} \tag{13.56}$$

Evaluating the derivatives of $\Psi = \Psi(x,t)$, which appear in Equation 13.55, we have

$$\frac{\partial \Psi}{\partial x} = \frac{\partial \Psi}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial \Psi}{\partial \eta}\frac{\partial \eta}{\partial x} = \frac{\partial \Psi}{\partial \xi} + \frac{\partial \Psi}{\partial \eta} \tag{13.57}$$

Then,

$$\frac{\partial^2 \Psi}{\partial x^2} = \frac{\partial}{\partial x}\frac{\partial \Psi}{\partial x} = \frac{\partial}{\partial x}\left(\frac{\partial \Psi}{\partial \xi} + \frac{\partial \Psi}{\partial \eta}\right)$$

$$= \frac{\partial}{\partial \xi}\left(\frac{\partial \Psi}{\partial \xi} + \frac{\partial \Psi}{\partial \eta}\right)\frac{\partial \xi}{\partial x} + \frac{\partial}{\partial \eta}\left(\frac{\partial \Psi}{\partial \xi} + \frac{\partial \Psi}{\partial \eta}\right)\frac{\partial \eta}{\partial x} \tag{13.58}$$

$$= \frac{\partial^2 \Psi}{\partial \xi^2} + 2\frac{\partial^2 \Psi}{\partial \xi \partial \eta} + \frac{\partial^2 \Psi}{\partial \eta^2}$$

Similarly, we find

$$\frac{1}{v}\frac{\partial \Psi}{\partial t} = \frac{\partial \Psi}{\partial \xi} - \frac{\partial \Psi}{\partial \eta} \tag{13.59}$$

and

$$\frac{1}{v^2}\frac{\partial^2 \Psi}{\partial t^2} = \frac{1}{v}\frac{\partial}{\partial t}\left(\frac{1}{v}\frac{\partial \Psi}{\partial t}\right) = \frac{1}{v}\frac{\partial}{\partial t}\left(\frac{\partial \Psi}{\partial \xi} - \frac{\partial \Psi}{\partial \eta}\right) \tag{13.60}$$

$$= \frac{\partial^2 \Psi}{\partial \xi^2} - 2\frac{\partial^2 \Psi}{\partial \xi \partial \eta} + \frac{\partial^2 \Psi}{\partial \eta^2}$$

But according to Equation 13.55, the right-hand sides of Equations 13.58 and 13.60 must be equal. This can be true only if

$$\frac{\partial^2 \Psi}{\partial \xi \, \partial \eta} \equiv 0 \tag{13.61}$$

The most general expression for Ψ that can satisfy this equation is a sum of two terms, one of which depends only on ξ and the other only on η ; no more complicated function of ξ and η permits Equation 13.61 to be valid. Thus,

$$\Psi = f(\xi) + g(\eta) \tag{13.62a}$$

or, substituting for ξ and η ,

$$\Psi = f(x + vt) + g(x - vt) \tag{13.62b}$$

where f and g are *arbitrary* functions of the variables x + vt and x - vt, respectively, which are not necessarily of a periodic nature, although they may be.

As time increases, the value of x must also increase in order to maintain a constant value for x - vt. The function g therefore retains its original form as time increases if we shift our viewpoint along the x-direction (in a positive sense) with a speed v. Thus, the function g must represent a disturbance that moves to the right (i.e., to larger values of x) with a speed v, whereas f represents the propagation of a disturbance to the left. We therefore conclude that Equation 13.55 does indeed describe wave motion and, in general, a traveling (or propagating) wave.

Let us now attempt to interpret Equation 13.62b in terms of the motion of a stretched string. At time t = 0, the displacement of the string is described by

$$q(x, 0) = f(x) + g(x)$$

If we take identical triangular forms for f(x) and g(x), the shape of the string at t = 0 is as shown at the top of Figure 13-3. As time increases, the disturbance represented by f(x + vt) propagates to the *left*, whereas the disturbance represented by g(x - vt) propagates to the *right*. This propagation of the individual disturbances to the left and right is illustrated in the lower part of Figure 13-3.

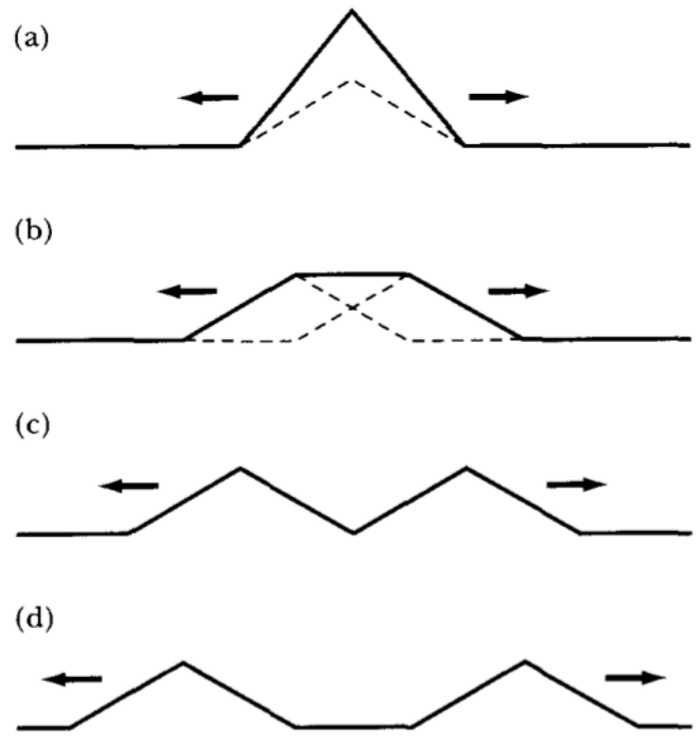


FIGURE 13-3 The propagation of a string is shown as a function of time from (a) to (d). At time $t = 0$ the string is described as $f(x) + g(x)$ as shown in (a). As time progresses, the disturbance $f(x + vt)$ propagates to the left and $g(x - vt)$ propagates to the right.

Consider next the left-going disturbance alone. If we terminate the string (at x = 0) by attaching it to a rigid support, we find the phenomenon of **reflection**. Because the support is rigid, we must have f(vt) = 0 for all values of time. This condition cannot be met by the function f alone (unless it trivially vanishes). We can satisfy the condition at x = 0 if we consider, in addition to f(x + vt), an imaginary disturbance, -f(-x + vt), which approaches the boundary point from the left, as in Figure 13-4. The disturbance f(x + vt) continues to propagate to the left, even into the imaginary section of the string (x < 0), while the disturbance -f(-x + vt) propagates across the boundary and along the real string. The net effect is that the original disturbance is reflected at the support and thereafter propagates to the right.

If the string is terminated by rigid supports at x = 0 and also at x = L, the disturbance propagates periodically back and forth with a period 2L/v.

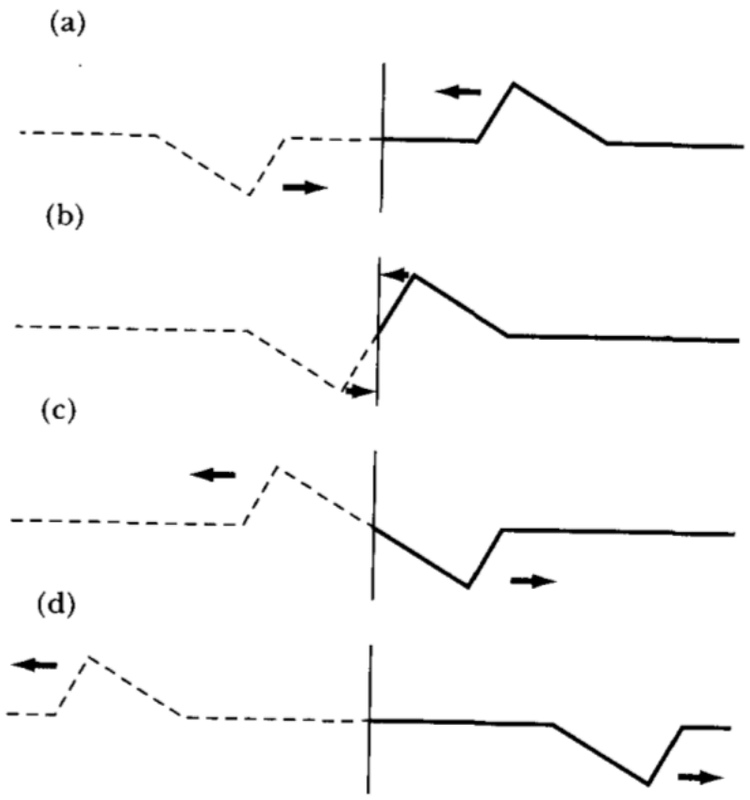


FIGURE 13-4 Consider only the left moving disturbance of Figure 13-3. The end of the string is fixed, and the wave reflects, because $f(vt) = 0$ always at the end. We can visualize the motion as if an imaginary disturbance (dashed line) was moving from the left to the right as time proceeds from (a) to (d).

13.7 Separation of the Wave Equation

If we require a general solution of the wave equation that is harmonic (as for the vibrating string or, for that matter, for a large number of problems of physical interest) , we can write

$$\Psi(x, t) = \psi(x)\, e^{i\omega t} \tag{13.63}$$

so that the one-dimensional wave equation (Equation 13.55) becomes

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\omega^2}{v^2}\psi = 0 \tag{13.64}$$

where ψ is now a function of x only.

The general wave motion of a system is not restricted to a single frequency (a. For a system with n degrees of freedom, there are n possible characteristic frequencies, and for a continuous string there is an infinite set of frequencies. An infinite set of frequencies would exist for a truly continuous string, but because a real string is composed fundamentally of atoms, there does exist an upper limit for ω . If we designate the r[th] frequency by ω_r , the wave function corresponding to this frequency is

$$\Psi_r(x, t) = \psi_r(x)\, e^{i\omega_r t} \tag{13.65}$$

The complete wave function is a superposition (recall that we are dealing with a linear system) of all the particular wave functions (or *modes*). Thus

$$\Psi(x, t) = \sum_r \Psi_r(x, t) = \sum_r \psi_r(x)\, e^{i\omega_r t} \tag{13.66}$$

In Equation 13.63, we assumed that the wave function was periodic in time. But now we see that this assumption entails no real restriction at all (apart from the usual assumptions regarding the continuity of the functions and the convergence of the series), because the summation in Equation 13.66 actually gives a Fourier representation of the wave function and is therefore the most general expression for the true wave function.

We now wish to show that Equation 13.65 results naturally from a powerful method that can often be used to obtain solutions to partial differential equations - the method of **separation of variables**. First, we express the solution as

$$\Psi(x, t) \equiv \psi(x) \cdot \chi(t) \tag{13.67}$$

that is, we assume that the variables are separable and therefore that the complete wave function can be expressed as the product of two functions, one of which is a spatial function only, and one of which is a temporal function only. It is not guaranteed that we will always find such functions, but many of the partial differential equations encountered in physical problems are separable in at least one coordinate system; some (such as those involving the Laplacian operator) are separable in many coordinate systems.

In short, the justification of the method of separation of variables, as is the case with many assumptions in physics, is in its success in producing mathematically acceptable solutions to a problem that eventually are found to properly describe the physical situation, i.e., are "experimentally verifiable."

Substituting $\Psi = \psi\chi$ into Equation 13.55, we have

$$\chi \frac{d^2\psi}{dx^2} - \frac{\psi}{v^2}\frac{d^2\chi}{dt^2} = 0$$

or

$$\frac{v^2}{\psi}\frac{d^2\psi}{dx^2} = \frac{1}{\chi}\frac{d^2\chi}{dt^2} \tag{13.68}$$

But, in view of the definitions of $\psi(x)$ and $\chi(t)$, the left-hand side of Equation 13.68 is a function of x alone, whereas the right-hand side is a function of t alone. This situation is possible only if each part of the equation is equal to the same constant. To be consistent with our previous notation, we choose this constant to be $-\omega^2$. Thus, we have

$$\frac{d^2\psi}{dx^2} + \frac{\omega^2}{v^2}\psi = 0 \tag{13.69a}$$

and

$$\frac{d^2\chi}{dt^2} + \omega^2\chi = 0 \tag{13.69b}$$

These equations are of a familiar form, and we know that the solutions are

$$\psi(x) = Ae^{i(\omega/v)x} + Be^{-i(\omega/v)x} \tag{13.70a}$$

$$\chi(t) = Ce^{i\omega t} + De^{-i\omega t} \tag{13.70b}$$

where the constants A, B, C, D are determined by the boundary conditions. We may write the solution $\Psi(x,t)$ in a shorthand manner as

$$\begin{aligned}\Psi(x,t) = \psi(x)\chi(t) &\sim \exp[\pm i(\omega/v)x]\exp[\pm i\omega t] \\ &\sim \exp[\pm i(\omega/v)(x \pm vt)]\end{aligned} \tag{13.71}$$

This notation means that the wave function Ψ varies as a linear combination of the terms

$$\exp[i(\omega/v)(x + vt)]$$
$$\exp[i(\omega/v)(x - vt)]$$
$$\exp[-i(\omega/v)(x + vt)]$$
$$\exp[-i(\omega/v)(x - vt)]$$

The *separation constant* for Equation 13.68 was chosen to be $-\omega^2$. There is nothing in the mathematics of the problem to indicate that there is a unique value of ω; hence, there must exist a set of equally acceptable frequencies ω_r. At this stage of the development, the set is in fact infinite, because no frequencies have yet been eliminated by boundary conditions. To each such frequency, there corresponds a wave function:

$$\Psi_r(x,t) \sim \exp[\pm i(\omega_r/v)(x \pm vt)]$$

The general solution is therefore not only a linear combination of the harmonic terms but also a sum over all possible frequencies;

$$\begin{aligned}\Psi(x,t) &\sim \sum_r a_r \Psi_r \\ &\sim \sum_r a_r \exp[\pm i(\omega_r/v)(x \pm vt)]\end{aligned} \tag{13.72}$$

The general solution of the wave equation therefore leads to a very complicated wave function. There are, in fact, an infinite number of arbitrary constants a_r. This is a general result for partial differential equations; but this infinity of constants must satisfy the physical requirements of the problem (the boundary conditions), and therefore they can be evaluated in the same manner that the coefficients of an infinite Fourier expansion can be evaluated.

For much of our discussion, it is sufficient to consider only one of the four possible combinations expressed by Equation 13.71; that is, we select a wave propagating in a particular direction and with a particular phase. Then, we can write, for example,

$$\Psi_r(x, t) \sim \exp[-i(\omega_r/v)(x - vt)]$$

This is the r^{th} Fourier component of the wave function, and the general solution is a summation over all such components. The functional form of each component is, however, the same, and so they can be discussed separately. Thus, we shall usually write, for simplicity,

$$\Psi(x, t) \sim \exp[-i(\omega/v)(x - vt)] \tag{13.73}$$

The general solution must be obtained by a summation over all frequencies that are allowed by the particular physical situation.

It is customary to write the differential equation for $\psi(x)$ as

$$\boxed{\frac{d^2\psi}{dx^2} + k^2\psi = 0} \tag{13.74}$$

which is the time-dependent form of the one-dimensional wave equation, also called the **Helmholtz equation**, and where

$$k^2 \equiv \frac{\omega^2}{v^2} \tag{13.75}$$

The quantity k, called the propagation constant or the wave number (i.e., proportional to the number of wavelengths per unit length), has dimensions $[L^{-1}]$. The wavelength λ is the distance required for one complete vibration of the wave,

$$\lambda = \frac{v}{\nu} = \frac{2\pi v}{\omega}$$

and thus the relation between k and λ is

$$k = \frac{2\pi}{\lambda}$$

We can therefore write, in general,

$$\Psi_r(x, t) \sim e^{\pm ik_r(x \pm vt)}$$

or, for the simplified wave function,

$$\Psi(x, t) \sim e^{-ik(x-vt)} = e^{i(\omega t - kx)} \tag{13.76}$$

If we superimpose two traveling waves of the type given by Equation 13.76 and if these waves are of equal magnitude (amplitude) but moving in opposite directions, then

$$\Psi = \Psi_+ + \Psi_- = Ae^{-ik(x+vt)} + Ae^{-ik(x-vt)} \tag{13.77}$$

or

$$\Psi = Ae^{-ikx}(e^{i\omega t} + e^{-i\omega t})$$

$$= 2Ae^{-ikx} \cos \omega t$$

the real part of which is

$$\Psi = 2A \cos kx \cos \omega t \tag{13.78}$$

Such a wave no longer has the property that it propagates; the wave form does not move forward with time. There are, in fact, certain positions at which there is no motion. These positions, the **nodes**, result from the complete cancellation of one wave by the other. The nodes of the wave function given by Equation 13.78 occur at $x = (2n + 1)\ \pi\ /2k$, where n is an integer Because there are fixed positions in waves of this type, they are called **standing waves**. Solutions to the problem of the vibrating string are of this form (but with a phase factor attached to the term kx such that the cosine is transformed into a sine function satisfying the boundary conditions)

EXAMPLE 13.3

Consider a string consisting of two densities, ρ_1 in region 1 where $x < 0$ and ρ_2 in region 2 where $x > 0$. A continuous wave train is incident from the left (i.e., from negative values of x). What are the ratios of the square of the amplitude magnitudes for the reflected and transmitted waves to the incident wave?

Solution. The wave will be both reflected and transmitted at $x = 0$ where the mass density discontinuity occurs. Therefore, in region 1 we have the superposition of the incident and reflected waves, and in region 2 we have only the transmitted wave. If the incident wave is $Ae^{i(\omega t - k_1 x)}$, then we have for the waves $\Psi_1(x,t)$ and $\Psi_2(x,t)$ in regions 1 and 2, respectively (see Equation 13.77)

$$\left.\begin{aligned} \Psi_1(x,\,t) &= \Psi_{\text{inc}} + \Psi_{\text{refl}} = Ae^{i(\omega t - k_1 x)} + Be^{i(\omega t + k_1 x)} \\ \Psi_2(x,\,t) &= \Psi_{\text{trans}} = Ce^{i(\omega t - k_2 x)} \end{aligned}\right\} \tag{13.79}$$

In Equation 13.79, we have explicitly taken into account the fact that the waves in both regions have the same frequency. But because the wave velocity on a string is given by

$$v = \sqrt{\frac{\tau}{\rho}}$$

we have $v_1 \neq v_2$, and therefore $k_1 \neq k_2$. We also have

$$k = \frac{\omega}{v} = \omega\sqrt{\frac{\rho}{\tau}} \tag{13.80}$$

so, in terms of the wave number of the incident wave,

$$k_2 = k_1\sqrt{\frac{\rho_2}{\rho_1}} \tag{13.81}$$

The amplitude A of the incident wave (see Equation 13.79) is given and is real. We must then obtain the amplitudes B and C of the reflected and transmitted waves to complete the solution of the problem. There are as yet no restrictions on B and C, and they may be complex quantities.

The physical requirements on the problem may be stated in terms of the boundary conditions. These are, simply, that the total wave function $\Psi = \Psi_1 + \Psi_2$ and its derivative must be continuous across the boundary. The continuity of Ψ results from the fact that the string is continuous. The condition on the derivative prevents the occurrence of a "kink" in the string, for if $\partial\Psi/\partial x_{0+} \neq \partial\Psi/\partial x_{0-}$ then $\partial^2\Psi/\partial x^2$ is infinite at x = 0; but the wave equation relates $\partial^2\Psi/\partial x^2$ and $\partial^2\Psi/\partial t^2$; and if the former is infinite, this implies an infinite acceleration, which is not allowed by the physical situation. We have, therefore, for all values of the time t,

$$\boxed{\Psi_1|_{x=0} = \Psi_2|_{x=0}} \tag{13.82a}$$

$$\boxed{\left.\frac{\partial\Psi_1}{\partial x}\right|_{x=0} = \left.\frac{\partial\Psi_2}{\partial x}\right|_{x=0}} \tag{13.82b}$$

From Equations 13.79 and 13.82a, we have

$$A + B = C \tag{13.83a}$$

and from Equations 13.79 and 13.82b we obtain

$$-k_1A + k_1B = -k_2C \tag{13.83b}$$

The solution of this pair of equations yields

$$B = \frac{k_1 - k_2}{k_1 + k_2} A \tag{13.84a}$$

and

$$C = \frac{2k_1}{k_1 + k_2} A \tag{13.84b}$$

The wave numbers k1 and k2 are both real, so amplitudes B and C are likewise real. Furthermore, k_1, k_2, and λ are all positive, so C is always positive. Thus, the transmitted wave is always in phase with the incident wave. Similarly, if $k_1 > k_2$, then the incident and reflected waves are in phase, but they are out of phase for $k_2 > k_1$, that is, for $\rho_2 > \rho_1$

The **reflection coefficient** R is defined as the ratio of the squared magnitudes of the amplitudes of the reflected and incident waves:

$$R \equiv \frac{|B|^2}{|A|^2} = \left(\frac{k_1 - k_2}{k_1 + k_2}\right)^2 \tag{13.85}$$

Because the energy content of a wave is proportional to the square of the amplitude of the wave function, R represents the ratio of the reflected energy to the incident energy. The quantity $|B|^2$ represents the intensity of the reflected wave.

No energy can be stored in the junction of the two strings, so the incident energy must be equal to the sum of the reflected and transmitted energies; that is, R+ T = 1.Thus,

$$T = 1 - R = \frac{4k_1k_2}{(k_1 + k_2)^2} \tag{13.86}$$

or

$$T = \frac{k_2}{k_1} \frac{|C|^2}{|A|^2} \tag{13.87}$$

In the study of the reflection and transmission of electromagnetic waves, we find quite similar expressions for R and T.

13.8 Phase Velocity, Dispersion, and Attenuation

We have seen in Equations 13.71 that the general solution to the wave equation produces, even in the one-dimensional case, a complicated system of exponential factors. For the purposes of further discussion, we restrict our attention to the particular combination

$$\Psi(x, t) = Ae^{i(\omega t - kx)} \tag{13.88}$$

This equation describes the propagation to the right (larger x) of a wave possessing a well-defined angular frequency ω . Certain physical situations can be quite adequately approximated by a wave function of this type - for example, the propagation of a monochromatic light wave in space or the propagation of a sinusoidal wave on a long (strictly, infinitely long) string.

If the argument of the exponential in Equation 13.88 remains constant, then the wave function $\Psi(x, t)$ also remains constant. The argument of the exponential is called the **phase** ϕ of the wave,

$$\phi \equiv \omega t - kx \tag{13.89}$$

If we move our viewpoint along the x-axis at a velocity such that the phase at every point is the same, we always see a stationary wave of the same shape. The velocity V with which we must move, called the **phase velocity** of the wave, corresponds to the velocity with which the **wave form** propagates. To ensure $\phi =$ constant, we set

$$d\phi = 0 \tag{13.90}$$

or

$$\omega\, dt = k\, dx$$

from which

$$V = \frac{dx}{dt} = \frac{\omega}{k} = v \tag{13.91}$$

so that the phase velocity in this case is just the quantity originally introduced as the velocity. It is possible to speak of a phase velocity only when the wave function has the same form throughout its length. This condition is necessary so we can measure the wavelength by taking the distance between *any* two successive wave crests (or between *any* two successive corresponding points on the wave). If the wave form were to change as a function of time or of distance along the wave, these measurements would not always yield the same results. The wavelength is not a function of time or space (i.e., that (ω is *pure*) only if the wave train is of infinite length. If the wave train is of finite length, there must be a spectrum of frequencies present in the wave, each with its own phase velocity. We will often assign a single frequency and phase velocity to a wave of finite length as a convenient approximation.

Let us return to the example of the loaded string and examine the properties of the phase velocity in that case. We have previously found (Equation12.152) that the frequency for the r^{th} mode of the loaded string when terminated at both ends is given by

$$\omega_r = 2\sqrt{\frac{\tau}{md}}\sin\left[\frac{r\pi}{2(n+1)}\right] \tag{13.92}$$

where the notation is the same as in Chapter 12. Recall that we take only positive values for the frequencies. When r = 1, there is a node at each end, and none between; hence, the length of the string is one-half of a wavelength. Similarly, when r = 2, then L = λ and, in general, λ_r = 2L/r. Therefore,

$$\frac{r\pi}{2(n+1)} = \frac{r\pi d}{2d(n+1)} = \frac{r\pi d}{2L} = \frac{\pi d}{\lambda_r} = \frac{k_r d}{2} \tag{13.93}$$

and

$$\omega_r = 2\sqrt{\frac{\tau}{md}}\sin\left(\frac{k_r d}{2}\right) \tag{13.94}$$

Because this expression no longer contains n or L, it applies equally well to a terminated or infinite loaded string.

To study the propagation of a wave in the loaded string, we initiate a disturbance by forcing one of the particles, say, the *zero*th one, to move according to

$$q_0(t) = Ae^{i\omega t} \tag{13.95}$$

If the string contains many particles(strictly, we need an infinite number of particles for this type of analysis, but we may approach the ideal conditions as closely as desired by increasing the finite number of particles), then any angular frequency less than is an allowed frequency (actually an eigenfrequency), satisfying Equation 13.95. After the transient effects have subsided and the steady-state conditions are attained, the phase velocity of the wave is given by

$$\boxed{V = \frac{\omega}{k} = \sqrt{\frac{\tau d}{m}} \frac{|\sin(kd/2)|}{kd/2} = V(k)} \tag{13.96}$$

In Equation 13.92 the values of r are required to be $\leq$ n (see Equation 12.144), so we automatically have $\omega_r \geq 0$ because $\sin[r\pi/2(n+1)] \geq 0$ for $0 \leq r \leq n$. We no longer have such a restriction on kd, so $\sin(kd/2)$ can become negative. We continue to consider only positive frequencies by always taking only the magnitude of $\sin(kd/2)$.

Thus the phase velocity is a function of the wave number; that is, Fis frequency-dependent. When V = V(k) for a given medium, that medium is said to be **dispersive**, and the wave exhibits **dispersion**. The best-known example of this phenomenon is the simple optical prism. The index of refraction of the prism depends on the wavelength of the incident light (i.e., the prism is a dispersive medium for optical light); on passing through the prism, the light is separated into a spectrum of wavelengths (i.e., the lightwave is dispersed).

For a longitudinal wave propagating down a long, slender rod, most of the energy is associated with the direction of the longitudinal wave propagation.

There is, however, a small amount of energy dissipated in a transverse wave moving at right angles. This lateral disturbance causes the phase velocity of the longitudinal wave to be decreased, and the effect depends on wavelength. For large wavelengths, the effect is small; for short wavelengths, especially those approaching the radius of the rod, the velocity dispersion is pronounced.

From Equation 13.96, we see that, as the wavelength becomes very long ($\lambda \to \infty$ or $k \to 0$), the phase velocity approaches the constant value

$$V(\lambda \to \infty) = \sqrt{\frac{\tau d}{m}} \tag{13.97}$$

Otherwise, V= V(k), and the wave is dispersive. We note that the phase velocity for the continuous string (see Equation 13.55) is

$$V_{\text{cont.}} = v = \sqrt{\frac{\tau}{\rho}} \tag{13.98}$$

and because m/d for the loaded string corresponds to ρ for the continuous string, the phase velocities for the two cases are equal in the long-wavelength limit (but *only* in this limit). This is a reasonable result because as λ becomes large compared with d, the properties of the wave are less sensitive to the spacing between particles, and in the limit, d may vanish without affecting the phase velocity.

In Equation 13.94, the restriction on r is $1 \le r \le n$. Then, because $k_r = r\pi/L$, we see that the value of k_r that maximizes ω_r in Equation 13.94 is

$$k_{\text{max}} = \pi/d \tag{13.99}$$

The corresponding frequency, from Equation 13.96, is $2\sqrt{\tau/md}$. What is the result of forcing the string to vibrate at a frequency *greater* than $2\sqrt{\tau/md}$? For this purpose, we allow k to become complex and investigate the consequences:

$$k \equiv \kappa - i\beta, \qquad \kappa, \beta > 0 \tag{13.100}$$

The expression for ω (Equation 13.94) then becomes

$$\begin{aligned}\omega &= 2\sqrt{\frac{\tau}{md}}\sin\left[\frac{d}{2}(\kappa - i\beta)\right] \\ &= 2\sqrt{\frac{\tau}{md}}\left(\sin\frac{d\kappa}{2}\cos\frac{i\beta d}{2} - \cos\frac{\kappa d}{2}\sin\frac{i\beta d}{2}\right) \\ &= 2\sqrt{\frac{\tau}{md}}\left(\sin\frac{\kappa d}{2}\cosh\frac{\beta d}{2} - i\cos\frac{\kappa d}{2}\sinh\frac{\beta d}{2}\right)\end{aligned} \tag{13.101}$$

If the frequency is to be a real quantity, the imaginary part of this expression must vanish. Thus, we may have either cos(kd/2) = 0 or sinh(β d/2) = 0. But the latter choice requires $\beta = 0$, contrary to the requirement that k be complex. We therefore have

$$\cos\frac{\kappa d}{2} = 0 \tag{13.102}$$

For this case, we must also have

$$\sin\frac{\kappa d}{2} = 1 \tag{13.103}$$

The expression for the angular frequency becomes

$$\omega = 2\sqrt{\frac{\tau}{md}}\cosh\frac{\beta d}{2} \tag{13.104}$$

Thus, we have the result that, for $\omega \leq 2\sqrt{\tau/md}$, the wave number k is real and the relation between k and ω is given by Equation 13.94; whereas, for $\omega > 2\sqrt{\tau/md}$, k is complex with the real part κ fixed by Equation 13.102 at the value $\kappa = \pi/d$ and with the imaginary part β given by Equation 13.104. The situation is shown in Figure 13-5.

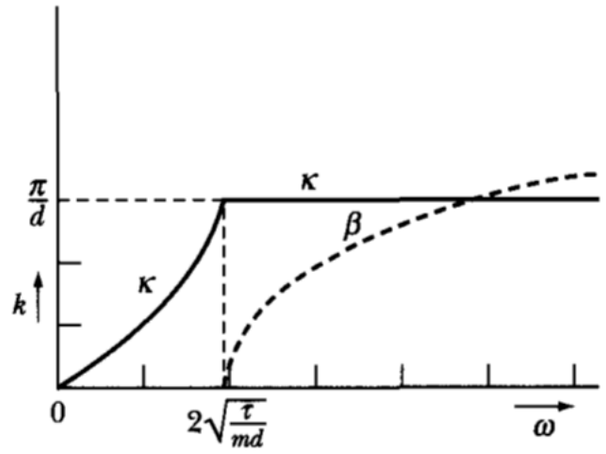


FIGURE 13-5 In the case of the loaded string the wave number k is real for angular frequencies $\omega \leq 2\sqrt{\tau/md}$. For $\omega > 2\sqrt{\tau/md}$, the wave number k is complex $(\kappa - i\beta)$ with the real part denoted by κ (displayed as fixed at π/d) and the imaginary part denoted by β (dashed line).

What is the physical significance of a complex wave number? Our original wave function was of the form

$$\Psi = Ae^{i(\omega t - kx)}$$

but, if k= $\kappa - i\beta$, then Ψ can be written as

$$\Psi = Ae^{-\beta x}e^{i(\omega t - \kappa x)} \tag{13.105}$$

and the factor $\exp(-\beta x)$ represents a damping, or *attenuation*, of the wave with increasing distance x. We therefore conclude that the wave is propagated without attenuation for $\omega \leq 2\sqrt{\tau/md}$ (this region is called the *passing band* of frequencies), and that attenuation sets in at $\omega_c = 2\sqrt{\tau/md}$ (called the *critical* or *cutoff frequency*) and increases with increasing frequency.

The physical significance of the real and imaginary parts of k is now apparent: β is the attenuation coefficient (and exists only if $\omega > \omega_c$), whereas κ is the wave number in the sense that the phase velocity V′ is given by

$$V' = \frac{\omega}{\kappa} = \frac{\omega}{\mathrm{Re}\, k} \tag{13.106}$$

rather than by V= w/k. If k is real, these expressions for V and V′are identical.

This example emphasizes the fact that the fundamental definition of the phase velocity is based on the requirement of the constancy of the phase and not on the ratio ω/k. Thus, in general, the phase velocity V and the so-called wave velocity v are distinct quantities. We note also that if ω is real and if the wave number k is complex, then the *wave velocity* v must also be complex so that the product kv yields a real quantity for the frequency through the relation $\omega = kv$. On the other hand, the *phase velocity*, which arises from the requirement that ϕ = constant, is necessarily always a real quantity.

In the preceding discussion, we considered the system to be conservative and argued that this requires ω to be a real quantity. We found that if ω exceeds the critical frequency ω_c attenuation results and the wave number becomes complex. If we relax the condition that the system is conservative, the frequency may then be complex and the wave number real. In such a case, the wave is damped in time rather than in *space*. Spatial attenuation (ω real, k complex) is of particular significance for traveling waves, whereas temporal attenuation (ω complex, k real) is important for standing waves.

Although attenuation occurs in the loaded string if $\omega > \omega_c$, the system is still conservative and no energy is lost. This seemingly anomalous situation results because the force applied to the particle in the attempt to initiate a traveling wave is (after the steady-state condition of an attenuated wave is set up) exactly 90° out of phase with the velocity of the particle, so that the power transferred, $P = \mathbf{F} \cdot \mathbf{v}$, is zero.

In this treatment of the loaded string, we have tacitly assumed an ideal situation; that is, the system was assumed to be lossless. As a result, we found that there was attenuation for $\omega > \omega_c$ but none for $\omega < \omega_c$. However, every real system is subject to loss, so in fact there is some attenuation even for $\omega < \omega_c$.

13.9 Group Velocity and Wave Packets

It was demonstrated in Section 3.9 that the superposition of various solutions of a linear differential equation is still a solution to the equation. Indeed, we formed the general solution to the problem of small oscillations (see Equation 12.43) by summing all the particular solutions. Let us assume, therefore, that we have two almost equal solutions to the wave equation represented by the wave functions Ψ_1 and Ψ_2 each of which has the same amplitude.

$$\left.\begin{aligned}\Psi_1(x,t) &= Ae^{i(\omega t - kx)}\\ \Psi_2(x,t) &= Ae^{i(\Omega t - Kx)}\end{aligned}\right\} \tag{13.107}$$

but whose frequencies and wave numbers differ by only small amounts

$$\left.\begin{aligned}\Omega &= \omega + \Delta\omega\\ K &= k + \Delta k\end{aligned}\right\} \tag{13.108}$$

Forming the solution that consists of the sum Ψ_1 and Ψ_2 we have

$$\begin{aligned}\Psi(x,t) = \Psi_1 + \Psi_2 &= A\big[\exp(i\omega t)\exp(-ikx)\\ &\quad + \exp\{i(\omega + \Delta\omega)t\}\exp\{-i(k + \Delta k)x\}\big]\\ &= A\left[\exp\left\{i\left(\omega + \frac{\Delta\omega}{2}\right)t\right\}\exp\left\{-i\left(k + \frac{\Delta k}{2}\right)x\right\}\right]\\ &\quad \cdot\left[\exp\left\{-i\left(\frac{(\Delta\omega)t - (\Delta k)x}{2}\right)\right\} + \exp\left\{i\left(\frac{(\Delta\omega)t - (\Delta k)x}{2}\right)\right\}\right]\end{aligned}$$

The second bracket is just twice the cosine of the argument of the exponential, and the real part of the first bracket is also a cosine. Thus, the real part of the wave function is

$$\Psi(x,t) = 2A\cos\left[\frac{(\Delta\omega)t - (\Delta k)x}{2}\right]\cos\left[\left(\omega + \frac{\Delta\omega}{2}\right)t - \left(k + \frac{\Delta k}{2}\right)x\right] \tag{13.109}$$

This expression is similar to that obtained in the problem of the weakly coupled oscillators (see Section 12.3), in which we found a slowly varying amplitude, corresponding to the term

$$2A\cos\left[\frac{(\Delta\omega)t - (\Delta k)x}{2}\right]$$

which modulates the wave function. The primary oscillation takes place at a frequency $\omega + (\Delta\omega/2)$, which, according to our assumption that $\Delta\omega$ is small, differs negligibly from ω. The varying amplitude gives rise to beats (Figure 13-6).

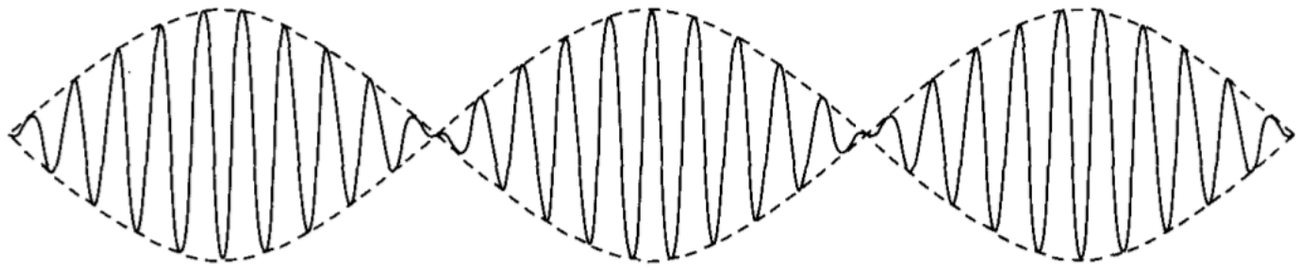


FIGURE 13-6 When two wave functions having frequencies very close together are summed, the phenomenon of beats (slowly varying amplitude) is observed.

The velocity U (called the **group velocity**) with which the modulations (or groups of waves) propagate is given by the requirement that the phase of the amplitude term be constant. Thus,

$$U = \frac{dx}{dt} = \frac{\Delta\omega}{\Delta k} \tag{13.110}$$

In a nondispersive medium $\Delta\omega/\Delta k = V$, so the group and phase velocities are identical. If dispersion is present, however, U and V are distinct.

Thus far, we have considered only the superposition of two waves. If we wish to superpose a system of n waves, we must write

$$\Psi(x,t) = \sum_{r=1}^{n} A_r \exp[i(\omega_r t - k_r x)] \tag{13.111a}$$

where A_r represents the amplitudes of the individual waves. In the event that n becomes very large (strictly, infinite), the frequencies are continuously distributed, and we may replace the summation by an integration, obtaining

$$\Psi(x,t) = \int_{-\infty}^{+\infty} A(k)\, e^{i(\omega t - kx)}\, dk \tag{13.111b}$$

where the factor A(k) represents the distribution amplitudes of the component waves with different frequencies, that is, the **spectral distribution** of the waves.

We have previously made the tacit assumption that $k \geq 0$. However, k is defined by $k^2 = \omega^2 / v^2$ (see Equation 13.75), so there is no mathematical reason why we may not also have $k < 0$. We may therefore extend the region of integration to include $-\infty < k < 0$ without mathematical difficulty. This procedure allows the identification of the integral representation of $\Psi(x,t)$ as a Fourier integral.

The most interesting cases occur when A(k) has a significant value only in the neighborhood of a particular wave number (say, k_0) and becomes vanishingly small for k outside a small range, denoted by $k_0 \pm \Delta k$. In such a case, the wave function can be written a

$$\Psi(x,t) = \int_{k_0-\Delta k}^{k_0+\Delta k} A(k)\, e^{i(\omega t - kx)}\, dk \tag{13.112}$$

A function of this type is called a **wave packet**. The concept of group velocity can be applied only to those cases that can be represented by a wave packet, that is, to wave functions containing a small range (or *band*) of frequencies.

For the case of the wave packet represented by Equation 13.112, the contributing frequencies are restricted to those lying near $\omega(k_0)$. We can therefore expand $\omega(k)$ about k = k_0.

$$\omega(k) = \omega(k_0) + \left(\frac{d\omega}{dk}\right)_{k=k_0} \cdot (k - k_0) + \cdots \tag{13.113a}$$

which we can abbreviate as

$$\omega = \omega_0 + \omega_0'(k - k_0) + \cdots \tag{13.113b}$$

The argument of the exponential in the wave packet integral becomes, approximately,

$$\omega t - kx = (\omega_0 t - k_0 x) + \omega_0'(k - k_0)t - (k - k_0)x$$

where we have added and subtracted the term k_0x. Thus,

$$\omega t - kx = (\omega_0 t - k_0 x) + (k - k_0)(\omega_0' t - x) \tag{13.114}$$

and Equation 13.112 becomes

$$\Psi(x, t) = \int_{k_0 - \Delta k}^{k_0 + \Delta k} A(k)\exp[i(k - k_0)(\omega_0' t - x)]\exp[i(\omega_0 t - k_0 x)]\,dk \tag{13.115}$$

The wave packet, expressed in this fashion, may be interpreted as follows.

The quantity

$$A(k)\exp[i(k - k_0)(\omega_0' t - x)]$$

constitutes an effective amplitude that, because of the small quantity (k - k_0) in the exponential, varies slowly with time and describes the motion of the wave packet (or envelope of a group of waves) in the same manner that the term

$$2A\cos\left[\frac{(\Delta\omega)t - (\Delta k)x}{2}\right]$$

describes the propagation of the packet formed from two superposed waves.

The requirement of constant phase for the amplitude term leads to

$$U = \omega_0' = \left(\frac{d\omega}{dk}\right)_{k=k_0} \tag{13.116}$$

for the group velocity. As stated earlier, only if the medium is dispersive does U differ from the phase velocity V. To show this explicitly, we write Equation 13.116 as

$$\frac{1}{U} = \left(\frac{dk}{d\omega}\right)_0$$

where the subscript zero means "evaluated at k= k_0, or, equivalently, at $\omega = \omega_0$ "

Because k = ω /v,

$$\frac{1}{U} = \left[\frac{d}{d\omega}\left(\frac{\omega}{v}\right)\right]_0 = \frac{v_0 - (\omega dv/d\omega)_0}{v_0^2}$$

Thus,

$$U = \frac{v_0}{1 - \dfrac{\omega_0}{v_0}\left(\dfrac{dv}{d\omega}\right)_0} \tag{13.117}$$

If the medium is nondispersive, v = V= constant, so $dv/d\omega = 0$ (see Equation 13.91); hence U= v_0= V.

The remaining quantity in Equation 13.115, $\exp[i(\omega_0 t - k_0 x)]$ varies rapidly with time; and if this were the only factor in Ψ , it would describe an infinite wave train oscillating at frequency ω_0 traveling with phase velocity V = ω_0 /k_0.

We should note that an infinite train of waves of a given frequency cannot transmit a signal or carry information from one point to another. Such transmission can be accomplished only by starting and stopping the wave train and thereby impressing a signal on the wave - in other words, by forming a wave packet. As a consequence of this fact, it is the group velocity, not the phase velocity, that corresponds to the velocity at which a signal may be transmitted.

CHAPTER 14 - Special Theory of Relativity

14.1 Introduction

In Section 2.7, it was pointed out that the Newtonian idea of the complete separability of space and time and the concept of the absoluteness of time break down when they are subjected to critical analysis. The final overthrow of the Newtonian system as the ultimate description of dynamics was the result of several crucial experiments, culminating with the work of Michelson and Morley in 1881-1887. The results of these experiments indicated that the speed of light is independent of any relative uniform motion between source and observer. This fact, coupled with the finite speed of light, required a fundamental reorganization of the structure of dynamics. This was provided during the period 1904-1905 by H. Poincare, H. A. Lorentz, and A. Einstein, who formulated the theory of relativity in order to provide a consistent description of the experimental facts. The basis of relativity theory is contained in two postulates:

I. *The laws of physical phenomena are the same in all inertial reference frames (that is, only the relative motion of inertial frames can be measured; the concept of motion relative to "absolute rest" is meaningless).*

II. *The velocity of light (in free space) is a universal constant, independent of any relative motion of the source and the observer.*

Using these postulates as a foundation, Einstein was able to construct a beautiful, logically precise theory. A wide variety of phenomena that take place at high velocity and cannot be interpreted in the Newtonian scheme are accurately described by relativity theory.

Postulate I, which Einstein called the *principle of relativity*, is the fundamental basis for the theory of relativity. Postulate II, the law of propagation of light, follows from Postulate I if we accept, as Einstein did, that Maxwell's equations are fundamental laws of physics. Maxwell's equations predict the speed of light in vacuum to be c, and Einstein believed this to be the case in all inertial reference frames.

We do not attempt here to give the experimental background for the theory of relativity; such information can be found in essentially every textbook on modem physics and in many others concerned with electrodynamics. Rather, we simply accept as correct the above two postulates and work out some of their consequences for the area of mechanics. The discussion here is limited to the case of **special relativity**, in which we consider only inertial reference frames, that is, frames that are in uniform motion with respect to one another. The more general treatment of accelerated reference frames is the subject of the **general theory of relativity**.

14.2 Galilean Invariance

In Newtonian mechanics, the concepts of space and time are completely separable; furthermore, time is assumed to be an absolute quantity susceptible of precise definition independent of the reference frame. These assumptions lead to the invariance of the laws of mechanics under coordinate transformations of the following type. Consider two inertial reference frames K and K', which move along their x_1 and x'_1-axes with a uniform relative velocity v (Figure 14-1).

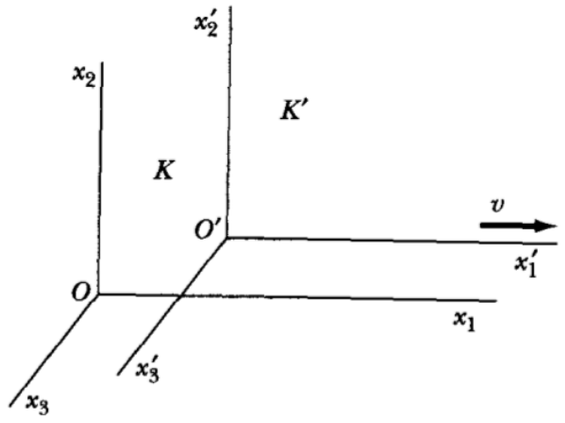


FIGURE 14-1 Two inertial reference frames K and K' move along their x_1- and x'_1-axes with a uniform relative velocity v.

The transformation of the coordinates of a point from one system to the other is clearly of the form

$$\left.\begin{aligned} x'_1 &= x_1 - vt \\ x'_2 &= x_2 \\ x'_3 &= x_3 \end{aligned}\right\} \tag{14.1a}$$

Also, we have

$$t' = t \tag{14.1b}$$

Equations 14.1 define a **Galilean transformation**. Furthermore, the element of length in the two systems is the same and is given by

$$\begin{aligned} ds^2 &= \sum_j dx_j^2 \\ &= \sum_j dx_j'^2 = ds'^2 \end{aligned} \tag{14.2}$$

The fact that Newton's laws are invariant with respect to Galilean transformations is termed **the principle of Newtonian relativity** or **Galilean invariance**. Newton's equations of motion in the two systems are

$$\begin{aligned} F_j &= m\ddot{x}_j \\ &= m\ddot{x}_j' = F'_j \end{aligned} \tag{14.3}$$

The form of the law of motion is then *invariant* to a Galilean transformation. The individual terms are not invariant, however, but they transform according to the same scheme and are said to be *covariant*.

We can easily show that the Galilean transformation is inconsistent with Postulate II. Consider a light pulse emanating from a flashbulb positioned in frame K'. The velocity transformation is found from Equation 14.1a, where we consider the light pulse only along x_1:

$$\dot{x}_1' = \dot{x}_1 - v \tag{14.4}$$

In system K', the velocity is measured $\dot{x}_1' = c$; Equation 14.4 therefore indicates the speed of the light pulse to be $\dot{x}_1 = c + v$, clearly in violation of Postulate II.

14.3 Lorentz Transformation

The principle of Galilean invariance predicts that the velocity of light is different in two inertial reference frames that are in relative motion. This result is in contradiction to the second postulate of relativity. Therefore, a new transformation law that renders physical laws *relativistically* covariant must be found. Such a transformation law is the **Lorentz transformation**. The original use of the Lorentz transformation preceded the development of Einsteinian relativity theory, but it also follows from the basic postulates of relativity; we derive it on this basis in the following discussion.

If a light pulse from a flashbulb is emitted from the common origin of the systems K and K' (see Figure 14-1) when they are coincident, then according to Postulate II, the wavefronts observed in the two systems must be described"' by

$$\left.\begin{aligned} \sum_{j=1}^{3} x_j^2 - c^2t^2 = 0 \\ \sum_{j=1}^{3} x_j'^2 - c^2t'^2 = 0 \end{aligned}\right\} \tag{14.5}$$

We can already see that Equations 14.5, which are consistent with the two postulates of the theory of relativity, cannot be reconciled with the Galilean transformations of Equations 14.1. The Galilean transformation allows a spherical light wavefront in one system but requires the center of the spherical wavefront in the second system to move at velocity v with respect to the first system. The interpretation of Equations 14.5, according to Postulate II, is that each observer believes that his spherical wavefront has its center fixed at his own coordinate origin as the wavefront expands.

We are faced with a quandary. We must abandon either the two relativity postulates or the Galilean transformation. Much experimental evidence, including the Michelson-Morley experiment and the aberration of starlight, requires the two postulates. However, the belief in the Galilean transformation is entrenched in our minds by our everyday experience.

The Galilean transformation had produced satisfactory results, including those of the preceding chapters of this book, for centuries. Einstein's great contribution was to realize that the Galilean transformation was *approximately* correct, but that we needed to reexamine our concepts of space and time.

Notice that we do not assume t = t' in Equations 14.5. Each system, K and K', has its own clocks, and we assume that a clock may be located at any point in space. These clocks are all identical, run the same way, and are synchronized. Because the flashbulb goes off when the origins are coincident and the systems move only in the x_1-direction with respect to each other, by direct observation we have

$$\left.\begin{aligned} x_2' &= x_2 \\ x_3' &= x_3 \end{aligned}\right\} \tag{14.6}$$

At time t = t' = 0, when the flashbulb goes off, the motion of the origin O' of K' is measured in K to be

$$x_1 - vt = 0 \tag{14.7}$$

and in system K', the motion of O' is

$$x_1' = 0 \tag{14.8}$$

At time t = t' = 0 we have $x'_1 = x_1 - vt$, but we know that Equation 14.1a is incorrect. Let us assume the next simplest transformation, namely,

$$x_1' = \gamma(x_1 - vt) \tag{14.9}$$

where γ is some constant that may depend on v and some constants, but not on the coordinates x_1, x'_1, t, or t'.

Equation 14.9 is a linear equation and assures us that each event in K corresponds to one and only one event in K'. This additional assumption in our derivation will be vindicated if we can produce a transformation that is consistent with all the experimental results.

Notice that γ must normally be very close to 1 to be consistent with the classical results discussed in earlier chapters.

We can use the preceding arguments to describe the motion of the origin O of system K in both K and K' to also determine

$$x_1 = \gamma'(x_1' + vt') \tag{14.10}$$

where we only have to change the relative velocities of the two systems.

Postulate I demands that the laws of physics be the same in both reference systems such that $\gamma = \gamma'$. By substituting x'_1 from Equation 14.9 into Equation 14.10, we can solve the remaining equation for t':

$$t' = \gamma t + \frac{x_1}{\gamma v}(1 - \gamma^2) \tag{14.11}$$

Postulate II demands that the speed of light be measured to be the same in both systems. Therefore, in both systems we have similar equations for the position of the flashbulb light pulse:

$$\left.\begin{aligned} x_1 &= ct \\ x_1' &= ct' \end{aligned}\right\} \tag{14.12}$$

Algebraic manipulation of Equations 14.9-14.12 gives

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \tag{14.13}$$

The complete transformation equations can now be written as

$$\left.\begin{aligned} x_1' &= \frac{x_1 - vt}{\sqrt{1 - v^2/c^2}} \\ x_2' &= x_2 \\ x_3' &= x_3 \\ t' &= \frac{t - \dfrac{vx_1}{c^2}}{\sqrt{1 - v^2/c^2}} \end{aligned}\right\} \tag{14.14}$$

These equations are known as the Lorentz (or Lorentz-Einstein) transformation in honor of the Dutch physicist H. A. Lorentz, who first showed that the equations are necessary so that the laws of electromagnetism have the same form in all inertial reference frames. Einstein showed that these equations are required for all the laws of physics.

The inverse transformation can easily be obtained by replacing v by -v and exchanging primed and unprimed quantities in Equations 14.14.

$$\left.\begin{aligned} x_1 &= \frac{x_1' + vt'}{\sqrt{1 - v^2/c^2}} \\ x_2 &= x_2' \\ x_3 &= x_3' \\ t &= \frac{t' + \dfrac{vx_1'}{c^2}}{\sqrt{1 - v^2/c^2}} \end{aligned}\right\} \tag{14.15}$$

As required, these equations reduce to the Galilean equations (Equations 14.1) when $v \to 0$ (or when $c \to \infty$).

In electrodynamics, the fields propagate with the speed of light, so Galilean transformations are never allowed. Indeed, the fact that the electrodynamic field equations (**Maxwell's equations**) are not covariant to Galilean transformations was a main factor in the realization of the need for a new theory. It seems rather extraordinary that Maxwell's equations, which are a complete set of equations for the electromagnetic field and are *covariant to Lorentz transformations*, were deduced from experiment long before the advent of relativity theory.

The velocities measured in each of the systems are denoted by u

$$u_i = \frac{dx_i}{dt}$$
$$u_i' = \frac{dx_i'}{dt'}$$ (14.16)

Using Equations 14.14, we determine

$$u_1' = \frac{dx_1'}{dt'} = \frac{dx_1 - v\,dt}{dt - \frac{v}{c^2}dx_1}$$

$$u_1' = \frac{u_1 - v}{1 - \frac{u_1 v}{c^2}} \tag{14.17a}$$

Similarly, we determine

$$u_2' = \frac{u_2}{\gamma\left(1 - \frac{u_1 v}{c^2}\right)} \tag{14.17b}$$

$$u_3' = \frac{u_3}{\gamma\left(1 - \frac{u_1 v}{c^2}\right)} \tag{14.17c}$$

Now we can determine whether Postulate II is satisfied directly. An observer in system K measures the speed of the light pulse from the flashbulb to be $u_1 = c$ in the x_1-direction. From Equation 14.17a, an observer in K' measures

$$u_1' = \frac{c - v}{1 - \frac{v}{c}} = c\left(\frac{c - v}{c - v}\right) = c$$

as required by Postulate II, independent of the relative system speed v.

EXAMPLE 14.1

Determine the relativistic length contraction using the Lorentz transformation.

Solution. Consider a rod of length lying along the x_1-axis of an inertial frame K. An observer in system K' moving with uniform speed v along the x_1-axis (as in Figure 14-1) measures the length of the rod in the observer's own coordinate system by determining at a given instant of time t' the difference in the coordinates of the ends of the rod, $x'_1(2) - x'_1(1)$. According to the transformation equations (Equations 14.14),

$$x'_1(2) - x'_1(1) = \frac{[x_1(2) - x_1(1)] - v[t(2) - t(1)]}{\sqrt{1 - v^2/c^2}} \tag{14.18}$$

where x1(2) — x1(1) = l . Note that times t(2) and t(l) are the times in the K system at which the observations are made; they do not correspond to the instants in K' at which the observer measures the rod. In fact, because t'(2) = t'(l), Equations 14.14 give

$$t(2) - t(1) = [x_1(2) - x_1(1)]\frac{v}{c^2}$$

The length l' as measured in the K' system is therefore

$$l' = x'_1(2) - x'_1(1)$$

Equation 14.18 now becomes

length contraction $\boxed{l' = l\sqrt{1 - v^2/c^2}}$ (14.19)

and, to a stationary observer in K, objects in K' also appear contracted. Thus, to an observer in motion relative to an object, the dimensions of objects are contracted by a factor $\sqrt{1 - \beta^2}$ in the direction of motion, in which β = v/c.

An interesting consequence of the FitzGerald-Lorentz contraction of length was reported in 1959 by James Terrell. Consider a cube of side A moving with uniform velocity v with respect to an observer some distance away. Figure 14-2a shows the projection of the cube on the plane containing the velocity vector **v** and the observer.

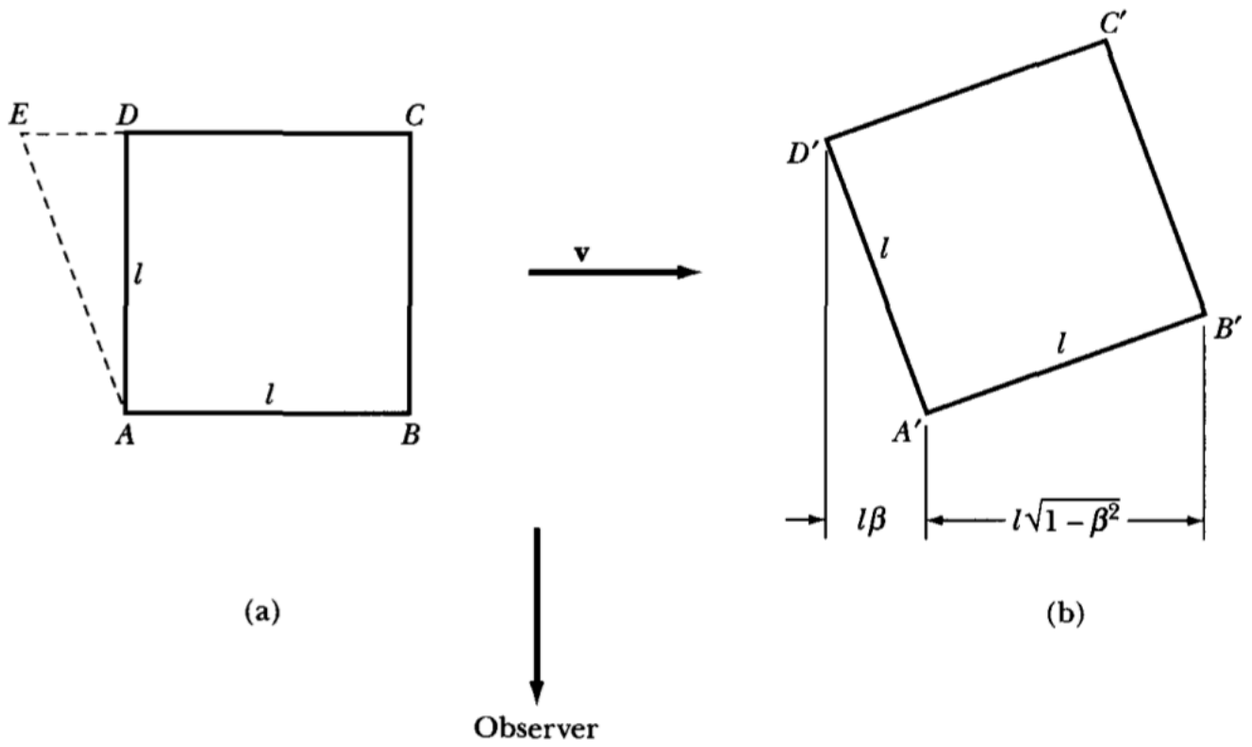


FIGURE 14-2 (a) An observer far away sees a cube of sides l at rest in system K. (b) Terrell pointed out that surprisingly, the same cube appears to be rotated if it is moving to the right with velocity **v** relative to system K.

The cube moves with its side AB perpendicular to the observer's line of sight. We wish to determine what the observer "sees"; that is, at a given instant of time in the observer's rest frame, we wish to determine the relative orientation of the corners A, B, C, and D. The traditional view (which went unquestioned for more than 50 years!) was that the only effect is a foreshortening of the sides AB and CD such that the observer sees a distorted tube of height l but of length $l\sqrt{1-\beta^2}$. Terrell pointed out that this interpretation overlooks certain facts: For light from corners A and D to reach the observer at the same instant, the light from D, which must travel a distance l farther than that from A, must have been emitted when comer D was at position E. The length DE is equal to $(l/c)v = l\beta$.

Therefore, the observer sees not only face AB, which is perpendicular to the line of sight, but also face AD, which is *parallel* to the line of sight. Also, the length of the side AB is foreshortened in the normal way to $l\sqrt{1-\beta^2}$. The net result (Figure 14-2b) corresponds exactly to the view the observer would have if the cube were rotated through an angle $\sin^{-1}\beta$. Therefore, the cube is not distorted; it undergoes an apparent rotation. Similarly, the customary statement that a moving sphere appears as an ellipsoid is incorrect; it appears still as a sphere. Computers can be used to show extremely interesting results of the type we have been discussing (Figure 14-3).

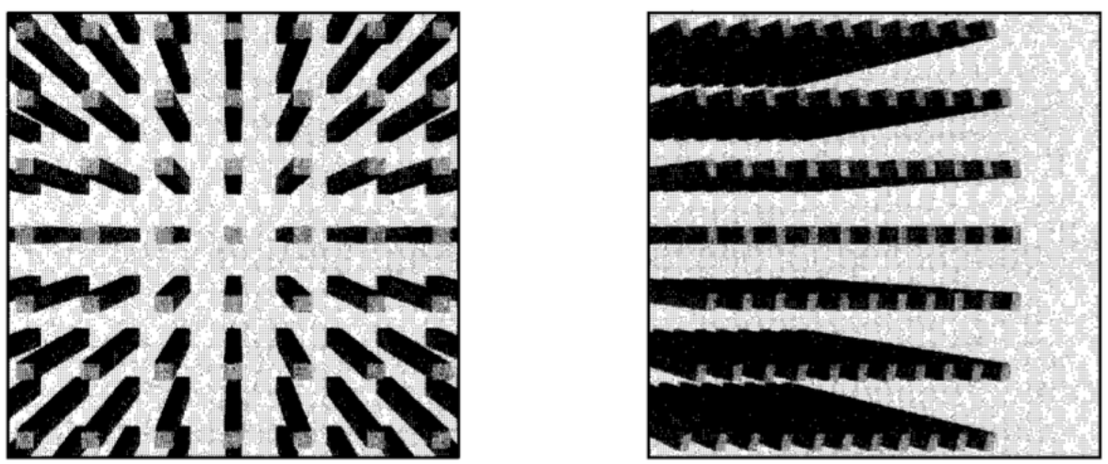


FIGURE 14-3 An array of rectangular bars is seen from above at rest in the figure on the left. In the right figure the bars are moving to the right with $v = 0.9c$. The bars appear to contract and rotate. Quoted from P.-K. Hsiung and R.H.P. Dunn, see *Science News* **137,** 232 (1990).

EXAMPLE 14.2

Use the Lorentz transformation to determine the time dilation effect.

Solution. Consider a clock fixed at a certain position (x_1) in the K system that produces signal indications with the interval

$$\Delta t = t(2) - t(1)$$

According to the Lorentz transformation (Equations 14.14), an observer in the moving system K' measures a time interval $\Delta t'$ (on the same clock) of

$$\Delta t' = t'(2) - t'(1)$$

$$= \frac{\left[t(2) - \frac{vx_1(2)}{c^2}\right] - \left[t(1) - \frac{vx_1(1)}{c^2}\right]}{\sqrt{1 - v^2/c^2}}$$

Because $x_1(2) = x_1(1)$ and because the clock is fixed in the K system, we have

$$\Delta t' = \frac{t(2) - t(1)}{\sqrt{1 - v^2/c^2}}$$

$$\boxed{\Delta t' = \frac{\Delta t}{\sqrt{1 - v^2/c^2}}} \tag{14.20}$$

Thus, to an observer in motion relative to the clock, the time intervals appear to be lengthened. This is the origin of the phrase "moving clocks run more slowly." Because the measured time interval on the moving clock is lengthened, the clock actually ticks slower. Notice that the clock is fixed in the K system, $x_1(1) = x_1(2)$, but not in the K' system, $x'_1(1) \neq x'_1(2)$.

The argument in the previous example can be reversed and the clock fixed in the K' system. The same result occurs; moving clocks run slower. The effect is called **time dilation**. It is important to note that the physical system is unimportant. The same effect occurs for a tuning fork, an hourglass, a quartz crystal, and a heartbeat. The problem is one of simultaneity. Events simultaneous in one system may not be simultaneous in another one moving with respect to the first. The same clock may be viewed from n different reference frames and found to be running at n different rates, simultaneously. Space and time are intricately interwoven. We shall return to this point later.

The time measured on a clock fixed in a system present at two events is called the **proper time** and given the symbol τ. For example, $\Delta t = \Delta\tau$ when a clock fixed in system K is present for both events, $x_1(1)$ and $x_1(2)$. Equation 14.20 becomes

$$\Delta t' = \gamma \Delta \tau \tag{14.21}$$

Notice that the proper time is always the minimum measurable time difference between two events. Moving observers always measure a longer time period

14.4 Experimental Verification of the Special Theory

The special theory of relativity explains the difficulties existing before 1900 with optics and electromagnetism. For example, the problems with stellar aberration and the Michelson-Morley experiment are solved by assuming no ether but requiring the Lorentz transformation.

But what about the new startling predictions of the special theory - length contraction and time dilation? These topics are addressed every day in the accelerator laboratories of nuclear and particle physics, where particles are accelerated to speeds close to that of light, and relativity must be considered. Other experiments can be performed with natural phenomena. We examine two of these.

Muon Decay

When cosmic rays enter the earth's outer atmosphere, they interact with particles and create cosmic showers. Many of the particles in these showers are π -mesons, which decay to other particles called muons. Muons are also unstable and decay according to the decay law, $N = N_0 e^{(-0.693t/t_{1/2})}$, where N_0 and N are the number of muons at time t = 0 and t, respectively, and $t_{1/2}$ is the half-life. However, enough muons reach the earth's surface that we can detect them easily.

Let us assume that we mount a detector on top of a 2,000-m mountain and count the number of muons traveling at a speed near v = 0.98c. Over a given period of time, we count 10^3 muons.

The half-life of muons is known to be 1.52 X 10~® s in their own rest frame (system K'). We move our detector to sea level and measure the number of muons (having v = 0.98c) detected during an equal period of time. What do we expect?

Determined classicjilly, muons traveling at a speed of 0.98c cover the 2,000 m in 6.8 X 10"® s, and 45 muons should survive the flight from 2,000 m to sea level according to the radioactive decay law. But experimental measurement indicates that 542 muons survive, a factor of 12 more.

This phenomenon must be treated relativistically. The decaying muons are moving at a high speed relative to the experimenters fixed on the earth. We therefore observe the muons' clock to be running slower. In the muons' rest frame, the time period of the muons' flight is not At = 6.8 X 10"® s but rather At/y. For v = 0.98c, y = 5, so we measure the flight time on a clock at rest in the muons' system to be 1.36 X 10"® s. The radioactive decay law predicts that 538 muons survive, much closer to our measurement and within the experimental uncertainties. An experiment similar to this has verified the time dilation prediction.

EXAMPLE 14.3

Examine the muon decay just discussed from the perspective of an observer moving with the muon.

Solution. The hjilf-life of the muon according to its own clock is 1.52 X 10^{-6} s But an observer moving with the muon would not measure the distance from the top of the mountain to sea level to be 2,000 m. According to that observer, the distance would be only 400 m. At a speed of 0.98c, it takes the muon only 1.36 X 10^{-6} s to travel the 400 m. An observer in the muon system would predict 538 muons to survive, in agreement with an observer on the earth.

Muon decay is an excellent example of a natural phenomenon that can be described in two systems moving with respect to each other. One observer sees time dilated and the other observer sees length contracted. Each, however, predicts a result in agreement with experiment.

Atomic Clock Time Measurements

An even more direct confirmation of special relativity was reported by two American physicists, J. C. Hafele and Richard E. Keating, in 1972. They used four extremely accurate cesium atomic clocks. Two clocks were flown on regularly scheduled commercial jet airplanes around the world, one eastward and one westward; the other two reference clocks stayed fixed on the earth at the U.S. Naval Observatory. A well-defined, hyperfine transition in the ground state of the ^{135}Cs atom has a frequency of 9,192,631,770 Hz and can be used as an accurate measurement of a time period.

The time measured on the two moving clocks was compared with that of the two reference clocks. The eastward trip lasted 65.4 hours with 41.2 flight hours. The westward trip, a week later, took 80.3 hours with 48.6 flight hours. The predictions are complicated by the rapid rotation of the earth and by a gravitational effect from the general theory of relativity.

We can gain some insight to the expected effect by neglecting the corrections and calculating the time difference as if the earth were not rotating. The circumference of the earth is about 4 X 10^7 m, and a typical jet airplane speed is almost 300 m/s. A clock fixed on the ground measures a flight time T_0 of

$$T_0 = \frac{4 \times 10^7\,\mathrm{m}}{300\,\mathrm{m/s}} = 1.33 \times 10^5\,\mathrm{s}(\approx 37\,\mathrm{hr}) \tag{14.22}$$

Because the moving clock runs more slowly, the observer on the earth would say that the moving clock measures only $T = T_0\sqrt{1-\beta^2}$. The time difference is

$$\begin{aligned} \Delta T &= T_0 - T = T_0(1 - \sqrt{1-\beta^2}) \\ &\approx \frac{1}{2}\beta^2 T_0 \end{aligned} \tag{14.23}$$

where only the first and second terms of the power series expansion for $\sqrt{1-\beta^2}$ are kept because β^2 is so small.

$$\Delta T = \frac{1}{2}\left(\frac{300\ \text{m/s}}{3 \times 10^8\ \text{m/s}}\right)^2 (1.33 \times 10^5\ \text{s}) \tag{14.24}$$

$$= 6.65 \times 10^{-8}\,\text{s} = 66.5\ \text{ns}$$

This time difference is greater than the uncertainty of the measurement. Notice that in this case, the clock left on the earth actually measures more time in seconds than the moving clock. This seems at variance with our earlier comments (see Equation 14.21 and discussion). But the time period referred to in Equation 14.21 is the time between two ticks, in this case, a transition in ^{133}Cs, which we measure in seconds. It is easy to remember that moving clocks run more slowly, so that in seconds the measured time difference involves fewer ticks and, according to the definition of a second, fewer seconds.

The actual predictions and observations for the time difference are

Travel	Predicted	Observed
Eastward	-40 ± 23 ns	-59 ± 10 ns
Westward	275 ± 21 ns	273 ± 7 ns

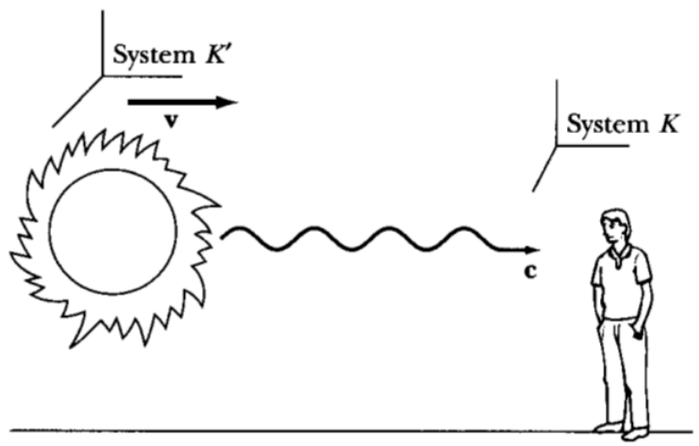
Again, the special theory of relativity is verified within the experimental uncertainties. A negative sign indicates that the time on the moving clock is less than the earth reference clock. The moving clocks lost time (ran slower) during the eastward trip and gained time (ran faster) during the westward trip. This difference is caused by the rotation of the earth, indicating that the flying clocks actually ticked faster or slower than the reference clocks on the earth. The overall positive time difference is a result of the gravitational potential effect (which we do not discuss here).

We have only briefly described two of the many experiments that have verified the special theory of relativity. There are no known experimental measurements that are inconsistent with the special theory of relativity. Einstein's work in this regard has so far withstood the test of time.

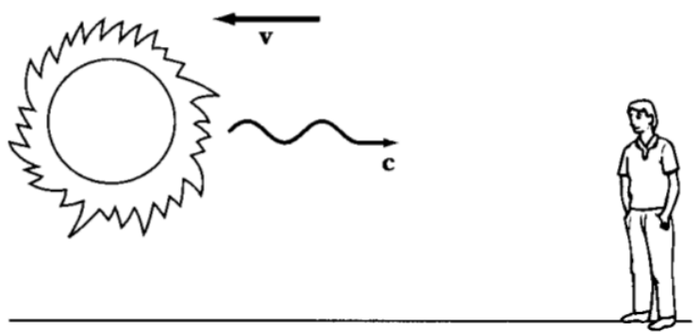
14.5 Relativistic Doppler Effect

The Doppler effect in sound is represented by an increased pitch of sound as a source approaches a receiver and a decrease of pitch as the source recedes. The change in frequency of the sound depends on whether the source or receiver is moving. This effect seems to violate Postulate I of the theory of relativity until we realize that there is a special frame for sound waves because there is a medium (e.g., air or water) in which the waves travel. In the case of light, however, there is no such medium. Only relative motion of source and receiver is meaningful in this context, and we should therefore expect some differences in the relativistic Doppler effect for light from the normal Doppler effect of sound.

Consider a source of light (e.g., a star) and a receiver approaching one another with relative speed v (Figure 14-4a).



(a) Source and receiver approaching



(b) Source and receiver receding

FIGURE 14-4 (a) An observer in system K sees light coming from a source fixed in system K'. System K' is moving toward the observer with speed v. The frequency of the light is observed in K to be increased over the value observed in K'. (b) When system K' is moving away from the observer, the frequency of the light decreases (the wavelength increases). This is the source of the term *redshifted*.

First, consider the receiver fixed in system K and the light source in system K' moving toward the receiver with speed v. During time Δt as measured by the receiver, the source emits n waves. During that time Δt , the total distance between the front and rear of the waves is

$$\text{length of wave train} = c\Delta t - v\Delta t \tag{14.25}$$

The wavelength is then

$$\lambda = \frac{c\Delta t - v\Delta t}{n} \tag{14.26}$$

and the frequency is

$$\nu = \frac{c}{\lambda} = \frac{cn}{c\Delta t - v\Delta t} \tag{14.27}$$

According to the source, it emits n waves of frequency ν_0 during the proper time $\Delta t'$:

$$n = \nu_0 \Delta t' \tag{14.28}$$

This proper time $\Delta t'$ measured on a clock in the source system is related to the time Δt measured on a clock fixed in system K of the receiver by

$$\Delta t' = \frac{\Delta t}{\gamma} \tag{14.29}$$

The clock moving with the source measures the proper time, because it is present at both the beginning and end of the waves.

Substituting Equation 14.29 into Equation 14.28, which in turn is substituted for n in Equation 14.27, gives

$$\begin{aligned} \nu &= \frac{1}{(1 - v/c)} \frac{\nu_0}{\gamma} \\ &= \frac{\sqrt{1 - v^2/c^2}}{1 - v/c} \nu_0 \end{aligned} \tag{14.30}$$

which can be written as

$$\nu = \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}}\nu_0 \quad \text{source and receiver approaching} \tag{14.31}$$

Equation 14.31 is also valid when the source is fixed and the receiver approaches it with speed v. Next, we consider the case in which the source and receiver recede from each other with velocity v (Figure 14-4b). The derivation is similar to the one just presented—with one small exception. In Equation 14.25, the distance between the beginning and end of the waves becomes

$$\text{length of wave train} = c\Delta t + v\Delta t \tag{14.32}$$

This change in sign is propagated through Equations 14.30 and 14.31, giving

$$\nu = \frac{\sqrt{1 - v^2/c^2}}{1 + v/c}\nu_0$$

$$\nu = \frac{\sqrt{1-\beta}}{\sqrt{1+\beta}}\nu_0 \quad \text{source and receiver receding} \tag{14.33}$$

Equations 14.31 and 14.33 can be combined into one equation.

$$\nu = \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}}\nu_0 \quad \text{relativistic Doppler effect} \tag{14.34}$$

if we agree to use a + sign for β (+v/ c) when the source and receiver are approaching each other and a - sign for β when they are receding.

The relativistic Doppler effect is important in astronomy. Equation 14.34 indicates that, if the source is receding at high speed from an observer, then a lower frequency (or longer wavelength) is observed for certain spectral lines or characteristic frequencies. This is the origin of the term red shift, the wavelengths of visible light are shifted toward longer wavelengths (red) if the source is receding from us.

Astronomical observations indicate that the universe is expanding. The farther away a star is, the faster it appears to be moving away (or the greater its red shift). These data are consistent with the "big bang" origin of the universe, which is estimated to have occurred some 13 billion years ago.

EXAMPLE 14.4

During a spaceflight to a distant star, an astronaut and her twin brother on the earth send radio signals to each other at annual intervals. What is the frequency of the radio signals each twin receives from the other during the flight to the star if the astronaut is moving at v = 0.8 c? What is the frequency during the return flight at the same speed?

Solution. We use Equation 14.34 to determine the frequency of radio signals that each receives from the other. The frequency ν_0 =1 signal/year. On the leg of the trip away from the earth, β = -0.8 and Equation 14.34 gives

$$\nu = \frac{\sqrt{1 - 0.8}}{\sqrt{1 + 0.8}}\nu_0$$

$$= \frac{\nu_0}{3}$$

The radio signals are received once every 3 years.

On the return trip, however, β = +0.8 and Equation 14.34 gives $\nu = 3\nu_0$, so the radio signals are received every 4 months. In this way, the twin on the earth can monitor the progress of his astronaut twin.

14.6 Twin Paradox

Consider twins who choose different career paths. Mary becomes an astronaut, and Frank decides to be a stockbroker At age 30, Mary leaves on a mission to a planet in a nearby star's system. Mary will have to travel at a high speed to reach the planet and return.

According to Frank, Mary's biological clock will tick more slowly during her trip, so she will age more slowly. He expects Mary to look and appear younger than he does when she returns. According to Mary, however, Frank will appear to be moving rapidly with respect to her system, and she thinks Frank will be younger when she returns. This is the paradox. Which twin, if either, is younger when Mary (the moving twin) returns to the earth where Frank (the fixed twin) has remained? Because the two expectations are so contradictory, doesn't Nature have a way to prove they will be the same age?

This paradox has existed almost since Einstein first published his special theory of relativity. Variations of the argument have been presented many times. The correct answer is that Mary, the astronaut, will return younger than her twin brother, Frank, who remains busy on Wall Street. The correct analysis is as follows. According to Frank, Mary's spaceship blasts off and quickly reaches a coasting speed of v = 0.8c, travels a distance of 8 ly (ly = a light year, the distance light travels in 1 year) to the planet, and quickly decelerates for a short visit to the planet. The acceleration and deceleration times are negligible compared with the total travel time of 10 years to the planet. The return trip also takes 10 years, so on Mary's return to Earth, Frank will be 30 + 10 + 10 = 50 years old. Frank calculates that Mary's clock is ticking slower and that each leg of the trip takes only $10\sqrt{1 - 0.8^2} = 6$ years. Mary therefore is only 30 + 6 + 6 = 42 years old when she returns. Frank's clock is (almost) in an inertial system.

When Mary performs the time measurements on her clock, they may be invalid according to the special theory because her system is not in an inertial frame of reference moving at a constant speed with respect to the earth. She accelerates and decelerates at both the earth and the planet, and to make valid time measurements to compare with Frank's clock, she must account for this acceleration and deceleration. The instantaneous rate of Mary's clock is still given by Equation 14.20, because the instantaneous rate is determined by the instantaneous speed v. Thus, there is no paradox if we obey the two postulates of the special theory. It is also clear which twin is in the inertial frame of reference. Mary will actually feel the forces of acceleration and deceleration. Frank feels no such forces. When Mary returns home, her twin brother has invested her 20 years of salary, making her a rich woman at the young age of 42. She was paid a 20-year salary for a job that took her only 12 years!

EXAMPLE 14.5

Mary and Frank send radio signals to each other at 1-year intervals after she leaves Earth. Analyze the times of receipt of the radio messages. Solution. In Example 14.4, we calculated that such radio signals are received every 3 years on the trip out and every $\frac{1}{3}$ year on the trip back. First, we examine the signals Mary receives from Frank. During the 6-year trip to the planet, Mary receives only two radio messages, but on the 6-year return trip, she receives eighteen signals, so she correctly concludes that her twin brother Frank has aged 20 years and is now 50 years old.

In Frank's system, Mary's trip to the planet takes 10 years. By the time Mary reaches the planet, Frank receives 10/3 signals (i.e., three signals plus one-third of the time to the next one). However, Frank continues to receive a signal every 3 years for the 8 years it takes the last signal Mary sends when she reaches the planet to travel to Frank. Thus, Frank receives signals every 3 years for 8 more years (total of 18 years) for a total of six radio signals from the period of travel to the planet. Frank has no way of knowing that Mary has stopped and turned around until the radio message, which takes 8 years, is received. Of the remaining 2 years of Mary's journey according to Frank (20 - 18 = 2), Frank receives signals every $\frac{1}{3}$ year, or six more signals. Frank correctly determines that Mary has aged 6 + 6 = 12 years during her journey because he receives a total of 12 signals.

Thus, both twins agree about their own ages and about each other's. Mary is 42 and Frank is 50 years old.

14.7 Relativistic Momentum

Newton's Second Law, $\mathbf{F} = d\mathbf{p}/dt$, is covariant under a Galilean transformation. Therefore, we do not expect it to keep its form under a Lorentz transformation. We can foresee difficulties with Newton's laws and the conservation laws unless we make some necessary changes. According to Newton's Second Law, for example, an acceleration at high speeds might cause a particle's velocity to exceed c, an impossible condition according to the special theory of relativity.

We begin by examining the conservation of linear momentum in a force-free (no external forces) collision. There are no accelerations. Observer A at rest in system K holds a ball of mass m, as does observer B in system K' moving to the right with relative speed v with respect to system K, as in Figure 14-1. The two observers throw their (identical) balls along their respective x_2-axes, which results in a perfectly elastic collision. The collision, according to observers in the two systems, is shown in Figure 14-5. Each observer measures the speed of his or her ball to be u_0.

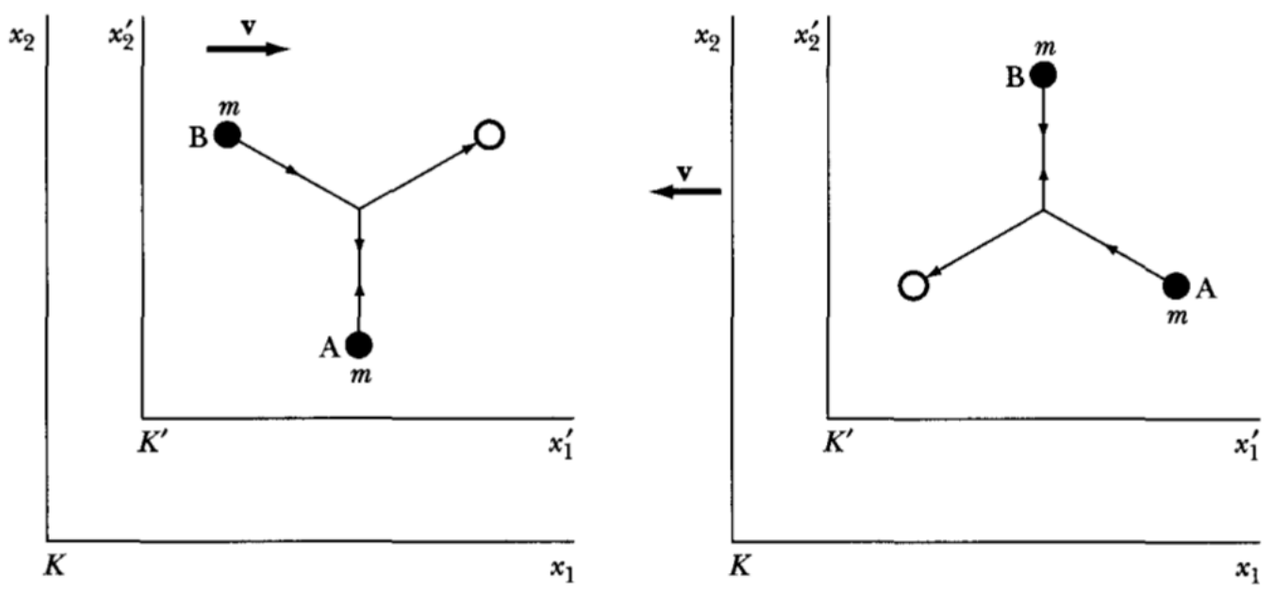


FIGURE 14-5 Observer A, at rest in fixed system K, throws a ball straight up in system K. Observer B, at rest in system K', which is moving to the right with velocity **v**, throws a ball straight down so that the two balls collide. (a) The collision according to observer A in system K. (b) The collision according to observer B in system K'. Each observer measures the speed of his or her ball to be u_0. We examine the linear momentum of the ball.

We first examine the conservation of momentum according to system K. The velocity of the ball thrown by observer A has components

$$\left.\begin{aligned} u_{A1} &= 0 \\ u_{A2} &= u_0 \end{aligned}\right\} \tag{14.35}$$

The momentum of ball A is in the x_2-direction:

$$p_{A2} = mu_0 \tag{14.36}$$

The collision is perfectly elastic, so the ball returns down with speed u_0. The change in momentum observed in system K is

$$\Delta p_{A2} = -2mu_0 \tag{14.37}$$

Does Equation 14.37 also represent the change in momentum of the ball thrown by observer B in the moving system K'? We use the inverse velocity transformation of Equations 14.17 (i.e., we interchange primes and unprimes and let $v \rightarrow -v$) to determine

$$\left.\begin{aligned} u_{B1} &= v \\ u_{B2} &= -u_0\sqrt{1 - v^2/c^2} \end{aligned}\right\} \tag{14.38}$$

where $u'_{B1} = 0$ and $u'_{B2} = u_0$. The momentum of ball B and its change in momentum during the collision become

$$p_{B2} = -mu_0\sqrt{1 - v^2/c^2} \tag{14.39}$$

$$\Delta p_{B2} = +2mu_0\sqrt{1 - v^2/c^2} \tag{14.40}$$

Equations 14.37 and 14.40 do not add to zero: *Linear momentum is not conserved according to the special theory if we use the conventions for momentum of classical physics*. Rather than abandoning the law of conservation of momentum, we look for a solution that allows us to retain both it and Newton's Second Law.

As we did for the Lorentz transformation, we assume the simplest possible change. We assume that the classical form of momentum mu is multiplied by a constant that may depend on speed k(u):

$$\mathbf{p} = k(u)\,m\mathbf{u} \tag{14.41}$$

In Example 14.6, we show that the value

$$k(u) = \frac{1}{\sqrt{1 - u^2/c^2}} \tag{14.42}$$

allows us to retain the conservation of linear momentum. Notice that the form of Equation 14.42 is the same as that found for the Lorentz transformation. In fact, the constant k(u) is given the same label: λ . However, this λ contains the speed of the particle u, whereas the Lorentz transformation contains the relative speed v between the two inertial reference frames. This distinction must be kept in mind; it often causes confusion.

We can make a plausible calculation for the relativistic momentum if we use the proper time τ (see Equation 14.21) rather than the normal time t. In this case

$$\mathbf{p} = m\frac{d\mathbf{x}}{d\tau} = m\frac{d\mathbf{x}}{dt}\frac{dt}{d\tau} \tag{14.43}$$

$$= m\frac{d\mathbf{x}}{dt}\frac{1}{\sqrt{1 - u^2/c^2}} \tag{14.44}$$

$$\boxed{\mathbf{p} = \frac{m\mathbf{u}}{\sqrt{1 - u^2/c^2}} = \gamma m\mathbf{u}} \quad \text{relativistic momentum} \tag{14.45}$$

where we retain **u** = d**x**/dt as used classically.

Although all observers do not agree as to d**x**/dt, they do agree as to d**x**/dτ , where the proper time dτ is measured by the moving object itself. The relation dt/dτ is obtained from Equation 14.21, where the speed u has been used in γ to represent the speed of a reference frame fixed in the object that is moving with respect to a fixed frame.

Equation 14.45 is our new definition of momentum, called **relativistic momentum**. Notice that it reduces to the classical result for small values of u/c. It was fashionable in past years to call the mass in Equation 14.45 the **rest mass** m_0 and to call the term

$$m = \frac{m_0}{\sqrt{1 - u^2/c^2}} \quad \textbf{(old-fashioned notation)} \tag{14.46}$$

the **relativistic mass**. The term rest mass resulted from Equation 14.46 when u = 0, and the classical form of momentum was thus retained: **p** = m**u**. Scientists spoke of the mass increasing at high speeds. We prefer to keep the concept of mass as an invariant, intrinsic property of an object. The use of the two terms *relativistic* and *rest mass* is now considered old-fashioned, although the terms are still sometimes used. *We always refer to the mass m, which is the same as the rest mass*. The use of relativistic mass often leads to mistakes when using classical expressions.

EXAMPLE 14.6

Show that linear momentum is conserved in the x_2-direction for the collision shown in Figure 14-5 if relativistic momentum is used.

Solution. We can modify the classical expressions for momentum already obtained for the two balls. The momentum for ball A becomes (from Equation 14.36)

$$p_{A2} = \frac{mu_0}{\sqrt{1 - u_0^2/c^2}} \tag{14.47}$$

and

$$\Delta p_{A2} = \frac{-2mu_0}{\sqrt{1 - u_0^2/c^2}} \tag{14.48}$$

Before modifying Equation 14.39 for the momentum of ball B, we must first find the speed of ball B as measured in system K We use Equation 14.38 to determine

$$\begin{aligned} u_B &= \sqrt{u_{B1}^2 + u_{B2}^2} \\ &= \sqrt{v^2 + u_0^2(1 - v^2/c^2)} \end{aligned} \tag{14.49}$$

The momentum p_{B2} is found by modifying Equation 14.39:

$$p_{B2} = -mu_0\gamma\sqrt{1 - v^2/c^2}$$

where

$$\gamma = \frac{1}{\sqrt{1 - u_B^2/c^2}}$$

$$p_{B2} = \frac{-mu_0\sqrt{1 - v^2/c^2}}{\sqrt{1 - u_B^2/c^2}} \tag{14.50}$$

Using u_B from Equation 14.49 gives

$$p_{B2} = \frac{-mu_0\sqrt{1 - v^2/c^2}}{\sqrt{(1 - u_0^2/c^2)(1 - v^2/c^2)}}$$

$$= \frac{-mu_0}{\sqrt{1 - u_0^2/c^2}} \tag{14.51}$$

$$\Delta p_{B2} = \frac{+2mu_0}{\sqrt{1 - u_0^2/c^2}} \tag{14.52}$$

Equations 14.48 and 14.52 add to zero, as required for the conservation of linear momentum.

14.8 Energy

With a new definition of linear momentum (Equation 14.45) in hand, we turn our attention to energy and force. We keep our former definition (Equation 2.86) of kinetic energy as being the work done on a particle. The work done is defined in Equation 2.84 to be

$$W_{12} = \int_1^2 \mathbf{F} \cdot d\mathbf{r} = T_2 - T_1 \tag{14.53}$$

Equation 2.2 for Newton's Second Law is modified to account for the new definition of linear momentum:

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d}{dt}(\gamma m\mathbf{u}) \tag{14.54}$$

If we start from rest, $T_1 = 0$, and the velocity **u** is initially along the direction of the force.

$$W = T = \int \frac{d}{dt}(\gamma m\mathbf{u}) \cdot \mathbf{u}\, dt \tag{14.55}$$

$$= m\int_0^u u\, d(\gamma u) \tag{14.56}$$

Equation 14.56 is integrated by parts to obtain

$$T = \gamma mu^2 - m\int_0^u \frac{udu}{\sqrt{1 - u^2/c^2}}$$

$$= \gamma mu^2 + mc^2\sqrt{1 - u^2/c^2}\,\Big|_0^u$$

$$= \gamma mu^2 + mc^2\sqrt{1 - u^2/c^2} - mc^2 \tag{14.57}$$

With algebraic manipulation. Equation 14.57 becomes

$$\boxed{T = \gamma mc^2 - mc^2} \quad \textbf{relativistic kinetic energy} \tag{14.58}$$

Equation 14.58 seems to resemble in no way our former result for kinetic energy, $T = \frac{1}{2}mu^2$. However, Equation 14.58 must reduce to $\frac{1}{2}mu^2$ for small values of velocity.

EXAMPLE 14.7

Show that Equation 14.58 reduces to the classical result for small speeds, $u << c$.

Solution. The first term of Equation 14.58 can be expanded in a power series:

$$T = mc^2 + \frac{1}{2}mu^2 - mc^2$$
$$= \frac{1}{2}mu^2 \tag{14.59}$$

where all terms of power $(u/c)^4$ or greater are neglected because $u << c$.

$$T = mc^2 + \frac{1}{2}mu^2 - mc^2$$
$$= \frac{1}{2}mu^2 \tag{14.60}$$

which is the classical result.

It is important to note that neither $\frac{1}{2}mu^2$ nor $\frac{1}{2}mu^2$ gives the correct relativistic value for the kinetic energy.

The term mc^2 in Equation 14.58 is called the **rest energy** and is denoted by Eq.

$$\boxed{E_0 \equiv mc^2} \quad \text{rest energy} \tag{14.61}$$

Equation 14.58 is rewritten

$$\gamma mc^2 = T + mc^2$$

Thus,

$$E = T + E_0 \tag{14.62}$$

where

$$\boxed{E \equiv \gamma mc^2 = T + E_0} \quad \text{total energy} \tag{14.63}$$

The total energy, $E = \gamma mc^2$, is defined as the sum of kinetic energy and the rest energy. Equations 14.58-14.63 are the origin of Einstein's famous relativistic result of the equivalence of mass and energy (energy = mc^2). These equations are consistent with this interpretation. Note that when a body is not in motion (u = 0 = T), Equation 14.63 indicates that the total energy is equal to the rest energy

If mass is simply another form of energy, then we must combine the classical conservation laws of mass and energy into one conservation law of mass-energy represented by Equation 14.63. This law is easily demonstrated in the atomic nucleus, where the mass of constituent particles is converted to the energy that binds the individual particles together.

EXAMPLE 14.8

Use the atomic masses of the particles involved to calculate the binding energy of a deuteron.

Solution. A deuteron is composed of a neutron and a proton. We use atomic masses, because the electron masses cancel.

$$\begin{aligned} \text{mass of neutron} &= 1.008665\ \text{u} \\ \text{mass of proton } (^1\text{H}) &= \underline{1.007825\ \text{u}} \\ \text{sum} &= 2.016490\ \text{u} \\ \text{mass of deuteron } (^2\text{H}) &= 2.014102\ \text{u} \\ \text{difference} &= 0.002388\ \text{u} \end{aligned}$$

This difference in mass-energy is equal to the binding energy holding the neutron and proton together as a deuteron. The mass units are atomic mass units (u), which can be converted to kilograms if necessary. However, the conversion of mass to energy is facilitated by the well-known relation between mass and energy:

$$1\ \text{u}c^2 = 931.5\ \text{MeV} \tag{14.64}$$

The binding energy of the deuteron is therefore

$$0.002388\ \text{u}c^2 \times 931.5 \frac{\text{MeV}}{\text{u}c^2} = 2.22\ \text{MeV}$$

Nuclear experiments of the form $\gamma + {}^2\text{H} \rightarrow \text{n} + \text{p}$ indicate that gamma rays of energy just greater than 2.22 MeV are required to break the deuteron apart into a neutron and a proton. Conversely, when a neutron and proton join at rest to form a deuteron, 2.22 MeV of energy is released in the form of kinetic energy

Because physicists believe that momentum is a more fundamental concept than kinetic energy (for example, there is no general law of conservation of kinetic energy), we would like a relation for mass-energy that includes momentum rather than kinetic energy. We begin with Equation 14.45 for momentum:

$$p = \gamma m u$$

$$p^2c^2 = \gamma^2 m^2 u^2 c^2$$

$$= \gamma^2 m^2 c^4 \left(\frac{u^2}{c^2}\right) \tag{14.65}$$

It is easy to show that

$$\frac{u^2}{c^2} = 1 - \frac{1}{\gamma^2} \tag{14.66}$$

so Equation 14.65 becomes

$$p^2c^2 = \gamma^2 m^2 c^4 \left(1 - \frac{1}{\gamma^2}\right)$$

$$= \gamma^2 m^2 c^4 - m^2 c^4$$

$$= E^2 - E_0^2$$

$$\boxed{E^2 = p^2c^2 + E_0^2} \tag{14.67}$$

Equation 14.67 is a very useful kinematic relationship. It relates the total energy of a particle to its momentum and rest energy.

Notice that a photon has no mass, so that Equation 14.67 gives

$$E = pc \quad \text{photon} \tag{14.68}$$

There is no such thing as a photon at rest

14.9 Spacetime and Four-Vectors

In Section 14.3 (Equation 14.5), we noticed that the quantities

$$\left.\begin{aligned}\sum_{j=1}^{3} x_j^2 - c^2t^2 = 0\\ \sum_{j=1}^{3} x_j'^2 - c^2t'^2 = 0\end{aligned}\right\}$$

are invariant because the speed of light is the same in all inertial systems in relative motion. Consider two events separated by space and time. In system K,

$$\Delta x_i = x_i(\text{event 2}) - x_i(\text{event 1})$$

$$\Delta t = t(\text{event 2}) - t(\text{event 1})$$

The interval Δs^2 is invariant in all inertial systems in relative motion

$$\Delta s^2 = \sum_{j=1}^{3} (\Delta x_j)^2 - c^2\Delta t^2 \tag{14.69}$$

$$\Delta s^2 = \Delta s'^2 = \sum_{j=1}^{3} (\Delta x_j')^2 - c^2\Delta t'^2 \tag{14.70}$$

Equation 14.69 can be written as a differential equation:

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 - c^2dt^2 \tag{14.71}$$

Consider the system K', where the particle is instantaneously at rest. Because $dx'_1 = dx'_2 = dx'_3 = 0$ in this case, $dt' = d\tau$, the proper time interval discussed above (Equation 14.21). Equation 14.70 becomes

$$-c^2d\tau^2 = dx_1^2 + dx_2^2 + dx_3^2 - c^2dt^2 \tag{14.72}$$

Using the Lorentz transformation, Equation 14.72 gives a similar result to Equation 14.21:

$$d\tau = \frac{dt}{\gamma} \tag{14.73}$$

The proper time τ is, along with the length quantity Δs^2, another Lorentz invariant quantity.

A useful concept in special relativity is that of the **light cone**. The invariant length Δs^2 suggests adding ct as a fourth dimension to the three space dimensions x_1, x_2, and x_3. In Figure 14-6, we plot ct versus one of the Euclidean space coordinates. The origin of (x, ct) is the present (0, 0). The solid lines represent the paths taken in the past and in the future by light. A particle traveling the path from A to B is said to be moving along its **worldline**. For time $t < 0$, the particle has been in the lower cone, the past. Similarly, for $t > 0$ the particle will move in the upper cone, the future. It is not possible for us to know about events outside the light cone; this region, called "elsewhere," requires $v > c$.

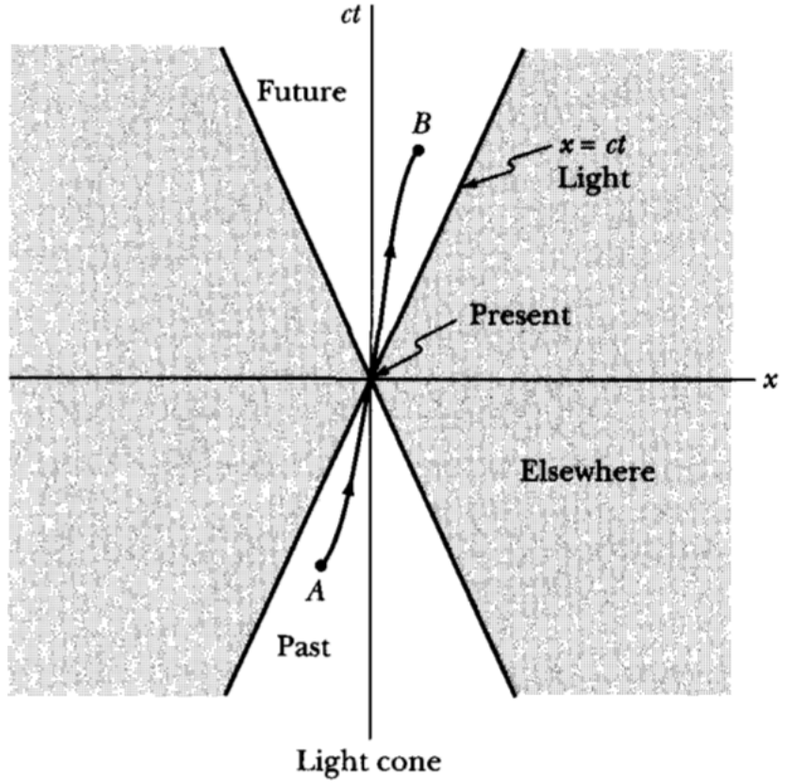


FIGURE 14-6 The variable *ct* is plotted versus *x* with the origin being the present. The heavy solid lines indicate the past and future paths of light and form a *light cone*. To the right and left of these lines is considered "elsewhere," because we cannot reach this region from the present. The path from *A* to *B* represents a *worldline*, a path that we can take traveling at speeds less than or equal to light.

There are two possibilities concerning the value of Δs^2. If $\Delta s^2 > 0$, the two events have a spacelike interval. One can always find an inertial frame traveling with v<csuch that the two events occur at different space coordinates but at the same time. When $\Delta s^2 < 0$, the two events are said to have a timelike interval. One can always find a suitable inertial frame in which the events occur at the same point in space but at different times. In the case $\Delta s^2 = 0$, the two events are separated by a light ray.

Only events separated by a timelike interval can be causally connected. The present event in the light cone can be causally related only to events in the past region of the light cone. Events with a spacelike interval cannot be causally connected. Space and time, although distinct, are nonetheless intricately related.

The previous discussion of space and time suggests using ct as a fourth dimensional parameter We continue this line of thought by defining $x_4 = ict$ and $x'_4 = ict'$. The use of the imaginary number i does not indicate that this component is imaginary. The imaginary number simply allows us to represent the relations in concise, mathematical form.

The rest of this section could just as well be carried out without the use of i(e.g., $x_4 = ct$), but the mathematics would be more cumbersome. The useful results are in terms of real, physical quantities.

Using $x_4 = ict$ and $x'_4 = ict'$, we can write Equations 14.5 as

$$\left.\begin{aligned} \sum_{\mu=1}^{4} x_\mu^2 &= 0 \\ \sum_{\mu=1}^{4} x_\mu'^2 &= 0 \end{aligned}\right\} \tag{14.74}$$

In accordance with standard convention, we use Greek indices (usually μ or ν) to indicate summations that run from 1 to 4; in relativity theory, Latin indices are usually reserved for summations that run from 1 to 3.

From these equations, it is clear that the two sums must be proportional(proof is given in Appendix G), and because the motion is symmetrical between the systems, the proportionality constant is unity. Thus,

$$\sum_{\mu} x_{\mu}^2 = \sum_{\mu} x_{\mu}'^2 \tag{14.75}$$

This relation is analogous to the three-dimensional, distance-preserving, orthogonal rotations we have studied previously (see Section 1.4) and indicates that the Lorentz transformation corresponds to a rotation in a four-dimensional space (called **world space** or **Minkowski space**). The Lorentz transformations are then orthogonal transformations in Minkowski space:

$$x_{\mu}' = \sum_{\nu} \lambda_{\mu\nu} x_{\nu} \tag{14.76}$$

where the $\lambda_{\mu\nu}$ are the elements of the Lorentz transformation matrix. From Equations 14.14, the transformation λ is

$$\boldsymbol{\lambda} = \begin{pmatrix} \gamma & 0 & 0 & i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \tag{14.77}$$

A quantity is called a **four-vector** if it consists of four components, each of which transforms according to the relation

$$A_{\mu}' = \sum_{\nu} \lambda_{\mu\nu} A_{\nu} \tag{14.78}$$

where the $\lambda_{\mu\nu}$ define a Lorentz transformation. Such a four-vector(openface capital letters) is

$$\mathbb{X} = (x_1, x_2, x_3, ict) \tag{14.79a}$$

or

$$\boxed{\mathbb{X} = (\mathbf{x}, ict)} \tag{14.79b}$$

where the notation of the last line means that the first three (space) components of $\mathbb{X}$ define the ordinary three-dimensional position vector **x** and that the fourth component is ict. Similarly, the differential of $\mathbb{X}$ is a four-vector:

$$d\mathbb{X} = (d\mathbf{x},\ ic\,dt) \tag{14.80}$$

In Minkowski space, the four-dimensional element of length is invariant. Its magnitude is unaffected by a Lorentz transformation, and such a quantity is called a four-scalar or world scalar. Equation 14.71 can be written as

$$ds = \sqrt{\sum_{\mu} dx_\mu^2} \tag{14.81}$$

and Equation 14.72 as

$$d\tau = \frac{i}{c}\sqrt{\sum_{\mu} dx_\mu^2} = \frac{i}{c}\,ds \tag{14.82}$$

The proper time dr is invariant because it is simply i/ c times the element of The proper time dτ is invariant because it is simply i/ c times the element of length ds. The ratio of the four-vector $d\mathbb{X}$ to the invariant dτ is therefore also a four-vector, called the four-vector velocity $\mathbb{V}$:

$$\boxed{\mathbb{V} = \frac{d\mathbb{X}}{d\tau} = \left(\frac{d\mathbf{x}}{d\tau},\ ic\frac{dt}{d\tau}\right)} \tag{14.83}$$

The components of the ordinary velocity u are

$$u_j = \frac{dx_j}{dt}$$

so, using Equations 14.71 and 14.82, dτ can be expressed as

$$d\tau = dt\sqrt{1 - \frac{1}{c^2}\sum_j \frac{dx_j^2}{dt^2}}$$

or

$$d\tau = dt\sqrt{1-\beta^2} \tag{14.84}$$

as we found in Equation 14.73. The four-vector velocity can therefore be written as

$$\boxed{\mathbb{V} = \frac{1}{\sqrt{1-\beta^2}}(\mathbf{u}, ic)} \tag{14.85}$$

where **u** represents the three space components of ordinary velocity, u_1, u_2, u_3. (Remember that the particle's velocity is now denoted by **u** to distinguish it from the moving frame velocity **v**.) The four-vector momentum is now simply the mass times four-vector velocity, because mass is invariant;

$$\mathbb{P} = m\mathbb{V} \tag{14.86}$$

$$\boxed{\mathbb{P} = \left(\frac{m\mathbf{u}}{\sqrt{1-\beta^2}}, ip_4\right)} \tag{14.87}$$

where

$$p_4 \equiv \frac{mc}{\sqrt{1-\beta^2}} \tag{14.88}$$

The first three components of the four-vector momentum $\mathbb{P}$ are the components of the relativistic momentum (Equation 14.45):

$$P_j = p_j = \gamma m u_j, \quad j = 1, 2, 3 \tag{14.89}$$

Using Equation 14.63, the fourth component of the momentum is related to the total energy E:

$$p_4 = \gamma mc = \frac{E}{c} \tag{14.90}$$

The four-vector momentum can therefore be written as

$$\mathbb{P} = \left(\mathbf{p}, i\frac{E}{c}\right) \tag{14.91}$$

where **p** stands for the three space components of momentum. Thus, in relativity theory, momentum and energy are linked in a manner similar to that which joins the concepts of space and time. If we apply the Lorentz transformation matrix (Equation 14.77) to the momentum $\mathbb{P}$, we find

$$\boxed{\begin{aligned} p_1' &= \frac{p_1 - (v/c^2)E}{\sqrt{1-\beta^2}} \\ p_2' &= p_2 \\ p_3' &= p_3 \\ E' &= \frac{E - vp_1}{\sqrt{1-\beta^2}} \end{aligned}} \tag{14.92}$$

EXAMPLE 14.9

Using the methods of this section, derive Equation 14.67.

Solution. If we place the origin of the moving system K' fixed on the particle, we have u = v. The square of the four-vector velocity (Equation 14.85) is invariant:

$$\mathbb{V}^2 = \sum_\mu V_\mu^2 = \frac{v^2 - c^2}{1-\beta^2} = -c^2 \tag{14.93}$$

Hence, the square of the four-vector momentum is also invariant:

$$\mathbb{P}^2 = \sum_\mu P_\mu^2 = m^2\mathbb{V}^2 = -m^2c^2 \tag{14.94}$$

From Equation 14.91, we also have, using $\mathbf{p}\cdot\mathbf{p} = p^2 = p_1^2 + p_2^2 + p_3^2$

$$\mathbb{P}^2 = p^2 - \frac{E^2}{c^2} \tag{14.95}$$

Combining the last two equations gives Equation 14.67.

$$E^2 = p^2c^2 + m^2c^4 = p^2c^2 + E_0^2$$

If we define an angle ϕ such that $\beta = \sin\phi$, the relativistic relations between velocity, momentum, and energy can be obtained by trigonometric relations involving the so-called "relativistic triangle" (Figure 14-7).

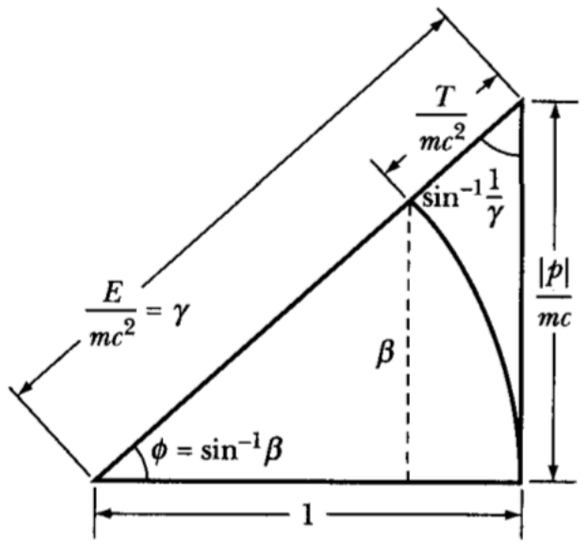


FIGURE 14-7 The relativistic triangle allows us to find relations between velocity, momentum, and energy by using trigonometric relations.

EXAMPLE 14.10

Derive the velocity addition rule.

Solution. Suppose that there are three inertial reference frames, K, K', and K", which are in collinear motion along their respective x_1-axes. Let the velocity of K' relative to K be v_1 and let the velocity of K" relative to K' be v_2. The speed of K" relative to K cannot be $v_1 + v_2$, because it must be possible to propagate a signal between any two inertial frames, and if both v_1 and v_2 are greater than c/2 (but less than c), then $v_1 + v_2 > c$. Therefore, the rule for the addition of velocities in relativity must be different from that in Galilean theory. The relativistic velocity addition rule can be obtained by considering the Lorentz transformation matrix connecting K and K". The individual transformation matrices are

$$\boldsymbol{\lambda}_{K'\to K} = \begin{pmatrix} \gamma_1 & 0 & 0 & i\beta_1\gamma_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta_1\gamma_1 & 0 & 0 & \gamma_1 \end{pmatrix}$$

$$\boldsymbol{\lambda}_{K''\to K'} = \begin{pmatrix} \gamma_2 & 0 & 0 & i\beta_2\gamma_2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta_2\gamma_2 & 0 & 0 & \gamma_2 \end{pmatrix}$$

The transformation from K" to K is just the product of these two transformations:

$$\boldsymbol{\lambda}_{K''\to K} = \boldsymbol{\lambda}_{K''\to K'}\boldsymbol{\lambda}_{K'\to K} = \begin{pmatrix} \gamma_1\gamma_2(1+\beta_1\beta_2) & 0 & 0 & i\gamma_1\gamma_2(\beta_1+\beta_2) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\gamma_1\gamma_2(\beta_1+\beta_2) & 0 & 0 & \gamma_1\gamma_2(1+\beta_1\beta_2) \end{pmatrix}$$

So that the elements of this matrix correspond to those of the normal Lorentz matrix (Equation 14.77), we must identify β and γ for the K" —> K transformation as

$$\left.\begin{aligned} \gamma &= \gamma_1\gamma_2(1+\beta_1\beta_2) \\ \beta\gamma &= \gamma_1\gamma_2(\beta_1+\beta_2) \end{aligned}\right\} \tag{14.96}$$

from which we obtain

$$\beta = \frac{\beta_1+\beta_2}{1+\beta_1\beta_2} \tag{14.97}$$

If we multiply this last expression by c, we have the usual form of the velocity (speed) addition rule:

$$\boxed{v = \frac{v_1+v_2}{1+(v_1v_2/c^2)}} \tag{14.98}$$

It follows that if $v_1 < c$ and $v_2 < c$, then $v < c$ also.

Even though *signal* velocities can never exceed c, there are other types of velocity that can be greater than c. For example, the phase velocity of a light wave in a medium for which the index of refraction is less than unity is greater than c, but the phase velocity does not correspond to the signal velocity in such a medium; the signal velocity is indeed less than c. Or consider an electron gun that emits a beam of electrons. If the gun is rotated, then the electron beam describes a certain path on a screen placed at some appropriate distance. If the angular velocity of the gun and the distance to the screen are sufficiently large, then the velocity of the spot traveling across the screen can be *any* velocity, arbitrarily large. Thus, the *writing speed* of an oscilloscope can exceed c, but again the writing speed does not correspond to the signal velocity; that is, information cannot be transmitted from one point on the screen to another by means of the electron beam. In such a device, a signal can be transmitted only from the gun to the screen, and this transmission takes place at the velocity of the electrons in the beam (i.e., < c).

EXAMPLE 14.11

Derive the relativistic Doppler effect if the angle between the light source and direction of relative motion of the observer is θ (Figure 14-8).

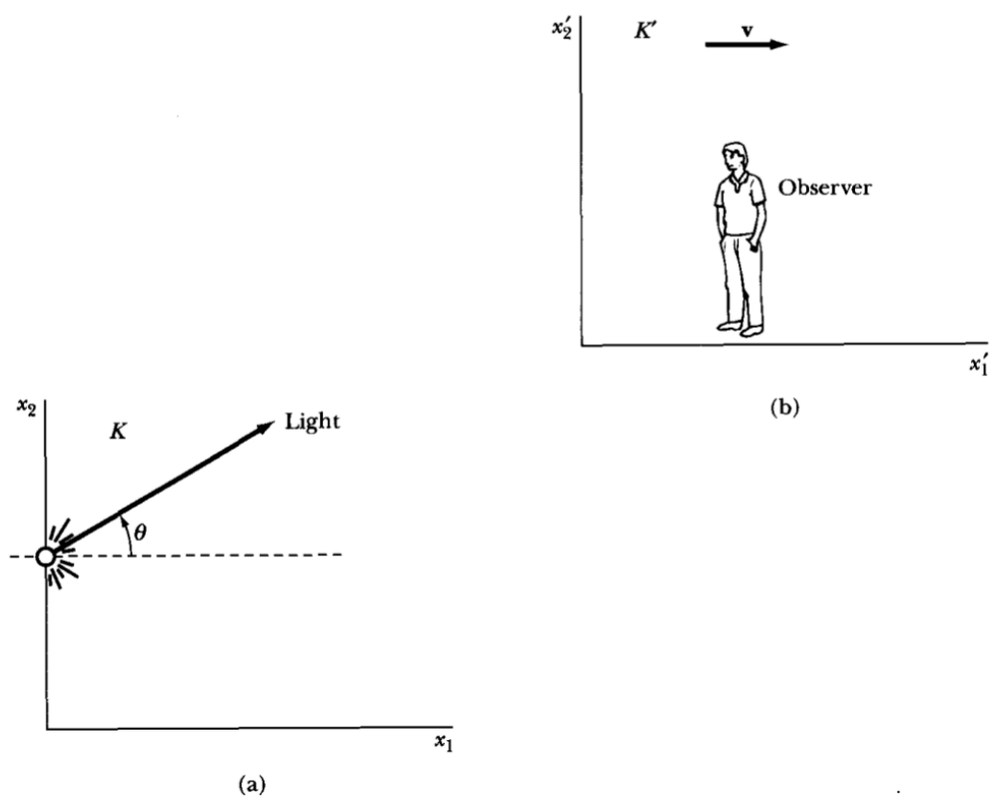


FIGURE 14-8 A light source fixed in system K emits light at a single frequency ν_0. An observer in system K', moving to the right at velocity **v** with respect to K, measures the light frequency to be ν'.

Solution. This example can easily be solved using the momentum-energy four-vector by treating the light as a photon with total energy E = $h\nu$. The light source is at rest in system K and emits a single frequency ν_0 .

$$E = h\nu_0 \tag{14.99}$$

$$p = \frac{E}{c} = \frac{h\nu_0}{c} \tag{14.100}$$

The observer moving to the right in system K' measures the energy E' for a photon of frequency ν' . From Equation 14.92, we have

$$E' = \gamma(h\nu_0 - vp_1) \tag{14.101}$$

$$h\nu' = \gamma\left(h\nu_0 - \frac{vh\nu_0}{c}\cos\theta\right) \tag{14.102}$$

where $p_1 = p\cos\theta$. Equation 14.102 reduces to

$$\nu' = \gamma\nu_0(1 - \beta\cos\theta) \tag{14.103}$$

which is equivalent to Equation 14.34, depending on the value of θ . For an early time, the observer is far to the left of the source, and as the observer approaches the source ($\theta = \pi$),

$$\nu' = \nu_0\frac{\sqrt{1+\beta}}{\sqrt{1-\beta}} \quad \text{observer approaching source} \tag{14.104}$$

as in Equation 14.31. At a much later time, the observer is receding ($\theta = 0$) and

$$\nu' = \nu_0\frac{\sqrt{1-\beta}}{\sqrt{1+\beta}} \quad \text{observer receding from source} \tag{14.105}$$

as in Equation 14.33. When the observer just passes the source ($\theta = \pi/2$),

$$\nu' = \frac{\nu_0}{\sqrt{1-\beta^2}} \qquad \text{observer passing source} \tag{14.106}$$

We can also treat the case where the observer is at rest and the source is moving. We still obtain Equations 14.104 -14.106 because, according to the principle of relativity, it is not possible to distinguish between the motion of the observer and the motion of the source.

14.10 Lagrangian Function in Special Relativity

Lagrangian and Hamiltonian dynamics (discussed in Chapter 7) must be adjusted in light of the new concepts presented here. We can extend the Lagrangian formalism into the realm of special relativity in the following way. For a single (non-relativistic) particle moving in a velocity-independent potential, the rectangular momentum components (see Equation 7.150) may be written as

$$p_i = \frac{\partial L}{\partial u_i} \tag{14.107}$$

According to Equation 14.87, the relativistic expression for the ordinary (i.e., space) momentum component is

$$p_i = \frac{m u_i}{\sqrt{1-\beta^2}} \tag{14.108}$$

We now require that the *relativistic* Lagrangian, when differentiated with respect to u_i as in Equation 14.107, yield the momentum components given by Equation 14.108:

$$\frac{\partial L}{\partial u_i} = \frac{m u_i}{\sqrt{1-\beta^2}} \tag{14.109}$$

This requirement involves only the *velocity* of the particle, so we expect that the velocity-*independent* part of the relativistic Lagrangian is unchanged from the nonrelativistic case. The velocity-*dependent* part, however, may no longer be equal to the kinetic energy. We therefore write

$$L = T^* - U \tag{14.110}$$

where $U = U(x_i)$ and $T^* = T^*(u_i)$. The function T^* must satisfy the relation

$$\frac{\partial T^*}{\partial u_i} = \frac{m u_i}{\sqrt{1 - \beta^2}} \tag{14.111}$$

It can be easily verified that a suitable expression for T^* (apart from a possible constant of integration that can be suppressed) is

$$T^* = -mc^2\sqrt{1 - \beta^2} \tag{14.112}$$

Hence, the relativistic Lagrangian can be written as

$$\boxed{L = -mc^2\sqrt{1 - \beta^2} - U} \tag{14.113}$$

and the equations of motion are obtained in the standard way from Lagrange's equations.

Notice that the Lagrangian is not given by T - U, because the relativistic expression for the kinetic energy (Equation 14.58) is

$$T = \frac{mc^2}{\sqrt{1 - \beta^2}} - mc^2 \tag{14.114}$$

The Hamiltonian (see Equation 7.153) can be calculated from

$$\begin{aligned} H &= \sum_i u_i p_i - L \\ &= \sum_i \frac{p_i^2 c^2}{\gamma mc^2} + \frac{mc^2}{\gamma} + U \end{aligned}$$

where we have used Equations 14.108 and 14.113 and changed $\sqrt{1 - \beta^2}$ to γ^{-1}. Thus,

$$H = \frac{p^2c^2}{\gamma mc^2} + \frac{mc^2}{\gamma} + U = \frac{1}{\gamma mc^2}(p^2c^2 + m^2c^4) + U$$

$$= \frac{E^2}{\gamma mc^2} + U$$

$$= E + U = T + U + E_0 \tag{14.115}$$

The relativistic Hamiltonian is equal to the total energy defined in Section 14.8 plus the potential energy. It differs from the total energy used previously in Chapter 7 by now including the rest energy.

14.11 Relativistic Kinematics

In the event that the velocities in a collision process are not negligible with respect to the velocity of light, it becomes necessary to use relativistic kinematics. In the discussion in Chapter 9, we took advantage of the properties of the center-of-mass coordinate system in deriving many of the kinematic relations. Because mass and energy are interrelated in relativity theory, it no longer is meaningful to speak of a "center-of-mass" system; in relativistic kinematics, one uses a "center-of-momentum" coordinate system instead. Such a system possesses the same essential property as the previously used center-of-mass system—the total linear momentum in the system is zero. Therefore, if a particle of mass m_1 collides elastically with a particle of mass m^, then in the center-of-momentum system we have

$$p_1' = p_2' \tag{14.116}$$

Using Equation 14.87, the space components of the momentum four-vector can be written as

$$m_1 u_1' \gamma_1' = m_2 u_2' \gamma_2' \tag{14.117}$$

where, as before, $\gamma = 1/\sqrt{1-\beta^2}$ and β = u/c.

In a collision problem, it is convenient to associate the laboratory coordinate system with the inertial system K and the center-of-momentum system with K' (see Figure 14-9).

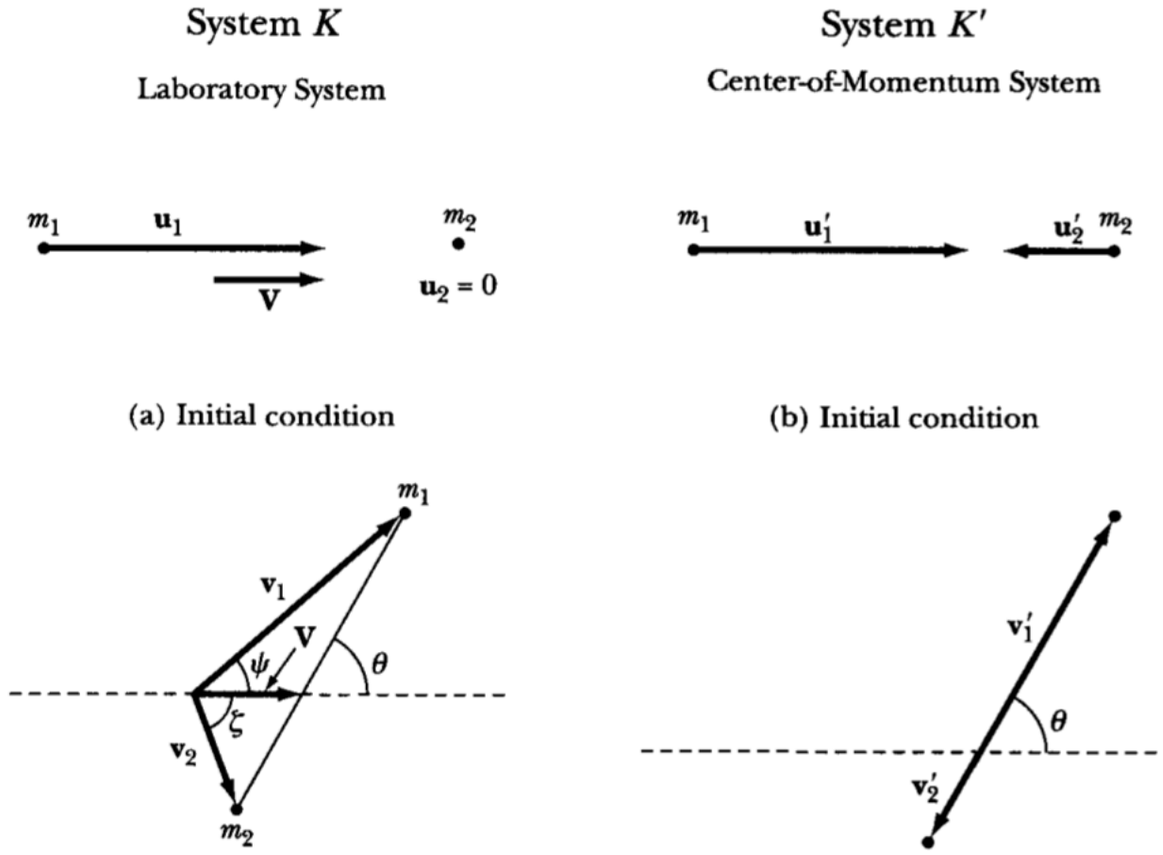


FIGURE 14-9 The elastic collision schematic of Figure 9-10 is redisplayed with systems K and K' indicated.

A simple Lorentz transformation then connects the two systems. To derive the relativistic kinematic expressions, the procedure is to obtain the center-of-momentum relations and then perform a Lorentz transformation back to the laboratory system. We choose the coordinate axes so that m_1 moves along the x-axis in K with speed u_1. Because m_2 is initially at rest in K, $u_2 = 0$. In K', m_2 moves with speed u'_2 and so K' moves with respect to K also with speed u'_2 and in the same direction as the initial motion of m_1.

Using the fact that $\beta\gamma = \sqrt{\gamma^2 - 1}$, we have

$$\begin{aligned} p_1' &= m_1 u_1' \gamma_1' = m_1 c \beta_1' \gamma_1' \\ &= m_1 c \sqrt{\gamma_1'^2 - 1} = m_2 c \sqrt{\gamma_2'^2 - 1} \\ &= p_2' \end{aligned} \tag{14.118}$$

which expresses the equality of the momenta in the center-of-momentum system.

According to Equation 14.92, the transformation of the momentum p_i (from K to K') is

$$p_1' = \left(p_1 - \frac{u_2'}{c^2} E_1\right)\gamma_2' \tag{14.119}$$

We also have

$$\left.\begin{aligned} p_1 &= m_1 u_1 \gamma_1 \\ E_1 &= m_1 c^2 \gamma_1 \end{aligned}\right\} \tag{14.120}$$

so Equation 14.118 can be used to obtain

$$\begin{aligned} m_1 c \sqrt{\gamma_1'^2 - 1} &= (m_1 c \beta_1 \gamma_1 - \beta_2' m_1 c \gamma_1)\gamma_2' \\ &= m_1 c \left(\gamma_2' \sqrt{\gamma_1^2 - 1} - \gamma_1 \sqrt{\gamma_2'^2 - 1}\right) \\ &= m_2 c \sqrt{\gamma_2'^2 - 1} \end{aligned} \tag{14.121}$$

These equations can be solved for γ_1' and γ_2' in terms of γ_1 :

$$\gamma_1' = \frac{\gamma_1 + \dfrac{m_1}{m_2}}{\sqrt{1 + 2\gamma_1\left(\dfrac{m_1}{m_2}\right) + \left(\dfrac{m_1}{m_2}\right)^2}} \tag{14.122a}$$

$$\gamma_2' = \frac{\gamma_1 + \dfrac{m_2}{m_1}}{\sqrt{1 + 2\gamma_1\left(\dfrac{m_2}{m_1}\right) + \left(\dfrac{m_2}{m_1}\right)^2}} \tag{14.122b}$$

Next, we write the equations of the transformation of the momentum components from K' back to K after the scattering. We now have both x- and y-components:

$$p_{1,x} = \left(p'_{1,x} + \frac{u'_2}{c^2} E'_1\right)\gamma'_2$$
$$= (m_1 c\beta'_1\gamma'_1 \cos\theta + m_1 c\beta'_2\gamma'_1)\gamma'_2$$
$$= m_1 c\gamma'_1\gamma'_2(\beta'_1 \cos\theta + \beta'_2) \qquad (14.123a)$$

(Note that, because the transformation is from K' to K, a plus sign occurs before the second term, in contrast to Equation 14.119.) Also,

$$p_{1,y} = m_1 c\beta'_1\gamma'_1 \sin\theta \qquad (14.123b)$$

The tangent of the laboratory scattering angle is given by $p_{i,y}/p_{i,x}$, therefore, dividing Equation 14.123b by Equation 14.123a, we obtain

$$\tan\psi = \frac{1}{\gamma'_2} \frac{\psi \sin\theta}{\cos\theta + (\beta'_2/\beta'_1)}$$

Using Equation 14.117 to express β'_2/β'_1, the result is

$$\tan\psi = \frac{1}{\gamma'_2} \frac{\sin\theta}{\cos\theta + (m_1\gamma'_1/m_2\gamma'_2)} \qquad (14.124)$$

For the recoil particle, we have

$$p_{2,x} = \left(p'_{2,x} + \frac{u'_2}{c^2} E'_2\right)\gamma'_2$$
$$= (-m_2 c\beta'_2\gamma'_2 \cos\theta + m_2 c\beta'_2\gamma'_2)\gamma'_2$$
$$= m_2 c\beta'_2\gamma'^2_2(1 - \cos\theta) \qquad (14.125a)$$

where a minus sign occurs in the first term because $p'_{2,x}$ is directed opposite to $p_{1,x}$. Also,

$$p_{2,y} = -m_2 c \beta_2' \gamma_2' \sin\theta \tag{14.125b}$$

As before, the tangent of the laboratory recoil angle ζ is given by $p_{2,y}/p_{2,x}$:

$$\tan\zeta = -\frac{1}{\gamma_2'} \frac{\sin\theta}{1 - \cos\theta} \tag{14.126}$$

The overall minus sign indicates that if m_1 is scattered toward positive values of ψ , then m_2 recoils in the negative ζ-direction.

A case of special interest is that in which $m_1 = m_2$. From Equations 14.122, we find

$$\gamma_1' = \gamma_2' = \sqrt{\frac{1 + \gamma_1}{2}}, \qquad m_1 = m_2 \tag{14.127}$$

The tangents of the scattering angles become

$$\tan\psi = \sqrt{\frac{2}{1 + \gamma_1}} \cdot \frac{\sin\theta}{1 + \cos\theta} \tag{14.128}$$

$$\tan\zeta = -\sqrt{\frac{2}{1 + \gamma_1}} \cdot \frac{\sin\theta}{1 - \cos\theta} \tag{14.129}$$

The product is therefore

$$\tan\psi \tan\zeta = -\frac{2}{1 + \gamma_1}, \qquad m_1 = m_2 \tag{14.130}$$

(The minus sign is of no essential importance; it only indicates that ψ and ζ are measured in opposite directions.)

We previously found that in the nonrelativistic limit there was always a right angle between the final velocity vectors in the scattering of particles of equal mass. Indeed, in the limit $\gamma_1 \to 1$, Equations 14.128 and 14.129 become equal to Equations 9.69 and 9.73, respectively, and so $\psi + \zeta = \pi/2$. Equation 14.130, however, shows that in the relativistic case $\psi + \zeta < \pi/2$; thus, the included angle in the scattering is always smaller than in the nonrelativistic limit. For equal scattering and recoil angles ($\psi = \zeta$), Equation 14.130 becomes

$$\tan \psi = \left(\frac{2}{1 + \gamma_1}\right)^{1/2}, \qquad m_1 = m_2$$

and the included angle between the directions of the scattered and recoil particles is

$$\begin{aligned} \phi &= \psi + \zeta = 2\psi \\ &= 2 \tan^{-1}\left(\frac{2}{1 + \gamma_1}\right)^{1/2}, \qquad m_1 = m_2 \end{aligned} \tag{14.131}$$

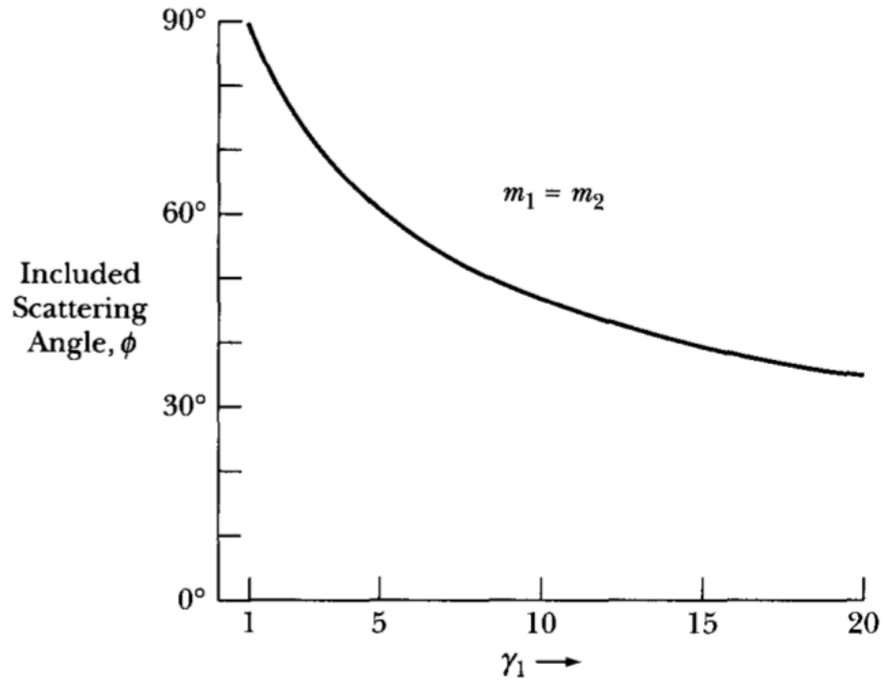


FIGURE 14-10 The included scattering angle, $\phi = \psi + \zeta$, is shown as a function of the relativistic parameter γ_1 for $m_1 = m_2$. For nonrelativistic scattering ($\gamma_1 = 1$), this angle is always 90°.

Figure 14-10 shows ϕ as a function of γ_1 up to $\gamma_1 = 20$. At $\gamma_1 = 10$, the included angle is approximately 46°. This value of γ_1 corresponds to an initial velocity that is 99.5% of the velocity of light. According to Equation 14.58, the kinetic energy is given by $T_1 = m_1c^2(\gamma_1 - 1)$; therefore, a proton with $\gamma_1 = 10$ would have a kinetic energy of approximately 8.4 GeV, whereas an electron with the same velocity would have $T_1 = 4.6$ MeV.

By using the transformation properties of the fourth component of the momentum four-vector (i.e., the total energy), it is possible to obtain the relativistic analogs of all the energy equations we have previously derived in the nonrelativistic limit.

APPENDIX A

Taylor's Theorem

A theorem of considerable importance in mathematical physics is Taylor's theorem, which relates to the expansion of an arbitrary function in a power series. In many instances, it is necessary to use this theorem to simplify' a problem to a tractable form.

Consider a function f(x) with continuous derivatives of all orders within a certain interval of the independent variable x. If this interval includes $x_0 \leq x \leq x_0 + h$, we may write

$$I \equiv \int_{x_0}^{x_0+h} f'(x)\,dx = f(x_0 + h) - f(x_0) \tag{A.1}$$

where f'(x) is the derivative of f(x) with respect to x. If we make the change of variable

$$x = x_0 + h - t \tag{A.2}$$

we have

$$I = \int_0^h f'(x_0 + h - t)\,dt \tag{A.3}$$

Integrating by parts

$$\begin{aligned} I &= tf'(x_0 + h - t)\Big|_0^h + \int_0^h tf''(x_0 + h - t)\,dt \\ &= hf'(x_0) + \int_0^h tf''(x_0 + h - t)\,dt \end{aligned} \tag{A.4}$$

Integrating the second term by parts, we find

$$I = hf'(x_0) + \frac{h^2}{2!}f''(x_0) + \int_0^h \frac{t^2}{2!}f'''(x_0 + h - t)\,dt \tag{A.5}$$

Continuing this process, we generate an infinite series for I. From the definition of I, we then have

$$f(x_0 + h) = f(x_0) + hf'(x_0) + \frac{h^2}{2!}f''(x_0) + \cdots \tag{A.6}$$

This is the Taylor series expansion of the function $f(x_0 + h)$. A more common form of the series results if we set $x_0 = 0$ and $h = x$ [i.e., the function f(x) is expanded about the origin]:

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \frac{x^3}{3!}f'''(0) + \cdots + \frac{x^n}{n!}f^{(n)}(0) + \cdots \tag{A.7}$$

where

$$f^{(n)}(0) \equiv \left.\frac{d^n}{dx^n}f(x)\right|_{x=0} \tag{A.8}$$

Equation A.7 is usually called the **Maclaurin's series** for the function f(x).

The series expansions given in Equations A.6 and A.7 possess two important properties. Under very general conditions, they may be differentiated or integrated term by term, and the resulting series converge to the derivative or integral of the original function.

EXAMPLE A.1

Find the Taylor series expansion of e^x.

Solution. Because the derivative of exp(x) of any order is just exp (x), the exponential series is

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \tag{A.9}$$

This result is of considerable importance and will be used often.

EXAMPLE A.2

Find the Taylor series expansion of sin x.

Solution. To expand f(x) = sin x, we need

$$\begin{aligned} f(x) &= \sin x, & f(0) &= 0 \\ f'(x) &= \cos x, & f'(0) &= 1 \\ f''(x) &= -\sin x, & f''(0) &= 0 \\ f'''(x) &= -\cos x, & f'''(0) &= -1 \end{aligned}$$

Therefore,

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \tag{A.10}$$

Similarly,

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \tag{A.11}$$

EXAMPLE A.3

Use the Taylor series expansion of $(1+t)^{-1}$ to integrate

$$\int_0^x \frac{dt}{1+t}$$

Solution. A series expansion can often be profitably used in the evaluation of a definite integral. (This is particularly true for those cases in which the indefinite integral cannot be found in closed form.)

$$\int_0^x \frac{dt}{1+t} = \int_0^x (1 - t^2 + t^3 - \cdots)\,dt, \quad |t| < 1 \tag{A.12}$$

Because

$$\frac{d}{dx}\ln(1+x)=\frac{1}{1+x} \tag{A.13}$$

We also have the result

$$\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\cdots \tag{A.14}$$

EXAMPLE A.4

Taylor's series can be used to restructure a function as well as to approximate it. For some applications, such a restructuring may be more useful to work with. We may, for example, want to expand the polynomial $f(x) = 4 + 6x + 3x^2 + 2x^3 + x^4$ about $x = 2$ rather than $x = 0$.

Solution. First, we compute the various derivatives and evaluate them at $x = 2$:

$$\begin{aligned}
f(2) &= 60\\
f'(2) &= (6+6x+6x^2+4x^3)\big|_{x=2} = 74\\
f''(2) &= (6+12x+12x^2)\big|_{x=2} = 78\\
f'''(2) &= (12+24x)\big|_{x=2} = 60\\
f^{iv}(2) &= 24\\
f^{v}(2) &= 0
\end{aligned}$$

Using Equation A.6 with h= (x - 2)

$$f(x)=60+74(x-2)+39(x-2)^2+10(x-2)^3+(x-2)^4 \tag{A.15}$$

EXAMPLE A.5

There are a great many important integrals arising in physics that cannot be integrated in closed form, that is, in terms of elementary functions (polynomials, exponentials, logarithms, trigonometric functions, and their inverses). Integrals with integrands

$$e^{-x^2}, \quad \frac{e^{-x}}{x}, \quad x \tan x, \quad \sin x^2, \quad 1/\ln x, \quad (\sin x)/x, \quad \text{or} \quad 1/\sqrt{1-x^3}$$

are a few such examples. Nevertheless, the values of the integrals or good approximations of their values are needed. A Taylor series expansion of all or part of the integrand followed by a term-by-term integration of the resulting series produces an answer as precise as is wished. As an example, solve the following integral:

$$\int_1^x \frac{e^t}{t} dt \tag{A.16}$$

Solution. Using Equation A.9,

$$\int_1^x \frac{e^t}{t} dt = \int_1^x \frac{\left(1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots\right) dt}{t} \tag{A.17}$$

$$= \int_1^x \frac{dt}{t} + \int_1^x dt + \int_1^x \frac{t}{2!} dt + \int_1^x \frac{t^2}{3!} dt + \cdots$$

$$= \ln x - (x-1) + \frac{1}{4}(x^2 - 1) + \frac{1}{18}(x^3 - 1) + \cdots \tag{A.18}$$

APPENDIX B

Elliptic Integrals

There is a large and important class of integrals called elliptic integrals that cannot be evaluated in closed form in terms of elementary functions. Elliptic integrals occur in many physical situations; for example, see the exact solution to the plane pendulum in Section 4.4. Any integral of the form

$$\int (a \sin\theta + b \cos\theta + c)^{\pm 1/2}\, d\theta, \quad \text{or} \quad \int R(x, \sqrt{y})\, dx \tag{B.1}$$

where R is a rational function, $y = ax^4 + bx^3 + cx^2 + dx + e$, with distinct linear factors and a, b, c, d, and e constants (with not both a, b zero) is an elliptic integral. It is customary, however, to transform all elliptic integrals into one or more of three standard forms. These standard forms have been much studied and tabulated. Several handbooks are available with tables of values for them

B.l Elliptic Integrals of the First Kind

$$F(k,\phi) = \int_0^{\phi} \frac{d\theta}{\sqrt{1 - k^2 \sin^2\theta}}, \quad k^2 < 1 \tag{B.2a}$$

or if $z = \sin\theta$

$$\bar{F}(k, x) = \int_0^{x} \frac{dz}{\sqrt{(1 - z^2)(1 - k^2 z^2)}}, \quad k^2 < 1 \tag{B.2b}$$

B.2 Elliptic Integrals of the Second Kind

$$E(k, \phi) = \int_0^{\phi} \sqrt{1 - k^2 \sin^2\theta}\, d\theta, \quad k^2 < 1 \tag{B.3a}$$

or if z = sin θ

$$\bar{E}(k, x) = \int_0^x \sqrt{\frac{1 - k^2 z^2}{1 - z^2}}\, dz, \quad k^2 < 1 \tag{B.3b}$$

B.3 Elliptic Integrals of the Third Kind

$$\Pi(n, k, \phi) = \int_0^{\phi} \frac{d\theta}{(1 + n \sin^2 \theta)\sqrt{1 - k^2 \sin^2 \theta}} \tag{B.4a}$$

or if z = sin θ

$$\bar{\Pi}(n, k, x) = \int_0^x \frac{dz}{(1 + nz^2)\sqrt{(1 - z^2)(1 - k^2 z^2)}} \tag{B.4b}$$

These standard forms obey the following identities, which are often helpful:

$$\left.\begin{aligned} F(k, \phi) &= F(k, \pi) - F(k, \pi - \phi) \\ E(k, \phi) &= E(k, \pi) - E(k, \pi - \phi) \end{aligned}\right\} \tag{B.5}$$

and

$$\left.\begin{aligned} F(k, m\pi + \phi) &= mF(k, \pi) + F(k, \phi) \\ E(k, m\pi + \phi) &= mE(k, \pi) + E(k, \phi) \end{aligned}\right\} \tag{B.6}$$

where m is an integer.

If tables are not handy or if ϕ or x is needed as a variable, the standard integrals may be approximated by expanding the integrand in an infinite series and integrating term by term. For example, consider

$$E(k, \phi) = \int_0^{\phi} \sqrt{1 - k^2 \sin^2 \theta}\, d\theta$$

Using the binomial theorem on the integrand

$$(1 - k^2 \sin^2 \theta)^{1/2} = 1 - \frac{1}{2} k^2 \sin^2 \theta - \frac{1}{8} k^4 \sin^4 \theta - \cdots$$

so

$$E(k, \phi) = \int_0^{\phi} \left[1 - \frac{1}{2} k^2 \sin^2 \theta - \frac{1}{8} k^4 \sin^4 \theta - \cdots \right.$$

$$\left. - \frac{1 \cdot 3 \cdot 5 \cdots (2n-3)}{2 \cdot 4 \cdot 6 \cdots (2n)} k^{2n} \sin^{2n} \theta - \cdots \right] d\theta$$

$$= \phi - \frac{k^2}{2} \int_0^{\phi} \sin^2 \theta d\theta - \cdots - \frac{1 \cdot 3 \cdot 5 \cdots (2n-3)}{2 \cdot 4 \cdot 6 \cdots (2n)} k^{2n}$$

$$\times \int_0^{\phi} \sin^{2n} \theta\, d\theta - \cdots \tag{B.7}$$

Similarly, the binomial theorem can be used to expand $(1 - k^2 \sin^2 \theta)^{-1/2}$ to yield

$$F(k, \phi) = \phi + \frac{1}{2} k^2 \int_0^{\phi} \sin^2 \theta d\theta + \frac{3}{8} k^4 \int_0^{\phi} \sin^4 \theta d\theta + \cdots$$

$$+ \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots (2n)} k^{2n} \int_0^{\phi} \sin^{2n} \theta d\theta + \cdots \tag{B.8}$$

EXAMPLE B.1

Put the integral $\int_{\phi_1}^{\phi_2} \sqrt{1 - k^2 \sin^2 \theta} d\theta$ into standard form.

Solution. Recall from calculus that for any integral $\int_a^b f(x)dx$ it is possible to write

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

so

$$\int_{\phi_1}^{\phi_2} \sqrt{1 - k \sin^2 \theta}\; d\theta = \int_{\phi_1}^{0} \sqrt{1 - k^2 \sin^2 \theta}\; d\theta + \int_{0}^{\phi_2} \sqrt{1 - k^2 \sin^2 \theta}\; d\theta$$

But there is another property of integrals:

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$$

so

$$\int_{\phi_1}^{\phi_2} \sqrt{1 - k^2 \sin^2 \theta}\, d\theta = \int_0^{\phi_2} \sqrt{1 - k^2 \sin^2 \theta}\, d\theta - \int_0^{\phi_1} \sqrt{1 - k^2 \sin^2 \theta}\, d\theta$$

or

$$\int_{\phi_1}^{\phi_2} \sqrt{1 - k^2 \sin^2 \theta}\; d\theta = E(k, \phi_2) - E(k, \phi_1) \tag{B.9}$$

The terms on the right can be looked up in a handbook.

EXAMPLE B.2

Transform the elliptic integral

$$\int_0^{\phi} \frac{d\theta}{\sqrt{1 - n^2 \sin^2 \theta}}, \quad \text{where } n^2 > 1$$

into a standard form.

Solution, To reduce this integral to standard form, the radical must be transformed to $\sqrt{1 - k^2 \sin^2 \theta}$, with $k^2 < 1$. To do this, consider the transformation $n \sin \theta = \sin \beta$. Differentiating, we have

$$n \cos \theta d\theta = \cos \beta \, d\beta$$

so

$$d\theta = \frac{\cos \beta \, d\beta}{n \cos \theta}$$

Using the identity $\sin^2 \theta + \cos^2 \theta = 1$ leads to

$$\cos \theta = \sqrt{1 - \sin^2 \theta} = \sqrt{1 - \left(\frac{\sin \beta}{n}\right)^2}$$

Also, $\cos \beta = \sqrt{1 - \sin^2 \beta}$, and $\sqrt{1 - n^2 \sin^2 \theta} = \sqrt{1 - \sin^2 \beta}$. Hence the integral becomes

$$\int_0^{\phi} \frac{d\theta}{\sqrt{1 - n^2 \sin^2 \theta}} = \int_0^{\sin^{-1}(n \sin \phi)} \frac{\sqrt{1 - \sin^2 \beta}}{n \sqrt{1 - \left(\frac{\sin \beta}{n}\right)^2} \sqrt{1 - \sin^2 \beta}} d\beta$$

$$= \frac{1}{n} \int_0^{\sin^{-1}(n \sin \phi)} \frac{d\beta}{\sqrt{1 - \left(\frac{1}{n^2}\right) \sin^2 \beta}}$$

so

$$\int_0^{\phi} \frac{d\theta}{\sqrt{1 - n^2 \sin^2 \theta}} = \frac{1}{n} \int_0^{\sin^{-1}(n \sin \phi)} \frac{d\beta}{\sqrt{1 - \left(\frac{1}{n^2}\right) \sin^2 \beta}} \tag{B.10}$$

where $l/n^2 < 1$. The integral on the right is now in standard form.

EXAMPLE B.3

Transform the elliptic integral

$$\int_0^{\phi} \frac{d\theta}{\sqrt{\cos 2\theta}}$$

into a standard form.

Solution. Let $\mu = \sin\theta$; then $d\mu = \cos\theta d\theta$.

Because $\sin^2\theta + \cos^2\theta = 1$, $\cos\theta = \sqrt{1 - \sin^2\theta} = \sqrt{1 - \mu^2}$, so $d\theta = d\mu/\sqrt{1 - \mu^2}$. By another trigonometric identity, $\cos 2\theta = 1 - 2\sin^2\theta = 1 - 2\mu^2$. Thus $\sqrt{\cos 2\theta} = \sqrt{1 - 2\mu^2}$, and

$$\int_0^{\phi} \frac{d\theta}{\sqrt{\cos 2\theta}} = \int_0^{\sin\phi} \frac{d\mu}{\sqrt{1 - \mu^2}\sqrt{1 - 2\mu^2}}$$

Let $z = \sqrt{2}\mu$; then $dz = \sqrt{2}d\mu$, so

$$\int_0^{\phi} \frac{d\theta}{\sqrt{\cos 2\theta}} = \frac{1}{\sqrt{2}} \int_0^{\sqrt{2}\sin\phi} \frac{dz}{\sqrt{(1 - z^2)\left(1 - \frac{1}{2}z^2\right)}} \tag{B.11}$$

The integral on the right is in standard form.

APPENDIX C

Ordinary Differential Equations of Second Order

C.1 Linear Homogeneous Equations

By far, the most important type of ordinary differential equation encountered in problems in mathematical physics is the second-order linear equation with constant coefficients. Equations of this type have the form

$$\frac{d^2y}{dx^2} + a\frac{dy}{dx} + by = f(x) \tag{C.1a}$$

or, denoting derivatives by primes,

$$y'' + ay' + by = f(x) \tag{C.1b}$$

A particularly important class of such equations are those for which f(x) = 0. These equations (called **homogeneous equations**) are important not only in themselves but also as *reduced* equations in the solution of the more general type of equation (Equation C.1).

We consider the linear homogeneous second-order equation with constant coefficients first.

$$y'' + ay' + by = 0 \tag{C.2}$$

These equations have the following important properties:

a. If $y_1(x)$ is a solution of Equation C.2, then $c_1y_1(x)$ is also a solution.
b. If $y_1(x)$ and $y_2(x)$ are solutions, then $y_1(x) + y_2(x)$ is also a solution (principle of *superposition*).
c. If $y_1(x)$ and $y_2(x)$ are *linearly* independent solutions, then the *general* solution to the equation is given by $c_1y_1(x) + c_2y_2(x)$. (The general solution always contains two arbitrary constants.)

The functions $y_1(x)$ and $y_2(x)$ are **linearly independent** if and only if the equation

$$\lambda y_1(x) + \mu y_2(x) \equiv 0 \tag{C.3}$$

is satisfied only by $\lambda = \mu = 0$. If Equation C.3 can be satisfied with λ and μ different from zero, then $y_1(x)$ and $y_2(x)$ are said to be **linearly dependent**.

The general condition (i.e., the necessary and sufficient condition) that a set of functions $y_1, y_2, y_3, \ldots$ be linearly dependent is that the Wronskian determinant of these functions vanish identically:

$$W = \begin{vmatrix} y_1 & y_2 & y_3 & \cdots & y_n \\ y_1' & y_2' & y_3' & \cdots & y_n' \\ y_1'' & y_2'' & y_3'' & \cdots & y_n'' \\ \vdots & & & & \\ y_1^{(n-1)} & y_2^{(n-1)} & y_3^{(n-1)} & \cdots & y_n^{(n-1)} \end{vmatrix} = 0 \tag{C.4}$$

where $y^{(n)}$ is the n^{th} derivative of y with respect to x.

The properties (a) and (b) above can be verified by direction substitution, but (c) is only asserted here to yield the general solution. These properties apply only to the homogeneous equation (Equation C.2) and not to the general equation (Equation C.l).

Equations of the type C.2 are reducible through the substitution

$$y = e^{rx} \tag{C.5}$$

Now

$$y' = re^{rx}, \quad y^n = r^2 e^{rx} \tag{C.6}$$

Using these expressions for y' and y" in Equation C.2, we find an algebraic equation called the **auxiliary equation**:

$$r^2 + ar + b = 0 \tag{C.7}$$

The solution of this quadratic in r is

$$r = -\frac{a}{2} \pm \frac{1}{2}\sqrt{a^2 - 4b} \tag{C.8}$$

We first assume that the two roots, denoted by r_1 and r_2, are not identical and write the solution as

$$y = e^{r_1 x} + e^{r_2 x} \tag{C.9}$$

Because the Wronskian determinant of $\exp(r_1 x)$ and $\exp(r_2 x)$ does not vanish, these functions are linearly independent. Thus, the general solution is

$$\boxed{y = c_1 e^{r_1 x} + c_2 e^{r_2 x}, \quad r_1 \neq r_2} \tag{C.10}$$

If it happens that $r_1 = r_2 = r$, then it can be verified by direct substitution that $x \exp(rx)$ is also a solution, and because $\exp(rx)$ and $x \exp(rx)$ are linearly independent, the general solution for identical roots is given by

$$\boxed{y = c_1 e^{rx} + c_2 x e^{rx}, \quad r_1 = r_2 \equiv r} \tag{C.11}$$

EXAMPLE C.1

Solve the equation

$$y'' - 2y' - 3y = 0 \tag{C.12}$$

Solution: The auxiliary equation is

$$r^2 - 2r - 3 = (r - 3)(r + 1) = 0 \tag{C.13}$$

The roots are

$$r_1 = 3, \quad r_2 = -1 \tag{C.14}$$

The general solution is therefore

$$y = c_1 e^{3x} + c_2 e^{-x} \tag{C.15}$$

EXAMPLE C.2

Solve the equation

$$y'' + 4y' + 4y = 0 \tag{C.16}$$

Solution, The auxiliary equation is

$$r^2 + 4r + 4 = (r + 2)^2 = 0 \tag{C.17}$$

The roots are equal, are r = -2. The general solution is therefore

$$y = c_1 e^{-2x} + c_2 x e^{-2x} \tag{C.18}$$

If the roots r_1 and r_2 of the auxiliary equation are imaginary, the solutions given by $c_1 \exp(r_1 x)$ and $c_2 \exp(r_2 x)$ are still correct.

To give the solutions entirely in terms of real quantities, we use the Euler relations to express the exponentials. Then,

$$\left.\begin{aligned} e^{r_1 x} &= e^{\alpha x} e^{i\beta x} &&= e^{\alpha x}(\cos \beta x + i \sin \beta x) \\ e^{r_2 x} &= e^{\alpha x} e^{-i\beta x} &&= e^{\alpha x}(\cos \beta x - i \sin \beta x) \end{aligned}\right\} \tag{C.19}$$

and the general solution is

$$\begin{aligned} y &= c_1 e^{r_1 x} + c_2 e^{r_2 x} \\ &= e^{\alpha x}[(c_1 + c_2)\cos \beta x + i(c_1 - c_2)\sin \beta x] \end{aligned} \tag{C.20}$$

Now Cj and ^2 are arbitrary, but these constants may be complex. However, not all four elements can be independent (because there would be four arbitrary constants rather than two). The number of independent elements can be reduced to the required two by making Cj and ^2 complex conjugates. Then the combinations A=Ci + C2 and B = ((ci — C2) become a pair of arbitrary, real constants. Using these quantities in the solution, we have

$$y = e^{\alpha x}(A \cos \beta x + B \sin \beta x) \tag{C.21}$$

Equation C.21 may be put into a form that is sometimes more convenient b multiplying and dividing by / i, = V A^ + B^:

$$y = \mu e^{\alpha x}[(A/\mu)\cos\beta x + (B/\mu)\sin\beta x] \tag{C.22}$$

Next, we define an angle 5 (see Figure C-1) such that

$$\sin\delta = A/\mu, \quad \cos\delta = B/\mu, \quad \tan\delta = A/B \tag{C.23}$$

Then, the solution becomes

$$\begin{aligned} y &= \mu e^{\alpha x}(\sin\delta\cos\beta x + \cos\delta\sin\beta x) \\ &= \mu e^{\alpha x}\sin(\beta x + \delta) \end{aligned}$$

Depending on the exact definition of the phase 5, we may write the solution alternatively as

$$\boxed{y = \mu e^{\alpha x}\sin(\beta x + \delta)} \tag{C.24a}$$

$$\boxed{y = \mu e^{\alpha x}\cos(\beta x + \delta)} \tag{C.24b}$$

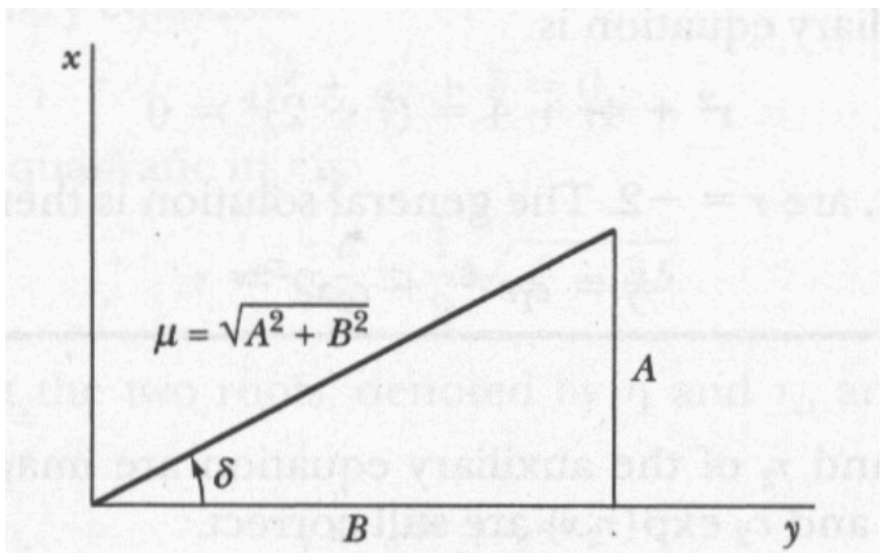


FIGURE C-1

EXAMPLE C.3

Solve the equation

$$y'' + 2y' + 4y = 0 \tag{C.25}$$

Solution. The auxiliary equation is

$$r^2 + 2r + 4 = 0 \tag{C.26}$$

with

$$r = \frac{-2 \pm \sqrt{4 - 16}}{2} = -1 \pm i\sqrt{3} \tag{C.27}$$

Hence,

$$\alpha = -1, \quad \beta = \sqrt{3} \tag{C.28}$$

and the general solution is

$$y = e^{-x}(c_1 \cos \sqrt{3}x + c_2 \sin \sqrt{3}x) \tag{C.29}$$

or

$$y = \mu e^{-x} \sin[(\sqrt{3}x + \delta)] \tag{C.30}$$

Summarizing, then, there are three possible types of general solutions to homogeneous second-order linear differential equations, as indicated in Table C-1.

TABLE C-1

Roots of the auxiliary equations	General solution
Real, unequal ($r_1 \neq r_2$)	$c_1 e^{r_1 x} + c_2 e^{r_2 x}$
Real, equal ($r_1 = r_2 \equiv r$)	$c_1 e^{rx} + c_2 x e^{rx}$
Imaginary ($\alpha \pm i\beta$)	$e^{\alpha x}(c_1 \cos \beta x + c_2 \sin \beta x)$ or $\mu e^{\alpha x} \sin(\beta x + \delta)$

C.2 Linear Inhomogeneous Equations

To solve the general (i.e., inhomogeneous) second-order linear differential equation, consider the following. Let y = u be the *general solution* of

$$y'' + ay' + by = 0 \tag{C.31}$$

and let y = v be any solution of

$$y'' + ay' + by = f(x) \tag{C.32}$$

Then, y = u + v is a solution of Equation C.32, because

$$\begin{aligned} y'' + ay' + by &= (u'' + au' + bu) + (v'' + av' + bv) \\ &= 0 + f(x) \end{aligned}$$

Because u contains the two arbitrary constants c_1 and c_2, the combinations u + v satisfies all the requirements of the general solution to Equation C.32. The function u is the **complementary function** and v is the **particular integral** of the equation. Because a general method of finding u has been given above, it only remains to find, by inspection or by trial, some function v that satisfies

$$v'' + av' + bv = f(x) \tag{C.33}$$

EXAMPLE C.4

Solve the equation

$$y'' + 5y' + 6y = x^2 + 2x \tag{C.34}$$

Solution, The auxiliary equation is

$$r^2 + 5r + 6 = (r + 3)(r + 2) = 0 \tag{C.35}$$

$$r_1 = -3, \quad r_2 = -2 \tag{C.36}$$

so the complementary function is

$$u = c_1 e^{-3x} + c_2 e^{-2x} \tag{C.37}$$

Because the right-hand side of the original equation is a second-degree polynomial, we guess a particular integral of the form

$$v = Ax^2 + Bx + C \tag{C.38}$$

Then,

$$v' = 2Ax + B \tag{C.39}$$

$$v'' = 2A \tag{C.40}$$

Substituting into the differential equation, we have

$$2A + 5(2Ax + B) + 6(Ax^2 + Bx + C) = x^2 + 2x \tag{C.41}$$

or

$$(6A)x^2 + (10A + 6B)x + (2A + 5B + 6C) = x^2 + 2x \tag{C.42}$$

Equating coefficients of like powers of x:

$$\left.\begin{aligned} 6A &= 1 \\ 10A + 6B &= 2 \\ 2A + 5B + 6C &= 0 \end{aligned}\right\} \tag{C.43}$$

Solving,

$$A = \frac{1}{6}, \quad B = \frac{1}{18}, \quad C = -\frac{11}{108} \tag{C.44}$$

Hence,

$$\begin{aligned} v &= \frac{1}{6}x^2 + \frac{1}{18}x - \frac{11}{108} \\ &= \frac{18x^2 + 6x - 11}{108} \end{aligned} \tag{C.45}$$

The general solution is therefore

$$y = u + v = c_1 e^{-3x} + c_2 e^{-2x} + \frac{18x^2 + 6x - 11}{108} \tag{C.46}$$

The type of solution illustrated in this example is called the method of undetermined coefficients.

EXAMPLE C.5

Solve the equation

$$y'' + 4y = 3x \cos x \tag{C.47}$$

Solution. The auxiliary equation is

$$r^2 + 4 = (r + 2i)(r - 2i) = 0 \tag{C.48}$$

with roots

$$\left.\begin{aligned} r_1 &= \alpha + i\beta = 0 + 2i \\ r_2 &= \alpha - i\beta = 0 - 2i \end{aligned}\right\} \tag{C.49}$$

so

$$\alpha = 0, \quad \beta = 2 \tag{C.50}$$

and the complementary function is

$$\begin{aligned} u &= e^{\alpha x}(c_1 \cos \beta x + c_2 \sin \beta x) \\ &= c_1 \cos 2x + c_2 \sin 2x \end{aligned} \tag{C.51}$$

To find a particular integral, we note that from x cos x and its derivatives it is possible to generate only terms involving the following functions:

$$x \cos x, \quad x \sin x, \quad \cos x, \quad \sin x$$

Therefore, because these functions are linearly independent, the trial particular integral is

$$v = Ax \cos x + Bx \sin x + C \cos x + D \sin x \tag{C.52}$$

$$v' = A(\cos x - x \sin x) + B(\sin x + x \cos x) - C \sin x + D \cos x \tag{C.53}$$

$$v'' = -A(2 \sin x + x \cos x) + B(2 \cos x - x \sin x) - C \cos x - D \sin x \tag{C.54}$$

Substituting into the original differential equation,

$$(3D - 2A)\sin x + (2B + 3C)\cos x + 3(A - 1)x\cos x + (3B)x\sin x = 0 \tag{C.55}$$

The coefficient of each term must vanish (because of the linear independence of the terms):

$$3D = 2A, \quad 2B = -3C, \quad A = 1, \quad 3B = 0 \tag{C.56}$$

from which

$$A = 1, \quad B = 0, \quad C = 0, \quad D = \frac{2}{3} \tag{C.57}$$

The general solution is therefore

$$y = c_1 \sin 2x + c_2 \cos 2x + x\cos x + \frac{2}{3}\sin x \tag{C.58}$$

If the right-hand side, f(x), of the general equation (Equation C.l or C.32) is such that f(x) and its first two derivatives (only second-order equations are being considered) contain only linearly independent functions, then a linear combination of these functions constitutes the trial particular integral. In the event that the trial function contains a term that already appears in the complementary function, use the term multiplied by x; if this combination also appears in the complementary function, use the term multiplied by x^2. No higher powers are needed because only second-order equations are being considered and only exp(rx) or x exp(rx) occur as solutions to the reduced equation; (x^2) exp(rx) never occurs.

APPENDIX D
Useful Formulas

D.1 Binomial Expansion

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3$$
$$+ \cdots + \binom{n}{r}x^r + \cdots, \qquad |x| < 1 \tag{D.1}$$

$$(1-x)^n = 1 - nx + \frac{n(n-1)}{2!}x^2 - \frac{n(n-1)(n-2)}{3!}x^3$$
$$+ \cdots + (-1)^r\binom{n}{r}x^r + \cdots, \qquad |x| < 1 \tag{D.2}$$

where the **binomial coefficient** is

$$\binom{n}{r} \equiv \frac{n!}{(n-r)!r!} \tag{D.3}$$

Some particularly useful cases of the above are

$$(1 \pm x)^{1/2} = 1 \pm \frac{1}{2}x - \frac{1}{8}x^2 \pm \frac{1}{16}x^3 - \cdots \tag{D.4}$$

$$(1 \pm x)^{1/3} = 1 \pm \frac{1}{3}x - \frac{1}{9}x^2 \pm \frac{5}{81}x^3 - \cdots \tag{D.5}$$

$$(1 \pm x)^{-1/2} = 1 \mp \frac{1}{2}x + \frac{3}{8}x^2 \mp \frac{5}{16}x^3 + \cdots \tag{D.6}$$

$$(1 \pm x)^{-1/3} = 1 \mp \frac{1}{3}x + \frac{2}{9}x^2 \mp \frac{14}{81}x^3 + \cdots \tag{D.7}$$

$$(1 \pm x)^{-1} = 1 \mp x + x^2 \mp x^3 + \cdots \tag{D.8}$$

$$(1 \pm x)^{-2} = 1 \mp 2x + 3x^2 \mp 4x^3 + \cdots \tag{D.9}$$

$$(1 \pm x)^{-3} = 1 \mp 3x + 6x^2 \mp 10x^3 + \cdots \tag{D.10}$$

For convergence of *all* the above series, we must have $|x| < 1$.

D.2 Trigonometric Relations

$$\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B \tag{D.11}$$

$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B \tag{D.12}$$

$$\sin 2A = 2 \sin A \cos A = \frac{2 \tan A}{1 + \tan^2 A} \tag{D.13}$$

$$\cos 2A = 2 \cos^2 A - 1 \tag{D.14}$$

$$\sin^2 \frac{A}{2} = \frac{1}{2}(1 - \cos A) \tag{D.15}$$

$$\cos^2 \frac{A}{2} = \frac{1}{2}(1 + \cos A) \tag{D.16}$$

$$\sin^2 A = \frac{1}{2}(1 - \cos 2A) \tag{D.17}$$

$$\sin^3 A = \frac{1}{4}(3 \sin A - \sin 3A) \tag{D.18}$$

$$\sin^4 A = \frac{1}{8}(3 - 4 \cos 2A + \cos 4A) \tag{D.19}$$

$$\cos^2 A = \frac{1}{2}(1 + \cos 2A) \tag{D.20}$$

$$\cos^3 A = \frac{1}{4}(3\cos A + \cos 3A) \tag{D.21}$$

$$\cos^4 A = \frac{1}{8}(3 + 4\cos 2A + \cos 4A) \tag{D.22}$$

$$\tan(A + B) = \frac{\tan A + \tan B}{1 - \tan A \tan B} \tag{D.23}$$

$$\tan^2 \frac{A}{2} = \frac{1 - \cos A}{1 + \cos A} \tag{D.24}$$

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i} \tag{D.25}$$

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \tag{D.26}$$

$$e^{ix} = \cos x + i \sin x \tag{D.27}$$

D.3 Trigonometric Series

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \tag{D.28}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \tag{D.29}$$

$$\tan x = x + \frac{x^3}{3} + \frac{2}{15}x^5 + \cdots, \quad |x| < \pi/2 \tag{D.30}$$

$$\sin^{-1} x = x + \frac{x^3}{6} + \frac{3}{40}x^5 + \cdots, \quad \begin{cases} |x| < 1 \\ |\sin^{-1} x| < \pi/2 \end{cases} \tag{D.31}$$

$$\cos^{-1} x = \frac{\pi}{2} - x - \frac{x^3}{6} - \frac{3}{40}x^5 - \cdots, \quad \begin{cases} |x| < 1 \\ 0 < \cos^{-1} x < \pi \end{cases} \tag{D.32}$$

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots, \quad |x| < 1 \tag{D.33}$$

D.4 Exponential and Logarithmic Series

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{D.34}$$

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots, \quad |x| < 1, \quad x = 1 \tag{D.35}$$

$$\ln[\sqrt{(x^2/a^2) + 1} + (x/a)] = \sinh^{-1} x/a \tag{D.36}$$

$$= -\ln[\sqrt{(x^2/a^2) + 1} - (x/a)] \tag{D.37}$$

D.5 Complex Quantities

Cartesian form: $z = x + iy$, complex conjugate $z^* = x - iy$, $i = \sqrt{-1}$ **(D.38)**

Polar form:
$$z = |z| e^{i\theta} \quad \textbf{(D.39)}$$

$$z^* = |z| e^{-i\theta} \quad \textbf{(D.40)}$$

$$zz^* = |z|^2 = x^2 + y^2 \quad \textbf{(D.41)}$$

Real part of z:
$$\text{Re } z = \frac{1}{2}(z + z^*) = x \quad \textbf{(D.42)}$$

Imaginary part of z:
$$\text{Im } z = -\frac{1}{2}(z - z^*) = y \quad \textbf{(D.43)}$$

Euler's formula:
$$e^{i\theta} = \cos\theta + i \sin\theta \quad \textbf{(D.44)}$$

D.6 Hyperbolic Functions

$$\sinh x = \frac{e^x - e^{-x}}{2} \quad \textbf{(D.45)}$$

$$\cosh x = \frac{e^x + e^{-x}}{2} \quad \textbf{(D.46)}$$

$$\tanh x = \frac{e^{2x} - 1}{e^{2x} + 1} \quad \textbf{(D.47)}$$

$$\sin ix = i \sinh x \quad \textbf{(D.48)}$$

$$\cos ix = \cosh x \quad \textbf{(D.49)}$$

$$\sinh ix = i \sin x \tag{D.50}$$

$$\cosh ix = \cos x \tag{D.51}$$

$$\sinh^{-1} x = \tanh^{-1}\left(\frac{x}{\sqrt{x^2 + 1}}\right) \tag{D.52}$$

$$= \ln(x + \sqrt{x^2 + 1}) \tag{D.53}$$

$$= \cosh^{-1}(\sqrt{x^2 + 1}), \quad \begin{cases} > 0, & x > 0 \\ < 0, & x < 0 \end{cases} \tag{D.54}$$

$$\cosh^{-1} x = \pm \tanh^{-1}\left(\frac{\sqrt{x^2 - 1}}{x}\right), \quad x > 1 \tag{D.55}$$

$$= \pm \ln(x + \sqrt{x^2 - 1}), \quad x > 1 \tag{D.56}$$

$$\cosh^{-1} x = \pm \sinh^{-1}(\sqrt{x^2 - 1}), \quad x > 1 \tag{D.57}$$

$$\frac{d}{dy} \sinh y = \cosh y \tag{D.58}$$

$$\frac{d}{dy} \cosh y = \sinh y \tag{D.59}$$

$$\sinh(x_1 + x_2) = \sinh x_1 \cosh x_2 + \cosh x_1 \sinh x_2 \tag{D.60}$$

$$\cosh(x_1 + x_2) = \cosh x_1 \cosh x_2 + \sinh x_1 \sinh x_2 \tag{D.61}$$

$$\cosh^2 x - \sinh^2 x = 1 \tag{D.62}$$

APPENDIX E
Useful Integrals

E.1 Algebraic Functions

$$\int \frac{dx}{a^2 + x^2} = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right), \qquad \left|\tan^{-1}\left(\frac{x}{a}\right)\right| < \frac{\pi}{2} \tag{E.1}$$

$$\int \frac{x\,dx}{a^2 + x^2} = \frac{1}{2}\ln(a^2 + x^2) \tag{E.2}$$

$$\int \frac{dx}{x(a^2 + x^2)} = \frac{1}{2a^2}\ln\left(\frac{x^2}{a^2 + x^2}\right) \tag{E.3}$$

$$\int \frac{dx}{a^2x^2 - b^2} = \frac{1}{2ab}\ln\left(\frac{ax - b}{ax + b}\right) \tag{E.4a}$$

$$= -\frac{1}{ab}\coth^{-1}\left(\frac{ax}{b}\right), \qquad a^2x^2 > b^2 \tag{E.4b}$$

$$= -\frac{1}{ab}\tanh^{-1}\left(\frac{ax}{b}\right), \qquad a^2x^2 < b^2 \tag{E.4c}$$

$$\int \frac{dx}{\sqrt{a + bx}} = \frac{2}{b}\sqrt{a + bx} \tag{E.5}$$

$$\int \frac{dx}{\sqrt{x^2 + a^2}} = \ln(x + \sqrt{x^2 + a^2}) \tag{E.6}$$

$$\int \frac{x^2\,dx}{\sqrt{a^2 - x^2}} = -\frac{x}{2}\sqrt{a^2 - x^2} + \frac{a^2}{2}\sin^{-1}\frac{x}{a} \tag{E.7}$$

$$\int \frac{dx}{\sqrt{ax^2 + bx + c}} = \frac{1}{\sqrt{a}}\ln(2\sqrt{a}\sqrt{ax^2 + bx + c} + 2ax + b), \quad a > 0 \tag{E.8a}$$

$$= \frac{1}{\sqrt{a}}\sinh^{-1}\left(\frac{2ax + b}{\sqrt{4ac - b^2}}\right), \quad \begin{cases} a > 0 \\ 4ac > b^2 \end{cases} \tag{E.8b}$$

$$= -\frac{1}{\sqrt{-a}}\sin^{-1}\left(\frac{2ax + b}{\sqrt{b^2 - 4ac}}\right), \quad \begin{cases} a < 0 \\ b^2 > 4ac \\ |2ax + b| < \sqrt{b^2 - 4ac} \end{cases}$$

(E.8c)

$$\int \frac{x\,dx}{\sqrt{ax^2 + bx + c}} = \frac{1}{a}\sqrt{ax^2 + bx + c} - \frac{b}{2a}\int \frac{dx}{\sqrt{ax^2 + bx + c}} \tag{E.9}$$

$$\int \frac{dx}{x\sqrt{ax^2 + bx + c}} = -\frac{1}{\sqrt{c}}\sinh^{-1}\left(\frac{bx + 2c}{|x|\sqrt{4ac - b^2}}\right), \quad \begin{cases} c > 0 \\ 4ac > b^2 \end{cases} \tag{E.10a}$$

$$= \frac{1}{\sqrt{-c}}\sin^{-1}\left(\frac{bx + 2c}{|x|\sqrt{b^2 - 4ac}}\right), \quad \begin{cases} c < 0 \\ b^2 > 4ac \end{cases} \tag{E.10b}$$

$$= -\frac{1}{\sqrt{c}}\ln\left(\frac{2\sqrt{c}}{x}\sqrt{ax^2 + bx + c} + \frac{2c}{x} + b\right), \quad c > 0$$

(E.10c)

$$\int \sqrt{ax^2 + bx + c}\,dx = \frac{2ax + b}{4a}\sqrt{ax^2 + bx + c} + \frac{4ac - b^2}{8a}\int \frac{dx}{\sqrt{ax^2 + bx + c}} \tag{E.11}$$

E.2 Trigonometric Functions

$$\int \sin^2 x\, dx = \frac{x}{2} - \frac{1}{4}\sin 2x \tag{E.12}$$

$$\int \cos^2 x\, dx = \frac{x}{2} + \frac{1}{4}\sin 2x \tag{E.13}$$

$$\int \frac{dx}{a + b\sin x} = \frac{2}{\sqrt{a^2 - b^2}} \tan^{-1}\left[\frac{a\tan(x/2) + b}{\sqrt{a^2 - b^2}}\right], \quad a^2 > b^2 \tag{E.14}$$

$$\int \frac{dx}{a + b\cos x} = \frac{2}{\sqrt{a^2 - b^2}} \tan^{-1}\left[\frac{(a - b)\tan(x/2)}{\sqrt{a^2 - b^2}}\right], \quad a^2 > b^2 \tag{E.15}$$

$$\int \frac{dx}{(a + b\cos x)^2} = \frac{b\sin x}{(b^2 - a^2)(a + b\cos x)} - \frac{a}{b^2 - a^2}\int \frac{dx}{a + b\cos x} \tag{E.16}$$

$$\int \tan x\, dx = -\ln|\cos x| \tag{E.17a}$$

$$\int \tanh x\, dx = \ln\cosh x \tag{E.17b}$$

$$\int e^{ax}\sin x\, dx = \frac{e^{ax}}{a^2 + 1}(a\sin x - \cos x) \tag{E.18a}$$

$$\int e^{ax}\sin^2 x\, dx = \frac{e^{ax}}{a^2 + 4}\left(a\sin^2 x - 2\sin x\cos x + \frac{2}{a}\right) \tag{E.18b}$$

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\pi/a} \tag{E.18c}$$

E.3 Gamma Functions

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x}\, dx \tag{E.19a}$$

$$= \int_0^1 [\ln(1/x)]^{n-1}\, dx \tag{E.19b}$$

$$\Gamma(n) = (n-1)!, \quad \text{for } n = \text{positive integer} \tag{E.19c}$$

$$n\Gamma(n) = \Gamma(n+1) \tag{E.20}$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \tag{E.21}$$

$$\Gamma(1) = 1 \tag{E.22}$$

$$\Gamma\left(1\frac{1}{4}\right) = 0.906 \tag{E.23}$$

$$\Gamma\left(1\frac{3}{4}\right) = 0.919 \tag{E.24}$$

$$\Gamma(2) = 1 \tag{E.25}$$

$$\int_0^1 \frac{dx}{\sqrt{1-x^n}} = \frac{\sqrt{\pi}}{n} \frac{\Gamma\left(\frac{1}{n}\right)}{\Gamma\left(\frac{1}{n}+\frac{1}{2}\right)} \tag{E.26}$$

$$\int_0^1 x^m (1-x^2)^n\, dx = \frac{\Gamma(n+1)\Gamma\left(\frac{m+1}{2}\right)}{2\Gamma\left(n+\frac{m+3}{2}\right)} \tag{E.27a}$$

$$\int_0^{\pi/2} \cos^n x\, dx = \frac{\sqrt{\pi}}{2} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}+1\right)}, \quad n > -1 \tag{E.27b}$$

APPENDIX F
Differential Relations in Different Coordinate Systems

F.1 Rectangular Coordinates

$$\mathbf{grad}\, U = \nabla U = \sum_i \mathbf{e}_i \frac{\partial U}{\partial x_i} \tag{F.1}$$

$$\mathrm{div}\, \mathbf{A} = \nabla \cdot \mathbf{A} = \sum_i \frac{\partial A_i}{\partial x_i} \tag{F.2}$$

$$\mathbf{curl}\, \mathbf{A} = \nabla \times \mathbf{A} = \sum_{i,j,k} \varepsilon_{ijk} \frac{\partial A_k}{\partial x_j} \mathbf{e}_i \tag{F.3}$$

$$\nabla^2 U = \nabla \cdot \nabla U = \sum_i \frac{\partial^2 U}{\partial x_i^2} \tag{F.4}$$

F.2 Cylindrical Coordinates

Refer to Figures F-1 and F-2.

$$x_1 = r \cos \phi, \quad x_2 = r \sin \phi, \quad x_3 = z \tag{F.5}$$

$$r = \sqrt{x_1^2 + x_2^2}, \quad \phi = \tan^{-1} \frac{x_2}{x_1}, \quad z = x_3 \tag{F.6}$$

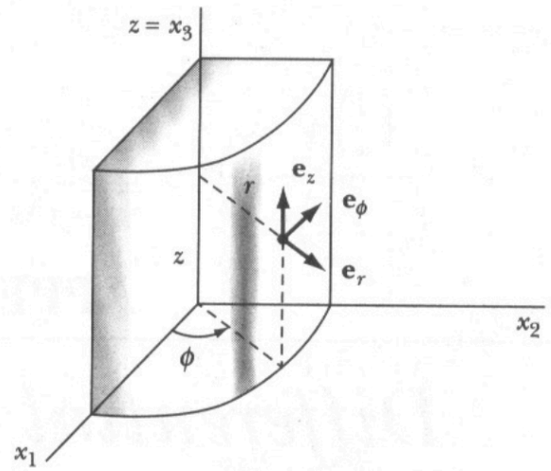


FIGURE F-1

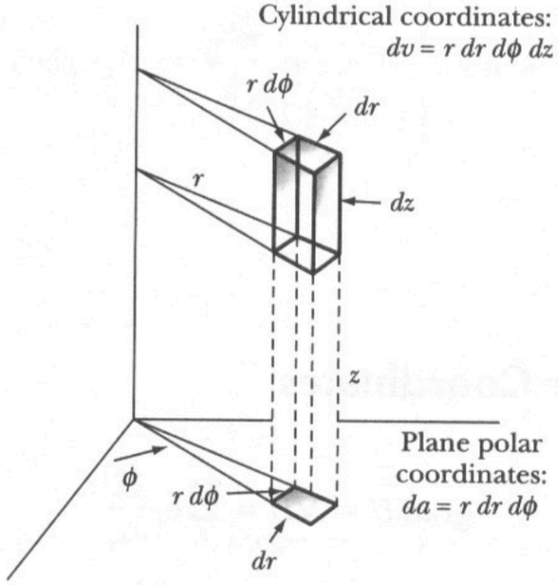


FIGURE F-2

$$ds^2 = dr^2 + r^2\,d\phi^2 + dz^2 \tag{F.7}$$

$$dv = r\,dr\,d\phi\,dz \tag{F.8}$$

$$\textbf{grad}\ \psi = \nabla\psi = \mathbf{e}_r\frac{\partial\psi}{\partial r} + \mathbf{e}_\phi\frac{1}{r}\frac{\partial\psi}{\partial\phi} + \mathbf{e}_z\frac{\partial\psi}{\partial z} \tag{F.9}$$

$$\text{div}\ \mathbf{A} = \frac{1}{r}\frac{\partial}{\partial r}(rA_r) + \frac{1}{r}\frac{\partial A_\phi}{\partial\phi} + \frac{\partial A_z}{\partial z} \tag{F.10}$$

$$\textbf{curl}\ \mathbf{A} = \mathbf{e}_r\left(\frac{1}{r}\frac{\partial A_z}{\partial\phi} - \frac{\partial A_\phi}{\partial z}\right) + \mathbf{e}_\phi\left(\frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r}\right) + \mathbf{e}_z\left(\frac{1}{r}\frac{\partial}{\partial r}(rA_\phi) - \frac{1}{r}\frac{\partial A_r}{\partial\phi}\right) \tag{F.11}$$

$$\nabla^2\psi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\psi}{\partial\phi^2} + \frac{\partial^2\psi}{\partial z^2} \tag{F.12}$$

F.3 Spherical Coordinates

Refer to Figures F-3 and F-4

$$x_1 = r \sin\theta \cos\phi, \quad x_2 = r \sin\theta \sin\phi, \quad x_3 = r \cos\theta \tag{F.13}$$

$$r = \sqrt{x_1^2 + x_2^2 + x_3^2}, \quad \theta = \cos^{-1}\frac{x_3}{r}, \quad \phi = \tan^{-1}\frac{x_2}{x_1} \tag{F.14}$$

$$ds^2 = dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2 \tag{F.15}$$

$$dv = r^2 \sin\theta\, dr\, d\theta\, d\phi \tag{F.16}$$

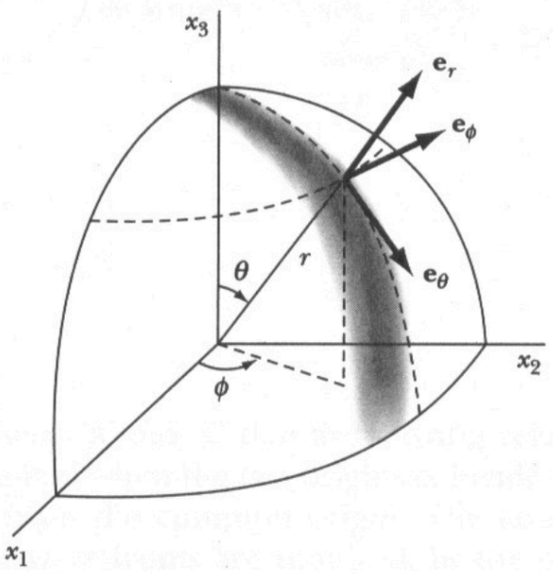


FIGURE F-3

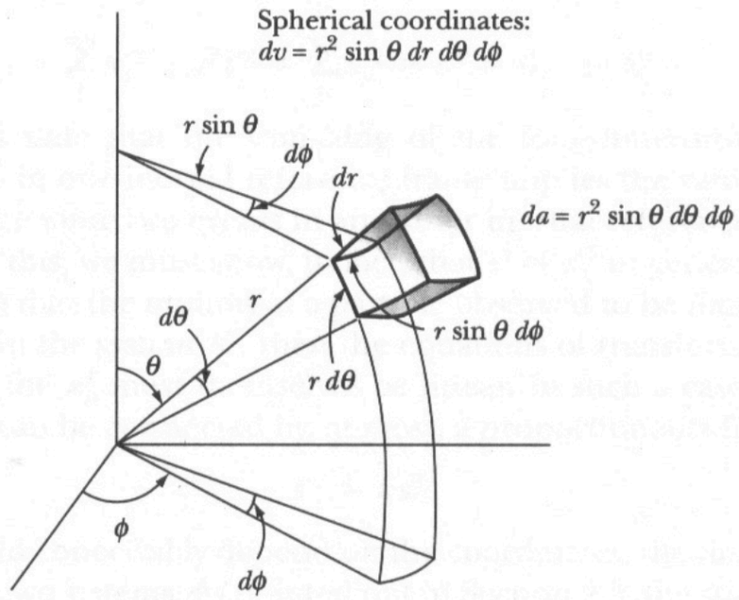


FIGURE F-4

$$\mathbf{grad}\,\psi = \nabla\psi = \mathbf{e}_r\frac{\partial\psi}{\partial r} + \mathbf{e}_\theta\frac{1}{r}\frac{\partial\psi}{\partial\theta} + \mathbf{e}_\phi\frac{1}{r\sin\theta}\frac{\partial\psi}{\partial\phi} \tag{F.17}$$

$$\mathrm{div}\mathbf{A} = \frac{1}{r^2}\frac{\partial}{\partial\mathbf{r}}(r^2A_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(A_\theta\sin\theta) + \frac{1}{r\sin\theta} + \frac{\partial A_\phi}{\partial\phi} \tag{F.18}$$

$$\mathbf{curl}\,\mathbf{A} = \mathbf{e}_r\frac{1}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(A_\phi\sin\theta) - \frac{\partial A_\theta}{\partial\phi}\right]$$

$$+ \mathbf{e}_\theta\frac{1}{r\sin\theta}\left[\frac{\partial A_r}{\partial\phi} - \sin\theta\frac{\partial}{\partial r}(rA_\phi)\right] + \mathbf{e}_\phi\frac{1}{r}\left[\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial\theta}\right] \tag{F.19}$$

$$\nabla^2\psi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2} \tag{F.20}$$

APPENDIX G

A "Proof of the Relation $\sum_\mu x_\mu^2 = \sum_\mu x_\mu'^2$

Consider the two inertial systems K and K' that are moving relative to one another with a speed v. At the instant when the two origins coincide (t = 0, t′= 0), let a light pulse be emitted from the common origin. The equations that describe the propagation of the wave fronts are required, by the second Einstein postulate, to be of the same form in the two systems:

$$\sum_j x_j^2 - c^2t^2 = \sum_\mu x_\mu^2 \equiv s^2 = 0, \quad \text{in } K \tag{G.1a}$$

$$\sum_j x_j'^2 - c^2t'^2 = \sum_\mu x_\mu'^2 \equiv s'^2 = 0, \quad \text{in } K' \tag{G.1b}$$

These equations state that the vanishing of the four-dimensional interval between two events in one inertial reference frame implies the vanishing of the interval between the same two events in any other inertial reference frame. But we need more than this; we must show, in fact, that $s^2 = s'^2$ in general.

If we require that the motion of a particle observed to be *linear* in the system K also be *linear* in the system K', then the equations of transformation that connect the x_μ and the x'_μ must themselves be linear. In such a case, the quadratic forms s^2 and s'^2 can be connected by, at most, a proportionality factor:

$$s'^2 = \kappa s^2 \tag{G.2a}$$

The factor k could conceivably depend on the coordinates, the time, and the relative speed of the two systems. As pointed out in Section 2.3, the space and time associated with an inertial reference frame are *homogeneous*, so the relation between s^2 and s'^2 cannot be different at different points in space nor at different instants of time. Therefore, the factor κ cannot depend on either the coordinates or the time. A dependence on v is still allowed, however, but the isotropy of space forbids a dependence on the direction of v. We have therefore reduced the possible dependence of s^2 on s'^2 to a factor that involves at most the magnitude of the speed v; that is, we have

$$s'^2 = \kappa(v) s^2 \tag{G.2b}$$

If we make the transformation from K' back to K, we have the result

$$s^2 = \kappa(-v) s'^2$$

where -v occurs because the velocity of K relative to K' is the negative of the velocity of K' relative to K. But we have already argued that the factor κ can depend only on the *magnitude* of v. We therefore have the two equations

$$\left.\begin{aligned} s'^2 &= \kappa(v) s^2 \\ s^2 &= \kappa(v) s'^2 \end{aligned}\right\} \tag{G.3}$$

Combining these equations, we conclude that $\kappa^2 = 1$, or $\kappa(\mathrm{v}) = \pm 1$. The value of $\kappa(\mathrm{v})$ must not be a discontinuous function of v; that is, if we change v at some rate, κ cannot suddenly jump from +1 to -1. In the limit of zero velocity, the systems K and K' become identical, so that $\kappa(\mathrm{v} = 0) = +1$. Hence,

$$\kappa = +1 \tag{G.4}$$

for all values of the velocity, and we have, finally,

$$s^2 = s'^2 \tag{G.5}$$

This important result states that the four-dimensional interval between two events is the same in all inertial reference frames.

APPENDIX H

Numerical Solution for Example 2.7

In this appendix, we show the MathCad solution that produced Figures 2-8 and 2-9 for Example 2.7. This program was written for MathCad for Windows, version 4.0.

$g := 9.8$ acceleration of gravity

$th := 60 \cdot \left(\frac{\pi}{180}\right)$ initial angle

$vo := 600$

$u := vo \cdot \cos(th)$ initial velocity

$v := vo \cdot \sin(th)$ initial horizontal velocity

$i := 1 .. 6$ initial vertical velocity

$k_i :=$

0.0000001
0.01
0.02
0.04
0.08
0.005

table of drag coefficients

$t := 0, 1 .. 130$ range of time values

$x(t, K) := \left(\left(\frac{u}{K}\right)\right) \cdot (1 - \exp(-K \cdot t))$ calculate horizontal position

$$y(t, K) := -g \cdot \frac{t}{K} + \frac{K \cdot v + g}{(K)^2} \cdot (1 - \exp(-K \cdot t)) \qquad \text{calculate vertical position}$$

[Now plot $y(t,k_j)$ versus $x(t,k_j)$ to produce Figure 2-8.]

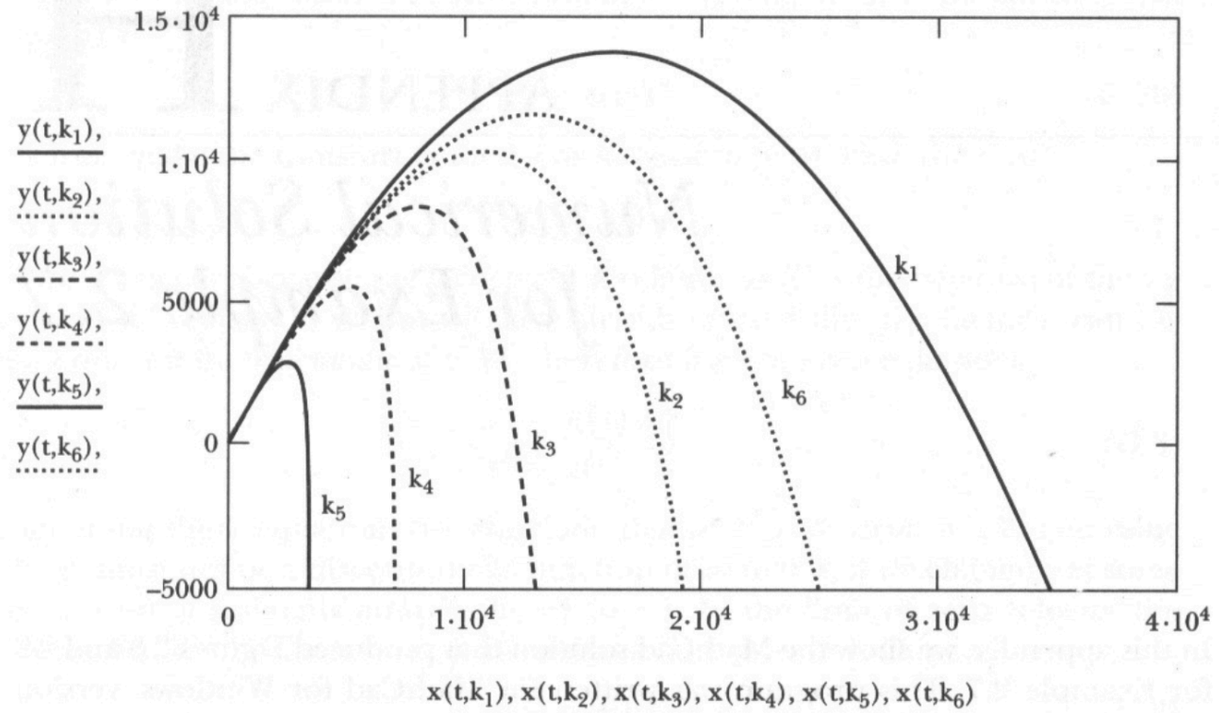


FIGURE 2-8

Now set up an equation to solve Equation 2.45 for T for any value of the retarding force constant k.

$$f(k, T) := \text{root}\left[T - \frac{k \cdot v + g}{g \cdot k} \cdot (1 - \exp(-k \cdot T)), T\right] \qquad \textbf{(2.45)}$$

$j := 1, 2 .. 81$ Set up range of values to calculate; 80 values.

$K_j := -0.001 + 0.001 \cdot j + 0.00000001$ This will allow us to calculate over a range of k values from 0 to 0.08.

$Tr_0 := 100$ The time value for k = 0 is 106 s. This is a guess to get the calculation started.

$Tr_j := f(K_j, Tr_{j-1})$ We now determine the solution for the time T for all the values of k. Solve Equation 2.45.

$Tr_1 := 106.074$ This is the value of T for k1. We do not bother to calculate all the others here.

Now we want to calculate the range R for all the values of T (as a function of k) that we have just found. To do this, we need to solve Equation 2.43 for each of the values of k and t = T that we have just found.

$x := 100$

This is the guess for the first value of x. The actual value of the guess does not matter.

$$f(k, T) := \text{root}\left[x - \frac{u}{k} \cdot (1 - \exp(-k \cdot T)), x\right]$$

This is the Equation 2.43 that we need to solve to find the range R.

$R_j := f(K_j, Tr_j)$

Now calculate the range R for all the values.

$R_1 = 3.182 \cdot 10^4$

We just list the first value and plot the remainder. This is the range for no air resistance, that is k = 0.

Now let's calculate and plot the range determined from the approximate calculation. Calculate Equation 2.55.

$$Rp_j := R_1 \cdot \left(1 - \frac{4 \cdot K_j \cdot v}{3 \cdot g}\right)$$

[Now plot R_j and R_{pj} versus K_j to produce Figure 2-9.]

Plot approximate and numerical solutions. Figure 2-9.

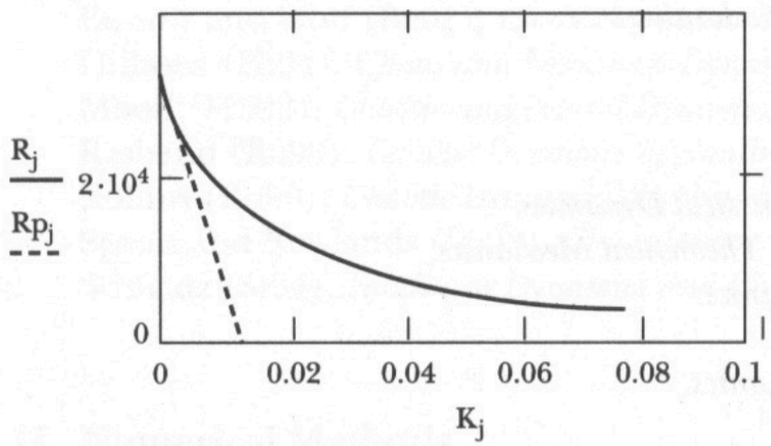


FIGURE 2-9