PROBLEMS FOR CHAPTER 5

1. Show that for any two complex numbers z_1 and z_2

$$|z_1 + z_2|^2 + |z_1 - z_2|^2 = 2|z_1|^2 + 2|z_2|^2$$

2. Given a complex number,

$$z = \frac{5 + i\,10}{(1 - i\,2)(2 - i)},$$

determine the real and imaginary parts of z.

3. Determine the real part, imaginary part, and the absolute magnitude of the following three expressions,

$$\sin(i \ln i), \qquad \sqrt{i} \qquad \text{and} \qquad \frac{1}{2i}\ln\left(\frac{1 + ia}{1 - ia}\right).$$

4. If $z = x + iy$, determine the real part, imaginary part and the absolute magnitude of the following three functions,

(a) $\sin z$, (b) $\cos z$ and (c) $\exp(iz)$.

5. For two complex numbers $z_1 = \sqrt{3} + i$ and $z_2 = -\sqrt{2} + i\sqrt{2}$, determine

(a) $|z_1/z_2|^4$ and (b) $(z_1/z_2)^4$.

6. Find all roots of the equation $z^6 - 1 = 0$ and plot the results in a complex plane.

7. The three cube roots of +1 are of the form 1, ω and ω^2. Determine $Re\,[\omega]$ and $Im\,[\omega]$.

8. The three cube roots of -1 are of the form α, α^3 and α^5. Determine $Re\,[\alpha]$ and $Im\,[\alpha]$.

9. Determine the fourth root of $z = -8 + i\,8\sqrt{3}$.

10. Determine the third (or, cube) root of $z = -2 + i\,2$.

11. Use DeMoivre's formula to express $\sin^3\theta$ and $\cos^4\theta$ in terms of sine and/or cosine of multiples of θ.

12. Use DeMoivre's formula to express $\cos(3\theta)$ and $\sin(3\theta)$ in terms of powers of $\sin\theta$ and $\cos\theta$.

Chapter 6: Determinants

In this chapter we will first define a determinant as well as its minors and cofactors. Next, we will outline a procedure, due to Laplace, for determining the value of the determinant in terms of its cofactors. Several properties of determinants will be described which help in simplifying the determinant so that it can be reduced and evaluated easily. Two final examples will help in understanding the utility of properties in the simplification process for a determinant.

6.1 DEFINITION OF A DETERMINANT

A determinant is a *square* array of numbers (or elements) such as

$$\mathbb{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nj} & \cdots & a_{nn} \end{vmatrix},$$

which are combined together according to certain rules to provide the value of the determinant. The element a_{ij} is located at the intersection of i^{th} row and j^{th} column. The number of rows (or columns) is called the order of the determinant. The determinant shown above is of order n. We also define the *minor* and the *cofactor* of each element of a determinant. Starting with a determinant of order n, one can obtain a smaller determinant by omitting one row and one column. The determinant of order $(n-1)$ obtained by omitting the row and column containing a_{ij} (that is, omitting i^{th} row and j^{th} column) is called the minor M_{ij} of a_{ij} and the minor multiplied by a sign of $(-1)^{i+j}$ is called the cofactor C_{ij} of a_{ij}.

Let us start evaluating some determinants. A single number can be considered as a determinant of order 1 and the value of this determinant is the value of the single number.

Next higher order determinant is of order 2. Consider a determinant of order 2,

$$\mathbb{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

According to the rules of determining the values of a determinant, the value of a determinant of order 2 is the difference between the products of elements along two diagonals as

$$\det(\mathbb{A}) = a_{11}a_{22} - a_{21}a_{12}\ .$$

Note, if we interchange the rows and columns of this determinant to form a new determinant $\mathbb{B}$,

$$\mathbb{B} = \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} ,$$

then

$$\det(\mathbb{B}) = \det(\mathbb{A}) \ .$$

Next higher order determinant is of order 3. A typical determinant of order 3 is

$$\mathbb{A} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \ .$$

Again, according to the rules of determining values of a determinant, the value of a determinant of order 3 can be written in terms of Levi-Civita symbols as

$$\det(\mathbb{A}) = \sum_{ijk} \epsilon_{ijk} a_{i1} a_{j2} a_{k3} = \epsilon_{123} a_{11} a_{22} a_{33} + \epsilon_{132} a_{11} a_{32} a_{23} + \epsilon_{213} a_{21} a_{12} a_{33} + \epsilon_{231} a_{21} a_{32} a_{13}$$
$$+ \epsilon_{312} a_{31} a_{12} a_{23} + \epsilon_{321} a_{31} a_{22} a_{13} = \sum_{ijk} \epsilon_{ijk} a_{1i} a_{2j} a_{3k} \ .$$

The last equality suggests that the value of a determinant of order 3 is not changed if its rows and columns are interchanged. In other words, if a determinant $\mathbb{B}$ is obtained from $\mathbb{A}$ by interchanging its rows and columns, then,

$$\det(\mathbb{B}) = \det(\mathbb{A}) \ .$$

Substituting explicitly the values of all Levi-Civita symbols, we get

$$\det(\mathbb{A}) = a_{11} a_{22} a_{33} - a_{11} a_{32} a_{23} - a_{21} a_{12} a_{33} + a_{21} a_{32} a_{13} + a_{31} a_{12} a_{23} - a_{31} a_{22} a_{13} \ .$$

Note that each product on the right-hand side contains only one element from each row and from each column of the determinant. Furthermore, factors on the right-hand side can be rearranged either as

$$\det(\mathbb{A}) = a_{11}(a_{22} a_{33} - a_{32} a_{23}) - a_{21}(a_{12} a_{33} - a_{13} a_{32}) + a_{31}(a_{12} a_{23} - a_{13} a_{22})$$

$$= a_{11}(-1)^{1+1} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{21}(-1)^{2+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}(-1)^{3+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

$$= a_{11}(-1)^{1+1} M_{11} + a_{21}(-1)^{2+1} M_{21} + a_{31}(-1)^{3+1} M_{31}$$

$$= a_{11}C_{11} + a_{21}C_{21} + a_{31}C_{31} \ ,$$

or as

$$\det(\mathbb{A}) = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{31}a_{22})$$

$$= a_{11}(-1)^{1+1}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{12}(-1)^{1+2}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}(-1)^{1+3}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$$= a_{11}(-1)^{1+1}M_{11} + a_{12}(-1)^{1+2}M_{12} + a_{13}(-1)^{1+3}M_{13}$$

$$= a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} \ .$$

These latter forms suggest an alternate way, due to Laplace, of obtaining the value of a determinant of any order n as

$$\det(\mathbb{A}) = \sum_{i=1}^{n} a_{i1}C_{i1} \ ,$$

where the expansion of the determinant is across first column, or as

$$\det(\mathbb{A}) = \sum_{i=1}^{n} a_{1i}C_{1i} \ ,$$

where the expansion of the determinant is across first row. It suggests that rows and columns of a determinant can be interchanged without affecting its value. In fact, the value of a determinant can be obtained by expanding across *any* row or *any* column. We illustrate this fact in the example below.

Example: Determine the value of the following determinant of order 3, by expanding it across any row or any column.

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{2} & \mathbf{1} \\ \mathbf{0} & \mathbf{-1} & \mathbf{2} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} \end{vmatrix} .$$

Solution: Expanding across first row,

$$\det(\mathbb{A}) = (1)(-1)^{1+1}\begin{vmatrix} -1 & 2 \\ 1 & 0 \end{vmatrix} + (2)(-1)^{1+2}\begin{vmatrix} 0 & 2 \\ 1 & 0 \end{vmatrix} + (1)(-1)^{1+3}\begin{vmatrix} 0 & -1 \\ 1 & 1 \end{vmatrix}$$

$$= 1(0-2) - 2(0-2) + 1(0+1) = +3 \ .$$

Expanding across second row,

$$\det(\mathbb{A}) = (0)(-1)^{2+1}\begin{vmatrix}2 & 1\\ 1 & 0\end{vmatrix} + (-1)(-1)^{2+2}\begin{vmatrix}1 & 1\\ 1 & 0\end{vmatrix} + (2)(-1)^{2+3}\begin{vmatrix}1 & 2\\ 1 & 1\end{vmatrix}$$

$$= -0(0-1) - 1(0-1) - 2(1-2) = +3 \ .$$

Expanding across third row,

$$\det(\mathbb{A}) = (1)(-1)^{3+1}\begin{vmatrix}2 & 1\\ -1 & 2\end{vmatrix} + (1)(-1)^{3+2}\begin{vmatrix}1 & 1\\ 0 & 2\end{vmatrix} + (0)(-1)^{3+3}\begin{vmatrix}1 & 2\\ 0 & -1\end{vmatrix}$$

$$= 1(4+1) - 1(2-0) + 0(-1+0) = +3 \ .$$

Expanding across first column,

$$\det(\mathbb{A}) = (1)(-1)^{1+1}\begin{vmatrix}-1 & 2\\ 1 & 0\end{vmatrix} + (0)(-1)^{2+1}\begin{vmatrix}2 & 1\\ 1 & 0\end{vmatrix} + (1)(-1)^{3+1}\begin{vmatrix}2 & 1\\ -1 & 2\end{vmatrix}$$

$$= 1(0-2) - 0(0-1) + 1(4+1) = +3 \ .$$

Expanding across second column,

$$\det(\mathbb{A}) = (2)(-1)^{1+2}\begin{vmatrix}0 & 2\\ 1 & 0\end{vmatrix} + (-1)(-1)^{2+2}\begin{vmatrix}1 & 1\\ 1 & 0\end{vmatrix} + (1)(-1)^{3+2}\begin{vmatrix}1 & 1\\ 0 & 2\end{vmatrix}$$

$$= -2(0-2) - 1(0-1) - 1(2-0) = +3 \ .$$

Expanding across third column,

$$\det(\mathbb{A}) = (1)(-1)^{1+3}\begin{vmatrix}0 & -1\\ 1 & 1\end{vmatrix} + (2)(-1)^{2+3}\begin{vmatrix}1 & 2\\ 1 & 1\end{vmatrix} + (0)(-1)^{3+3}\begin{vmatrix}1 & 2\\ 0 & -1\end{vmatrix}$$

$$= 1(0+1) - 2(1-2) + 0(-1-0) = +3 \ .$$

Thus, in general,

$$\sum_k a_{ik}C_{jk} = \sum_k C_{ki}a_{kj} = \delta_{ij}\det(\mathbb{A}) \ . \qquad Eq.(6.1)$$

6.2 PROPERTIES OF DETERMINANT $\mathbb{A}$

There are several very useful properties of determinants which are helpful in determining the value of a large determinant. Since the value of a determinant does not change on interchanging the rows and the columns, all the remarks below for columns of a determinant will apply equally well to the rows of the determinant.

As a shorthand notation for writing the determinant, let us write A_l for the l^{th} column of $\mathbb{A}$. Then

$$\mathbb{A} = |A_1, A_2, \ldots A_l, \ldots A_n| \ .$$

Property Number 1: If any two columns of a determinant $\mathbb{A}$ are interchanged, then the value of the resulting determinant $\mathbb{B}$ is same as the value of the original determinant except for an overall change of sign. In other words, if

$$\mathbb{A} = |A_1, A_2 \ldots A_l, \ldots A_m, \ldots A_n| \ ,$$

and

$$\mathbb{B} = |A_1, A_2 \ldots A_m, \ldots A_l, \ldots A_n| \ ,$$

then

$$\det \mathbb{B} = -\det \mathbb{A} \ .$$

To prove this fact, we first assume that the l^{th} column and the m^{th} column of $\mathbb{A}$ are two adjacent columns, that is,

$$\mathbb{A} = |A_1, A_2 \ldots A_l, A_m, \ldots A_n|$$

and

$$\mathbb{B} = |A_1, A_2 \ldots A_m, A_l, \ldots A_n| \ .$$

Now, to determine the values of these two determinants, let us expand each determinant across the column that contains elements of the original l^{th} column. Each element of this column will have the same minor both in $\mathbb{A}$ and in $\mathbb{B}$; however, the cofactor of each element will differ in sign only since the l^{th} column is moved from left of m^{th} column to the right of m^{th} column. So, $\det \mathbb{B} = -\det \mathbb{A}$. This result is true even if the l^{th} column and the m^{th} column of $\mathbb{A}$ are not adjacent columns. If the l^{th} column and the m^{th} column are separated by j columns, such as

$$\mathbb{A} = |A_1, A_2 \ldots A_l, \leftarrow j \text{ columns } \rightarrow, A_m, \ldots A_n| \ ,$$

then a total of $(2j + 1)$ interchanges of adjacent columns are required to exchange the l^{th} column and the m^{th} column and since $(-1)^{2j+1} = (-1)$ for any j, the fact that $\det \mathbb{B} = -\det \mathbb{A}$ is proved.

Example: A determinant $\mathbb{B}$ is obtained by interchanging the first two columns of another determinant $\mathbb{A}$ as shown below:

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} & \mathbf{-2} \\ \mathbf{3} & \mathbf{1} & \mathbf{-4} \end{vmatrix} \textbf{ and } \mathbb{B} = \begin{vmatrix} \mathbf{-1} & \mathbf{1} & \mathbf{1} \\ \mathbf{3} & \mathbf{2} & \mathbf{-2} \\ \mathbf{1} & \mathbf{3} & \mathbf{-4} \end{vmatrix} .$$

Show that $\det \mathbb{B} = -\det \mathbb{A}$.

Solution: Expanding across first row, we get

$$\det(\mathbb{A}) = (1)(-1)^{1+1}\begin{vmatrix} 3 & -2 \\ 1 & -4 \end{vmatrix} + (-1)(-1)^{1+2}\begin{vmatrix} 2 & -2 \\ 3 & -4 \end{vmatrix} + (1)(-1)^{1+3}\begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix}$$

$$= 1(-12+2) + 1(-8+6) + 1(2-9) = -19$$

and

$$\det(\mathbb{B}) = (-1)(-1)^{1+1}\begin{vmatrix} 2 & -2 \\ 3 & -4 \end{vmatrix} + (1)(-1)^{1+2}\begin{vmatrix} 3 & -2 \\ 1 & -4 \end{vmatrix} + (1)(-1)^{1+3}\begin{vmatrix} 3 & 2 \\ 1 & 3 \end{vmatrix}$$

$$= -1(-8+6) - 1(-12+2) + 1(9-2) = +19 \ .$$

Property Number 2: If a determinant has two identical columns, such as

$$\mathbb{A} = |A_1 A_2 \dots A_l \dots A_l \dots A_n| \quad ,$$

then $\det \mathbb{A} = 0$.

The proof of this statement easily follows from Property Number 1, since on interchanging the two columns we get a negative sign, while the determinant stays the same since the two exchanged columns are identical. Thus,

$$\det(\mathbb{A}) = -\det(\mathbb{A}) \ .$$

Or

$$\det(\mathbb{A}) = 0 \ .$$

Example: Two columns of a determinant $\mathbb{A}$ are identical. Determine the value of this determinant.

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{-1} \\ \mathbf{2} & \mathbf{3} & \mathbf{3} \\ \mathbf{3} & \mathbf{1} & \mathbf{1} \end{vmatrix} .$$

Solution: Expanding the determinant across first row, we get

$$\det(\mathbb{A}) = (1)(-1)^{1+1}\begin{vmatrix}3 & 3\\ 1 & 1\end{vmatrix} + (-1)(-1)^{1+2}\begin{vmatrix}2 & 3\\ 3 & 1\end{vmatrix} + (-1)(-1)^{1+3}\begin{vmatrix}2 & 3\\ 3 & 1\end{vmatrix}$$

$$= 1(3-3) + 1(2-9) - 1(2-9) = 0 \ .$$

Property Number 3: If each element of a column of a determinant is multiplied by a constant k, then the value of the whole determinant is multiplied by the same constant k. Or, saying differently, multiplying a determinant by a constant k amounts to multiplying any one of the columns of the determinant by k.

Suppose determinant $\mathbb{B}$ is constructed by multiplying the l^{th} column of determinant $\mathbb{A}$ by a constant number k. Then

$$\mathbb{A} = \ |A_1, A_2 \dots A_l, \dots A_n\,|$$

and

$$\mathbb{B} = \ |A_1, A_2 \dots kA_l, \dots A_n\,| \ .$$

Also,

$$\det(\mathbb{A}) = \sum_{i=1}^{n} a_{il}C_{il}$$

and

$$\det(\mathbb{B}) = \sum_{i=1}^{n} (ka_{il})C_{il} = k\sum_{i=1}^{n} a_{il}C_{il} = k\,\det(\mathbb{A}) \ .$$

Example: A new determinant $\mathbb{B}$ is formed by multiplying the third column of another determinant $\mathbb{A}$ by a constant number $\boldsymbol{k = 5}$, as shown:

$$\mathbb{A} = \begin{vmatrix}\mathbf{1} & \mathbf{-1} & \mathbf{1}\\ \mathbf{2} & \mathbf{3} & \mathbf{-2}\\ \mathbf{3} & \mathbf{1} & \mathbf{-4}\end{vmatrix} \textbf{ and } \mathbb{B} = \begin{vmatrix}\mathbf{1} & \mathbf{-1} & \mathbf{5}\\ \mathbf{2} & \mathbf{3} & \mathbf{-10}\\ \mathbf{3} & \mathbf{1} & \mathbf{-20}\end{vmatrix} \ .$$

Show that $\det \mathbb{B} = \boldsymbol{k}\ \det \mathbb{A}$.

Solution: Determinant $\mathbb{A}$ is the same determinant as in the previous example. Its value, as determined previously, is -19.

Now, expanding the determinant $\mathbb{B}$ across the first column, we get

$$\det(\mathbb{B}) = (1)(-1)^{1+1}\begin{vmatrix} 3 & -10 \\ 1 & -20 \end{vmatrix} + (2)(-1)^{2+1}\begin{vmatrix} -1 & 5 \\ 1 & -20 \end{vmatrix} + (3)(-1)^{3+1}\begin{vmatrix} -1 & 5 \\ 3 & -10 \end{vmatrix}$$

$$= 1(-60 + 10) - 2(20 - 5) + 3(10 - 15) = -95 = 5(-19)\ .$$

Property Number 4: If each element of the l^{th} column of a determinant $\mathbb{A}$ can be expressed as

$a_{jl} = \alpha p_{jl} + \beta q_{jl}$ with $j = 1,2 \cdots n$, then, columnwise,

$$A_l = \alpha P_l + \beta Q_l\ .$$

Now, expanding the determinant $\mathbb{A}$ across its l^{th} column

$$\det(\mathbb{A}) = \det|A_1 A_2 \cdots A_{l-1}\ \alpha P_l + \beta Q_l\ A_{l+1} \cdots A_n|$$

$$= \sum_{j=1}^{n} (\alpha p_{jl} + \beta q_{jl}) C_{jl} = \alpha \sum_{j=1}^{n} p_{jl} C_{jl} + \beta \sum_{j=1}^{n} q_{jl} C_{jl}$$

$$= \alpha \det|A_1 A_2 \cdots A_{l-1}\ P_l\ A_{l+1} \cdots A_n| + \beta \det|\ A_1 A_2 \cdots A_{l-1}\ Q_l\ A_{l+1} \cdots A_n|$$

$$= \alpha\ \det(\mathbb{P}) + \beta\ \det(\mathbb{Q})\ .$$

Note, the determinants $\mathbb{A}$, $\mathbb{P}$ and $\mathbb{Q}$ share all columns except the l^{th} column.

Example: Two new determinants, $\mathbb{P}$ and $\mathbb{Q}$, are created by splitting the third column of another determinant $\mathbb{A}$, as shown:

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} & \mathbf{-2} \\ \mathbf{3} & \mathbf{1} & \mathbf{-4} \end{vmatrix} \textbf{ with } \mathbb{P} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} & \mathbf{-1} \\ \mathbf{3} & \mathbf{1} & \mathbf{-2} \end{vmatrix} \textbf{ and } \mathbb{Q} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{0} \\ \mathbf{2} & \mathbf{3} & \mathbf{-1} \\ \mathbf{3} & \mathbf{1} & \mathbf{-2} \end{vmatrix}\ .$$

Show that $\det(\mathbb{A}) = \det(\mathbb{P}) + \det(\mathbb{Q})$.

Solution: Determinant $\mathbb{A}$ is the determinant of previous few examples. Its value, as determined previously, is -19.

Now, expanding the determinants $\mathbb{P}$ and $\mathbb{Q}$ across the first row, we get

$$\det(\mathbb{P}) = (1)(-1)^{1+1}\begin{vmatrix} 3 & -1 \\ 1 & -2 \end{vmatrix} + (-1)(-1)^{1+2}\begin{vmatrix} 2 & -1 \\ 3 & -2 \end{vmatrix} + (1)(-1)^{1+3}\begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix}$$

$$= 1(-6 + 1) + 1(-4 + 3) + 1(2 - 9) = -5 - 1 - 7 = -13\ ,$$

$$\det(\mathbb{Q}) = (1)(-1)^{1+1}\begin{vmatrix} 3 & -1 \\ 1 & -2 \end{vmatrix} + (-1)(-1)^{1+2}\begin{vmatrix} 2 & -1 \\ 3 & -2 \end{vmatrix} + (0)(-1)^{1+3}\begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix}$$

$$= 1(-6+1) + 1(-4+3) - 0 = -5 - 1 = -6 \ .$$

Clearly, $\det(\mathbb{A}) = \det(\mathbb{P}) + \det(\mathbb{Q})$.

Property Number 5: If a determinant $\mathbb{B}$ is obtained from determinant $\mathbb{A}$ by adding to the l^{th} column of $\mathbb{A}$ a scalar multiple of any other column of $\mathbb{A}$, then $\det \mathbb{B} = \det \mathbb{A}$.

Explicitly,

$$\mathbb{B} = |A_1 A_2 \cdots A_l + \alpha A_m \cdots A_m \cdots A_n| \ .$$

Using Property Number 4,

$$\det \mathbb{B} = \det|A_1 A_2 \cdots A_l \cdots A_m \cdots A_n| + \alpha \det|A_1 A_2 \cdots A_m \cdots A_m \cdots A_n| \ .$$

Or, using Property Number 2,

$$\det \mathbb{B} = \det|A_1 A_2 \cdots A_l \cdots A_m \cdots A_n| + 0 = \det \mathbb{A} \ .$$

Example: A new determinant $\mathbb{B}$ is created by adding first column to the third column of another determinant $\mathbb{A}$ as shown:

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} & \mathbf{-2} \\ \mathbf{3} & \mathbf{1} & \mathbf{-4} \end{vmatrix} \textbf{ and } \mathbb{B} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{2} \\ \mathbf{2} & \mathbf{3} & \mathbf{0} \\ \mathbf{3} & \mathbf{1} & \mathbf{-1} \end{vmatrix} .$$

Show that $\det \mathbb{B} = \det \mathbb{A}$.

Solution: Determinant $\mathbb{A}$ is the same determinant as in the previous examples. Its value, as determined previously, is -19.

Now, expanding the determinant $\mathbb{B}$ across the third column, we get

$$\det(\mathbb{B}) = (2)(-1)^{1+3}\begin{vmatrix} 2 & 3 \\ 3 & 1 \end{vmatrix} + (0)(-1)^{2+3}\begin{vmatrix} 1 & -1 \\ 3 & 1 \end{vmatrix} + (-1)(-1)^{3+3}\begin{vmatrix} 1 & -1 \\ 2 & 3 \end{vmatrix}$$

$$= 2(2-9) - 0 - 1(3+2) = -14 - 5 = -19 \ .$$

Finally, we provide two examples in which we use properties of determinants to first simplify the determinant, then evaluate it.

Example: Using properties of determinants, find out the value of

$$\mathbb{A} = \begin{vmatrix} \mathbf{1} & \mathbf{-1} & \mathbf{1} \\ \mathbf{2} & \mathbf{3} & \mathbf{-2} \\ \mathbf{3} & \mathbf{1} & \mathbf{-4} \end{vmatrix}.$$

Solution: We have already seen that the value of $\mathbb{A}$ is -19. Now we will use various properties of determinants to create as many zeros in a single row or in a single column as possible. We will do so for the first row of determinant $\mathbb{A}$.

First, we add column 2 to column 3 and use property number 5 to get

$$\begin{vmatrix} 1 & -1 & 1 \\ 2 & 3 & -2 \\ 3 & 1 & -4 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 3 & 1 \\ 3 & 1 & -3 \end{vmatrix}.$$

Next, we add column 1 to column 2 and use property number 5 again to get

$$\begin{vmatrix} 1 & -1 & 1 \\ 2 & 3 & -2 \\ 3 & 1 & -4 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 3 & 1 \\ 3 & 1 & -3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 5 & 1 \\ 3 & 4 & -3 \end{vmatrix}.$$

Now, we expand across row 1 to get,

$$\begin{vmatrix} 1 & -1 & 1 \\ 2 & 3 & -2 \\ 3 & 1 & -4 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 3 & 1 \\ 3 & 1 & -3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 5 & 1 \\ 3 & 4 & -3 \end{vmatrix} = 1 \begin{vmatrix} 5 & 1 \\ 4 & -3 \end{vmatrix}.$$

The determinant of order 2 is easy to evaluate as

$$\begin{vmatrix} 1 & -1 & 1 \\ 2 & 3 & -2 \\ 3 & 1 & -4 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 3 & 1 \\ 3 & 1 & -3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 5 & 1 \\ 3 & 4 & -3 \end{vmatrix} = 1 \begin{vmatrix} 5 & 1 \\ 4 & -3 \end{vmatrix} = -19.$$

As seen in the above example, the trick of simplifying the evaluation of a determinant is to create as many zeros in a single row (or column) as possible and then use this row (or column) to evaluate the determinant.

Example: Given that $\begin{vmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ \boldsymbol{d} & \boldsymbol{e} & \boldsymbol{f} \\ \boldsymbol{g} & \boldsymbol{h} & \boldsymbol{i} \end{vmatrix} = 5$, use the properties of general determinants to find the value of

$$\mathbb{D} = \begin{vmatrix} \boldsymbol{-2a} & \boldsymbol{3d-g} & \boldsymbol{g+a} \\ \boldsymbol{-2b} & \boldsymbol{3e-h} & \boldsymbol{h+b} \\ \boldsymbol{-2c} & \boldsymbol{3f-i} & \boldsymbol{i+c} \end{vmatrix}.$$

Solution: The determinant to be evaluated is

$$\mathbb{D} = \begin{vmatrix} -2a & 3d-g & g+a \\ -2b & 3e-h & h+b \\ -2c & 3f-i & i+c \end{vmatrix} .$$

First, take -2 common out of column 1 to get

$$\mathbb{D} = (-2)\begin{vmatrix} a & 3d-g & g+a \\ b & 3e-h & h+b \\ c & 3f-i & i+c \end{vmatrix} .$$

Now, subtract column 1 from column 3 to get

$$\mathbb{D} = (-2)\begin{vmatrix} a & 3d-g & g \\ b & 3e-h & h \\ c & 3f-i & i \end{vmatrix} .$$

Adding column 3 to column 2 gives

$$\mathbb{D} = (-2)\begin{vmatrix} a & 3d & g \\ b & 3e & h \\ c & 3f & i \end{vmatrix} .$$

Taking a common factor of 3 out of column 2, we have

$$\mathbb{D} = (-2)(3)\begin{vmatrix} a & d & g \\ b & e & h \\ c & f & i \end{vmatrix} .$$

Finally, interchange rows and columns of the determinant,

$$\mathbb{D} = (-6)\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = (-6)(5) = -30 .$$

PROBLEMS FOR CHAPTER 6

1. Determine the values of the following two determinants

$$\begin{vmatrix} 5 & 2 & 3 \\ -1 & 3 & 4 \\ 0 & -2 & 3 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} -2 & 3 & 5 \\ 1 & 2 & 4 \\ 0 & 3 & -2 \end{vmatrix}.$$

2. Using the properties of determinants, first simplify and then determine the value of the following determinant,

$$\begin{vmatrix} a-b & a^2-b^2 & a^3-b^3 \\ b-c & b^2-c^2 & b^3-c^3 \\ c-a & c^2-a^2 & c^3-a^3 \end{vmatrix}.$$

3. Without expanding the determinant, and using the properties of determinants only, show that

$$\begin{vmatrix} bc & a^2 & a^2 \\ b^2 & ca & b^2 \\ c^2 & c^2 & ab \end{vmatrix} = \begin{vmatrix} bc & ab & ca \\ ab & ca & bc \\ ca & bc & ab \end{vmatrix}.$$

4. Without expanding the determinant, and using the properties of determinants only, show that

$$\begin{vmatrix} n\,a_1 + b_1 & n\,a_2 + b_2 & n\,a_3 + b_3 \\ n\,b_1 + c_1 & n\,b_2 + c_2 & n\,b_3 + c_3 \\ n\,c_1 + a_1 & n\,c_2 + a_2 & n\,c_3 + a_3 \end{vmatrix} = (n+1)(n^2 - n + 1) \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

5. Without expanding the determinant, and using the properties of determinants only, show that

$$\begin{vmatrix} a^2 & (a+1)^2 & (a+2)^2 \\ b^2 & (b+1)^2 & (b+2)^2 \\ c^2 & (c+1)^2 & (c+2)^2 \end{vmatrix} = -4(a-b)(b-c)(c-a) \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

Chapter 7: Matrices

In this chapter we will learn about matrices and their properties. Just like determinants, matrices are also arrays of numbers or elements, but that is where the similarity between a determinant and a matrix ends. A determinant, by design, is a square array of numbers while a matrix, in general, is a rectangular array of numbers. With determinants, the elements are combined together to provide a single number, which is the value of the determinant. On the other hand, in the case of a matrix there is no concept of a single value of the matrix available by any combination of its elements. However, only for a square matrix, one can mention a corresponding determinant whose elements are the same as the elements of the square matrix, or one can loosely talk about determinant of a square matrix. As an application of matrices, in this chapter we will explore their use in solving n coupled inhomogeneous algebraic equations.

7.1 A GENERAL MATRIX

A general matrix is a rectangular array of numbers or elements. For example,

$$\mathbb{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

is a matrix with m rows and n columns. It is common to say that $\mathbb{A}$ is an $m \times n$ matrix. The element a_{ij} is located at the intersection of i^{th} row and j^{th} column of the matrix. If $\mathbb{A} = \mathbb{B}$, then all elements of matrix $\mathbb{A}$ are individually equal to all elements of matrix $\mathbb{B}$; that is, $a_{ij} = b_{ij}$ for all i and j. Clearly, if $\mathbb{A}$ is an $m \times n$ matrix, then so is $\mathbb{B}$. Matrices consisting of only one row or one column are referred to as vectors. For example, an $m \times 1$ matrix is a column vector of dimension m, and a $1 \times n$ matrix is a row vector of dimension n.

Various matrices combine (add, subtract or multiply) according to some well-defined rules. First, in order to add or subtract two matrices, they must be of the same size; that is, must have same number of rows and same number of columns. Furthermore, matrices add or subtract element-by-element. For example, if $\mathbb{C} = \mathbb{A} \pm \mathbb{B}$, then $c_{ij} = a_{ij} \pm b_{ij}$. Matrix addition is commutative as well as associative, that is,

$$\mathbb{A} + \mathbb{B} = \mathbb{B} + \mathbb{A} \text{ and } \mathbb{A} + (\mathbb{B} + \mathbb{C}) = (\mathbb{A} + \mathbb{B}) + \mathbb{C} \ .$$

Multiplication of two matrices, $\mathbb{A}$ and $\mathbb{B}$, leads to a new matrix $\mathbb{C}$, that is, $\mathbb{C} = \mathbb{A}\mathbb{B}$. The element c_{ij}, located at the intersection of i^{th} row and j^{th} column of matrix $\mathbb{C}$, is given by

$$c_{ij} = \sum_k a_{ik} b_{kj} \quad . \qquad Eq.(7.1)$$

In other words, element c_{ij} of $\mathbb{C}$ is obtained by multiplying, element-by-element, the i^{th} row of $\mathbb{A}$ with the j^{th} column of $\mathbb{B}$, and then adding all the products. Note that for the multiplication of two general rectangular matrices to have any meaning, the number of columns of $\mathbb{A}$ (namely, dummy index k in the sum in Eq. (7.1)) must be equal to the number of rows of $\mathbb{B}$ (again, the same dummy index k in the same sum). So, if $\mathbb{A}$ is an $m \times p$ matrix and $\mathbb{B}$ is a $p \times n$ matrix, then the product $\mathbb{C}$ is an $m \times n$ matrix, while the product $\mathbb{BA}$ is undefined. However, as long as matrix multiplication is allowed, it is both distributive and associative, that is,

$$\mathbb{A}(\mathbb{B} + \mathbb{C}) = \mathbb{AB} + \mathbb{AC} \ ,$$

$$(\mathbb{AB})\mathbb{C} = \mathbb{A}(\mathbb{BC}) \ .$$

The use of the word $vector$ for matrices with only one row or one column is justified since multiplying a 1×3 row matrix containing elements $A_x, A_y,$ and A_z with a 3×1 column matrix containing elements $B_x, B_y, and\ B_z$ is equivalent to taking the scalar or dot product of two vectors. Explicitly,

$$(A_x \quad A_y \quad A_z)\begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix} = A_x B_x + A_y B_y + A_z B_z = \mathbf{A \bullet B} \ .$$

Multiplication of a matrix $\mathbb{A}$ by a constant *k* simply multiplies each and every element of $\mathbb{A}$ by k; that is,

$$k\mathbb{A} = \begin{pmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{pmatrix} \ . \qquad Eq.(7.2)$$

If the number of rows and the number of columns of a matrix are equal, say to n, then the matrix is a $square$ matrix of order n. For two square matrices $\mathbb{A}$ and $\mathbb{B}$ of the same order, one can form the products $\mathbb{AB}$ as well as $\mathbb{BA}$. However, in general, $\mathbb{AB} \neq \mathbb{BA}$. If $\mathbb{AB} = \mathbb{BA}$, then the matrices $\mathbb{A}$ and $\mathbb{B}$ are said to $commute$.

Example: Consider two square matrices $\mathbb{A}$ and $\mathbb{B}$ of order 2,

$$\mathbb{A} = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{pmatrix} \qquad \mathbb{B} = \begin{pmatrix} \mathbf{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} .$$

Check whether the products $\mathbb{AB}$ and $\mathbb{BA}$ are equal or not.

Solution: The products $\mathbb{AB}$ and $\mathbb{BA}$ are

$$\mathbb{AB} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

$$\mathbb{BA} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Clearly, in this case $\mathbb{AB} \neq \mathbb{BA}$.

Example: Given two square matrices, $\mathbb{A} = \begin{pmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \\ \boldsymbol{\beta} & \boldsymbol{\alpha} \end{pmatrix}$ and $\mathbb{B} = \begin{pmatrix} \boldsymbol{\gamma} & \boldsymbol{\delta} \\ \boldsymbol{\delta} & \boldsymbol{\gamma} \end{pmatrix}$, determine whether they commute or not.

Solution: In this case,

$$\mathbb{AB} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}\begin{pmatrix} \gamma & \delta \\ \delta & \gamma \end{pmatrix} = \begin{pmatrix} \alpha\gamma + \beta\delta & \alpha\delta + \beta\gamma \\ \beta\gamma + \alpha\delta & \beta\delta + \alpha\gamma \end{pmatrix},$$

$$\mathbb{BA} = \begin{pmatrix} \gamma & \delta \\ \delta & \gamma \end{pmatrix}\begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} = \begin{pmatrix} \gamma\alpha + \delta\beta & \gamma\beta + \delta\alpha \\ \alpha\delta + \beta\gamma & \beta\delta + \gamma\alpha \end{pmatrix}.$$

Since $\mathbb{AB} = \mathbb{BA}$, then $\mathbb{A}$ and $\mathbb{B}$ commute.

A matrix with all of its elements as zero is called a null or a zero matrix and it is denoted by $\mathbb{0}$. We note in passing that, in general, if $\mathbb{AB} = \mathbb{0}$, it does not imply either $\mathbb{A} = \mathbb{0}$ or $\mathbb{B} = \mathbb{0}$. As an example, consider two matrices,

$$\mathbb{A} = \begin{pmatrix} 2 & 1 \\ -6 & -3 \end{pmatrix} \quad \mathbb{B} = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}.$$

Their product is

$$\mathbb{AB} = \begin{pmatrix} 2 & 1 \\ -6 & -3 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \mathbb{0}.$$

Note that the product $\mathbb{AB}$ is a zero matrix while neither $\mathbb{A}$ nor $\mathbb{B}$ is a zero matrix!

7.2 PROPERTIES OF MATRICES

From this point onwards, we will focus our attention only on square matrices. First, we define some special square matrices. A square matrix whose elements are all zero except the elements along the diagonal is called a $diagonal\ matrix$. For such a matrix $\mathbb{A}$, $a_{ij} = 0$ if $i \neq j$, and it looks like

$$\mathbb{A} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}. \qquad Eq.(7.3)$$

In particular, a square diagonal matrix which has only 1 as its diagonal element is a $unit\ or\ identity\ matrix\ \mathbb{1}$,

$$\mathbb{1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} . \qquad Eq.(7.4)$$

Thus, various elements of the unit matrix can be written in terms of a Kronecker delta as, $a_{ij} = \delta_{ij}$. Multiplying any general matrix $\mathbb{A}$, either on the left or on the right, by $\mathbb{1}$ gives back the matrix $\mathbb{A}$. So,

$$\mathbb{A}\mathbb{1} = \mathbb{1}\mathbb{A} = \mathbb{A} .$$

Transpose of a Matrix $(\widetilde{\mathbb{A}})$

A square matrix obtained by interchanging rows and columns of a matrix $\mathbb{A}$ is its transpose, $\widetilde{\mathbb{A}}$. (Another common notation for the transpose of a matrix which you may encounter is $\mathbb{A}^T$.) Thus,

$$\left(\widetilde{\mathbb{A}}\right)_{ij} = (\mathbb{A})_{ji} \text{ or } \tilde{a}_{ij} = a_{ji} .$$

Obviously, $(\widetilde{\widetilde{\mathbb{A}}}) = \mathbb{A}$. There is an interesting property of transpose of a product of two general square matrices. The transpose of a product of two matrices is equal to the product of transposes of the two matrices in reverse order, that is, if $\mathbb{C} = \mathbb{A}\mathbb{B}$ then $\tilde{\mathbb{C}} = \widetilde{\mathbb{B}}\widetilde{\mathbb{A}}$. As a proof, note

$$c_{ij} = \sum_k a_{ik} b_{kj} = \sum_k \tilde{a}_{ki}\, \tilde{b}_{jk} = \sum_k \tilde{b}_{jk} \tilde{a}_{ki} = \left(\widetilde{\mathbb{B}}\widetilde{\mathbb{A}}\right)_{ji} .$$

Also, $\tilde{c}_{ji} = c_{ij}$ by definition of transpose. Thus $\tilde{c}_{ji} = \left(\widetilde{\mathbb{B}}\widetilde{\mathbb{A}}\right)_{ji}$ or $\tilde{\mathbb{C}} = \widetilde{\mathbb{B}}\widetilde{\mathbb{A}}$.

Symmetric and Antisymmetric Matrices

A matrix $\mathbb{A}$ for which $\widetilde{\mathbb{A}} = \mathbb{A}$ is called a symmetric matrix since for such a matrix $a_{ij} = a_{ji}$. On the other hand, a matrix $\mathbb{A}$ for which $\widetilde{\mathbb{A}} = -\mathbb{A}$ is called an antisymmetric matrix since for such a matrix $a_{ij} = -a_{ji}$. It implies that the diagonal elements a_{ii} of an antisymmetric matrix are zero. Now, for any general square matrix $\mathbb{A}$, $\mathbb{A} + \widetilde{\mathbb{A}}$ is a symmetric matrix while $\mathbb{A} - \widetilde{\mathbb{A}}$ is an antisymmetric matrix. It is so because

$$\left(\mathbb{A} + \widetilde{\mathbb{A}}\right)_{ij} = a_{ij} + \tilde{a}_{ij} = \tilde{a}_{ji} + a_{ji} = \left(\widetilde{\mathbb{A}} + \mathbb{A}\right)_{ji}$$

$$\left(\mathbb{A} - \widetilde{\mathbb{A}}\right)_{ij} = a_{ij} - \tilde{a}_{ij} = \tilde{a}_{ji} - a_{ji} = -\left(\mathbb{A} - \widetilde{\mathbb{A}}\right)_{ji} .$$

Since, for any general matrix $\mathbb{A}$, we can write

$$\mathbb{A} = \frac{1}{2}\left(\mathbb{A} + \widetilde{\mathbb{A}}\right) + \frac{1}{2}\left(\mathbb{A} - \widetilde{\mathbb{A}}\right) \ ,$$

it means that any general matrix can be written as a sum of a symmetric and an antisymmetric matrix.

Inverse of a Matrix $\left(\mathbb{A}^{-1}\right)$

If $\mathbb{AB} = \mathbb{BA} = \mathbb{1}$ (unit matrix), then matrix $\mathbb{B}$ is the inverse of matrix $\mathbb{A}$ and matrix $\mathbb{A}$ is the inverse of matrix $\mathbb{B}$. Thus $\mathbb{B} = \mathbb{A}^{-1}$ and $\mathbb{A} = \mathbb{B}^{-1}$. Not every square matrix has an inverse; however, the inverse of a matrix, if it exists, is unique. As a proof of the uniqueness of the inverse of a matrix, let us suppose that there are two different matrices $\mathbb{B}$ and $\mathbb{C}$ which are inverse of the same matrix $\mathbb{A}$. Then $\mathbb{AB} = \mathbb{1}$ and $\mathbb{CA} = \mathbb{1}$. Now consider the triple product of matrices, $\mathbb{CAB}$. Using the associative nature of matrix multiplication, it can be rewritten as,

$$\mathbb{CAB} = (\mathbb{CA})\mathbb{B} = \mathbb{C}(\mathbb{AB}) \qquad \text{or} \qquad (\mathbb{1})\mathbb{B} = \mathbb{C}(\mathbb{1}) \qquad \text{or} \qquad \mathbb{B} = \mathbb{C} \ .$$

Thus, the inverse of $\mathbb{A}$ is unique.

There is an interesting property of the inverse of a product of two general square matrices. The inverse of a product of two matrices is equal to the product of inverses of the two matrices in reverse order; that is, if $\mathbb{C} = \mathbb{AB}$, then $\mathbb{C}^{-1} = \mathbb{B}^{-1}\mathbb{A}^{-1}$. As a proof, notice that

$$\mathbb{C}^{-1}\mathbb{C} = \mathbb{B}^{-1}\mathbb{A}^{-1}\mathbb{AB} = \mathbb{B}^{-1}\mathbb{1B} = \mathbb{1} \qquad \text{and} \qquad \mathbb{CC}^{-1} = \mathbb{ABB}^{-1}\mathbb{A}^{-1} = \mathbb{A1A}^{-1} = \mathbb{1} \ .$$

Since the inverse of a matrix is unique, $\mathbb{C}^{-1}$ is the true inverse of $\mathbb{C}$.

Adjoint of a Matrix $(\mathbb{A}^{a})$

In preparation for describing a technique for obtaining the inverse of a square matrix, we first define the adjoint of a square matrix. Note that for a square matrix $\mathbb{A}$ we may think of its determinant as the corresponding determinant whose elements are identical to the elements of $\mathbb{A}$ and whose value is $\det(\mathbb{A})$. Thus, if $\mathbb{A}$ is an $n \times n$ square matrix, its adjoint matrix is obtained by first replacing each element a_{ij} by the cofactor C_{ij} of the corresponding element in the determinant $\det(\mathbb{A})$, with appropriate sign, and then transposing the matrix. Specifically, adjoint of $\mathbb{A}$ is

$$\mathbb{A}^{a} = \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{pmatrix} ,$$

where C_{ij} is the cofactor of a_{ij}. From its construction, we infer $(\mathbb{A}^{a})_{ij} = C_{ji}$. Using Eq. (6.1), then

$$(\mathbb{A}\mathbb{A}^a)_{ij} = \sum_k (\mathbb{A})_{ik}(\mathbb{A}^a)_{kj} = \sum_k a_{ik}C_{jk} = \delta_{ij}\det(\mathbb{A}) \ .$$

or,

$$\mathbb{A}\mathbb{A}^a = \det(\mathbb{A})\ \mathbb{1} \ ,$$

where $\mathbb{1}$ is the unit matrix.

Similarly, using Eq. (6.1) again,

$$(\mathbb{A}^a\mathbb{A})_{ij} = \sum_k (\mathbb{A}^a)_{ik}(\mathbb{A})_{kj} = \sum_k C_{ki}a_{kj} = \delta_{ij}\det(\mathbb{A}) \ .$$

Thus,

$$\mathbb{A}\mathbb{A}^a = \mathbb{A}^a\mathbb{A} \ = \det(\mathbb{A})\mathbb{1} \ .$$

This can be rewritten as

$$\left(\frac{\mathbb{A}^a}{\det(\mathbb{A})}\right)\mathbb{A} = \mathbb{A}\left(\frac{\mathbb{A}^a}{\det(\mathbb{A})}\right) = \mathbb{1} \ ,$$

so that

$$\mathbb{A}^{-1} = \frac{\mathbb{A}^a}{det(\mathbb{A})} \ . \qquad Eq.(7.5)$$

Note that if $\det(\mathbb{A}) = 0$ then $\mathbb{A}^{-1}$ is undefined. In other words, the condition for $\mathbb{A}^{-1}$ to exist is that $\det(\mathbb{A}) \neq 0$. A matrix $\mathbb{A}$ for which $\det(\mathbb{A}) \neq 0$ is referred to as a nonsingular matrix while a matrix for which $\det(\mathbb{A}) = 0$ is a singular matrix. Thus, the inverse of a singular matrix does not exist.

Example: Find the inverse of the matrix

$$\mathbb{A} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{2} \\ \mathbf{2} & \mathbf{-1} & \mathbf{3} \\ \mathbf{0} & \mathbf{2} & \mathbf{3} \end{pmatrix}$$

using the adjoint method.

Solution: First, we note that $\det\mathbb{A} = +1(-9) + 2(4) = -1 \neq 0$, so that the matrix $\mathbb{A}$ is nonsingular and its inverse exists.

The cofactors of various matrix elements are

$$C_{11} = -9, C_{12} = -6, C_{13} = +4 \ ,$$

$$C_{21} = +4, C_{22} = +3, C_{23} = -2 \ ,$$

$$C_{31} = +2, C_{32} = +1, C_{33} = -1 \ .$$

The adjoint matrix is

$$\mathbb{A}^a = \begin{pmatrix} -9 & 4 & 2 \\ -6 & 3 & 1 \\ 4 & -2 & -1 \end{pmatrix} .$$

Finally, the inverse of $\mathbb{A}$ is

$$\mathbb{A}^{-1} = \frac{\mathbb{A}^a}{\det(\mathbb{A})} = \begin{pmatrix} 9 & -4 & -2 \\ 6 & -3 & -1 \\ -4 & 2 & 1 \end{pmatrix} .$$

Just to verify that our inverse is correct, we multiply the original matrix $\mathbb{A}$ by its inverse matrix $\mathbb{A}^{-1}$ and make sure that we get the identity matrix, $\mathbb{1}$.

$$\mathbb{A}^{-1}\mathbb{A} = \begin{pmatrix} 9 & -4 & -2 \\ 6 & -3 & -1 \\ -4 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 0 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbb{1} \ .$$

Example: Given the matrix

$$\mathbb{A} = \begin{pmatrix} \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \mathbf{2} & \mathbf{3} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{2} \end{pmatrix} ,$$

find $\mathbb{A}^{-1}$ using the adjoint method.

Solution: The determinant of this matrix is $\det \mathbb{A} = +1(6) - 2(1) = 4 \neq 0,$ hence the matrix is nonsingular. The cofactors of various elements a_{ij} are

$$C_{11} = +6, C_{12} = -4, C_{13} = +2 \ ,$$

$$C_{21} = -1, C_{22} = +2, C_{23} = -1 \ ,$$

$$C_{31} = -9, C_{32} = +6, C_{33} = -1 \ .$$

The adjoint matrix is

$$\mathbb{A}^a = \begin{pmatrix} 6 & -1 & -9 \\ -4 & 2 & 6 \\ 2 & -1 & -1 \end{pmatrix},$$

and the inverse matrix is

$$\mathbb{A}^{-1} = \frac{\mathbb{A}^a}{\det(\mathbb{A})} = \frac{1}{4}\begin{pmatrix} 6 & -1 & -9 \\ -4 & 2 & 6 \\ 2 & -1 & -1 \end{pmatrix}.$$

7.3 SOLVING INHOMOGENEOUS ALGEBRAIC EQUATIONS

The idea of matrix inversion is very useful in solving a set of algebraic equations. Consider the following set of n algebraic equations containing n unknown variables $x_1, x_2, \cdots, x_n$,

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = y_1 \ ,$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = y_2 \ ,$$

$$\vdots$$

$$a_{n1}x_1 + a_{n2x_2} + \cdots + a_{nn}x_n = y_n \ .$$

Here all the coefficients a_{ij} as well as the numbers on the right-hand sides of these equations, namely y_n, are known. If all the numbers y_n are zero, then these algebraic equations are called $homogeneous$ algebraic equations. On the other hand, if all or some of the numbers y_n are nonzero, then these algebraic equations are called $inhomogeneous$ algebraic equations. The equations can be rewritten in a matrix form as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

or

$$\mathbb{A}\mathbb{X} = \mathbb{Y} \ , \qquad Eq.(7.6a)$$

where $\mathbb{X}$ and $\mathbb{Y}$ are column matrices (or vectors), $\mathbb{Y}$ is known and $\mathbb{X}$ is unknown. Then, using the idea of matrix inversion,

$$\mathbb{X} = \mathbb{A}^{-1}\mathbb{Y} = \frac{\mathbb{A}^a}{\det(\mathbb{A})}\mathbb{Y} \ . \qquad Eq.(7.6b)$$

Example: Solve the following three inhomogeneous algebraic equations of three variables, $x, y, \text{and } z$ using the inverse of a matrix.

$$\boldsymbol{x + 2y + 3z = 4} \ ,$$

$$\boldsymbol{2x + 3y = 8} \ ,$$

$$\boldsymbol{y + 2z = 12} \ .$$

Solution: Rewriting these equations in a matrix form, we get

$$\mathbb{AX} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \\ 12 \end{pmatrix} = \mathbb{Y} \ .$$

The inverse of this matrix $\mathbb{A}$ was determined in the previous example. Using it, we get

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbb{A}^{-1} \begin{pmatrix} 4 \\ 8 \\ 12 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 6 & -1 & -9 \\ -4 & 2 & +6 \\ 2 & -1 & -1 \end{pmatrix} \begin{pmatrix} 4 \\ 8 \\ 12 \end{pmatrix}$$

$$= \begin{pmatrix} 6 & -1 & -9 \\ -4 & 2 & +6 \\ 2 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -23 \\ 18 \\ -3 \end{pmatrix} .$$

Thus, $x = -23, y = 18 \text{ and } z = -3$.

In case of homogenous algebraic equations, $\mathbb{Y} = 0$. Then, if $\det(\mathbb{A}) \neq 0$, we get the solution $\mathbb{X} = 0$, which is called the trivial solution. A nontrivial solution of homogenous algebraic equations is available only when $\det(\mathbb{A}) = 0$.

7.4 SPECIAL MATRICES

Now, we introduce some well-known and useful matrices. An $orthogonal\ matrix\ \mathbb{A}$ is the one that satisfies

$$\mathbb{A}\widetilde{\mathbb{A}} = \widetilde{\mathbb{A}}\mathbb{A} = \mathbb{1} \ .$$

Clearly, for the orthogonal matrix,

$$\widetilde{\mathbb{A}} = \mathbb{A}^{-1} \ .$$

As an example of an orthogonal matrix, consider

$$\mathbb{A} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} .$$

For this matrix

$$\widetilde{\mathbb{A}} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

and

$$\mathbb{A}\,\widetilde{\mathbb{A}} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} = \mathbb{1}$$

as well as

$$\widetilde{\mathbb{A}}\,\mathbb{A} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}\begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} = \mathbb{1} .$$

So, we conclude that $\widetilde{\mathbb{A}} = \mathbb{A}^{-1}$ and $\mathbb{A}$ is an orthogonal matrix.

If the elements of a matrix consist of some complex numbers, then we can obtain the $complex\ conjugate\ matrix$, $\mathbb{A}^*$, by simply replacing each element of matrix $\mathbb{A}$ by its complex conjugate. Thus, if $(\mathbb{A})_{ij} = a_{ij}$, then $(\mathbb{A}^*)_{ij} = a_{ij}^*$. As an example, consider

$$\mathbb{A} = \begin{pmatrix} 2 & 1+i \\ 2-i & 3 \end{pmatrix} .$$

Its complex conjugate matrix is

$$\mathbb{A}^* = \begin{pmatrix} 2 & 1-i \\ 2+i & 3 \end{pmatrix} .$$

Now, we will introduce $Hermitian$ and $unitary\ matrices$. First, we define the Hermitian adjoint of a general matrix $\mathbb{A}$. It is denoted by the symbol † (called dagger). The Hermitian adjoint is obtained by first taking the complex conjugate of the matrix, followed by taking the transpose of the matrix. Thus,

$$\mathbb{A}^\dagger = \widetilde{\mathbb{A}^*}$$

is the definition of the Hermitian adjoint of $\mathbb{A}$. Now, a matrix $\mathbb{A}$ is a Hermitian matrix if its Hermitian adjoint is equal to the matrix itself. In other words, if

$$\mathbb{A}^\dagger = \mathbb{A} ,$$

then $\mathbb{A}$ is a Hermitian matrix.

Next, a matrix $\mathbb{A}$ is a unitary matrix if its Hermitian adjoint is equal to the inverse, $\mathbb{A}^{-1}$, of the matrix. In other words, if $\mathbb{A}^{\dagger} = \mathbb{A}^{-1}$, or if

$$\mathbb{A}\mathbb{A}^{\dagger} = \mathbb{A}^{\dagger}\mathbb{A} = \mathbb{1} \ ,$$

then $\mathbb{A}$ is a unitary matrix.

As an example, consider the matrix

$$\mathbb{A} = \frac{1}{\sqrt{a^2 + b^2 + c^2}} \begin{pmatrix} a & b - ic \\ b + ic & -a \end{pmatrix} \ ,$$

with $a, b, \text{and } c$ real. One can even visually see that $\mathbb{A}^{\dagger} = \mathbb{A}$ for this matrix, so that it is a Hermitian matrix. Furthermore, one can easily check that this matrix satisfies

$$\mathbb{A}\mathbb{A}^{\dagger} = \mathbb{A}^{\dagger}\mathbb{A} = \mathbb{1} \ ,$$

so that this matrix is also a unitary matrix.

7.5 CHARACTERISTIC OR SECULAR EQUATION OF A MATRIX

If $\mathbb{A}$ is an $n \times n$ matrix and λ is a scalar parameter, then

$$\mathbb{K} = \mathbb{A} - \lambda\mathbb{1}$$

is called the characteristic matrix of $\mathbb{A}$. The equation

$$\det(\mathbb{K}) = \det(\mathbb{A} - \lambda\mathbb{1}) = 0$$

is called the characteristic or secular equation of $\mathbb{A}$. If $\mathbb{A}$ is an $n \times n$ matrix, then the secular equation is of the form

$$f(\lambda) = 1 + a_1\lambda + a_2\lambda^2 + \cdots + a_n\lambda^n = 0$$

where the coefficients $a_i (i = 1, \ldots n)$ depend on the elements of $\mathbb{A}$. The n roots of this algebraic equation $(\lambda_1, \lambda_2, \ldots\ldots, \lambda_n)$ are called the characteristic roots of the matrix $\mathbb{A}$.

Eigenvalues/Eigenvectors of Matrices

Given a $n \times n$ matrix $\mathbb{A}$, there exist certain column matrices (or vectors) $\mathbb{X}$ such that $\mathbb{A}\mathbb{X}$ is a simple multiple of $\mathbb{X}$. In other words,

$$\mathbb{A}\mathbb{X} = \lambda\mathbb{X} \qquad Eq.(7.7)$$

where λ is a number (real or complex). Then λ is said to be an eigenvalue of $\mathbb{A}$ belonging to the eigenvector $\mathbb{X}$. Here $\mathbb{X}$ is a column matrix with n rows,

$$\mathbb{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} .$$

Thus, Eq. (7.7) gives

$$\sum_{j=1}^{n} a_{ij}x_j = \lambda x_i \ .$$

Or,

$$\sum_{j=1}^{n} (a_{ij} - \lambda\delta_{ij})x_j = 0 \ .$$

These are n homogenous algebraic equations, for $i = 1, 2, \dots n$, in variables $x_1, x_2, \dots, x_n$. A nontrivial solution exists for these equations only if

$$\det(\mathbb{A} - \lambda\mathbb{1}) = 0 \ .$$

This, of course, is the secular equation of matrix $\mathbb{A}$. The characteristic roots of $\mathbb{A}$ are, then, the eigenvalues of $\mathbb{A}$, and the corresponding vectors $\mathbb{X}$ are the eigenvectors. We briefly note that eigenvalues and eigenvectors are quite useful in quantum physics and this fact will be mentioned again in Chapter 13.

Example: Find the eigenvalues and eigenvectors of the matrix

$$\mathbb{A} = \begin{pmatrix} 0 & 3 & 4 \\ 3 & 0 & 0 \\ 4 & 0 & 0 \end{pmatrix} .$$

Solution: The characteristic equation is $\det(\mathbb{A} - \lambda\mathbb{1}) = 0$, or

$$\det\begin{pmatrix} -\lambda & 3 & 4 \\ 3 & -\lambda & 0 \\ 4 & 0 & -\lambda \end{pmatrix} = 0 \ .$$

On expanding the determinant, we get

$$-\lambda^3 + 9\lambda + 16\lambda = 0 \qquad \text{or} \qquad \lambda(\lambda + 5)(\lambda - 5) = 0 \ .$$

Or,

$$\lambda = 0, +5, -5 \ .$$

These are the three eigenvalues of matrix $\mathbb{A}$. Next, we determine the eigenvectors one-by-one. For $\lambda = 0$, $\mathbb{A}\mathbb{X} = \lambda\mathbb{X}$ gives,

$$\begin{pmatrix} 0 & 3 & 4 \\ 3 & 0 & 0 \\ 4 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} .$$

Or,

$$3x_2 + 4x_3 = 0, \qquad 3x_1 = 0, \qquad 4x_1 = 0 \ .$$

Thus, within an arbitrary constant, the eigenvector is

$$\begin{pmatrix} 0 \\ 4 \\ -3 \end{pmatrix} .$$

A *normalized* eigenvector is analogous to a unit vector since they both have magnitudes of 1. To normalize any eigenvector, we simply divide it by the magnitude of the eigenvector. So, for $\lambda = 0$, the magnitude of the eigenvector is $\sqrt{0 + 4^2 + (-3)^2} = 5$, and, therefore, the normalized eigenvector is

$$\frac{1}{5}\begin{pmatrix} 0 \\ 4 \\ -3 \end{pmatrix} .$$

For $= +5$, $\mathbb{A}\mathbb{X} = \lambda\mathbb{X}$ gives

$$\begin{pmatrix} 0 & 3 & 4 \\ 3 & 0 & 0 \\ 4 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = +5 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} .$$

Or,

$$-5x_1 + 3x_2 + 4x_3 = 0$$

$$+3x_1 - 5x_2 = 0 \qquad \text{or} \qquad x_2 = \frac{3}{5}\, x_1$$

$$+4x_1 - 5x_3 = 0 \qquad \text{or} \qquad x_3 = \frac{4}{5}x_1 \ .$$

Again, within an arbitrary constant, the eigenvector is

$$\begin{pmatrix} 5 \\ 3 \\ 4 \end{pmatrix},$$

and the normalized eigenvector is $\frac{1}{5\sqrt{2}}\begin{pmatrix} 5 \\ 3 \\ 4 \end{pmatrix}$.

Finally, for $\lambda = -5,\ \mathbb{A}\mathbb{X} = \lambda\mathbb{X}$ gives

$$\begin{pmatrix} 0 & 3 & 4 \\ 3 & 0 & 0 \\ 4 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = -5\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} .$$

Or

$$+5x_1 + 3x_2 + 4x_3 = 0$$

$$+3x_1 + 5x_2 = 0 \qquad \text{or} \qquad x_2 = -\frac{3}{5}\ x_1$$

$$+4x_1 + 5x_3 = 0 \qquad \text{or} \qquad x_3 = -\frac{4}{5}x_1 \ .$$

So, the eigenvector is, within an arbitrary constant,

$$\begin{pmatrix} 5 \\ -3 \\ -4 \end{pmatrix},$$

and the normalized eigenvector is $\frac{1}{5\sqrt{2}}\begin{pmatrix} 5 \\ -3 \\ -4 \end{pmatrix}$.

PROBLEMS FOR CHAPTER 7

1. Given the following two matrices $\mathbb{A}$ and $\mathbb{B}$, determine $\mathbb{AB}$, $\mathbb{BA}$, $\mathbb{A}^2$, and $\mathbb{B}^2$ if they exist. If a product does not exist, make a note of it.

$$\mathbb{A} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \\ 3 & 2 \end{pmatrix} \qquad \text{and} \qquad \mathbb{B} = \begin{pmatrix} 3 & -1 \\ 4 & -3 \end{pmatrix} .$$

2. Given the three matrices,

$$\mathbb{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbb{B} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbb{C} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} ,$$

find all possible products of pairs of matrices (either $\mathbb{A}$ and $\mathbb{B}$, or $\mathbb{B}$ and $\mathbb{C}$, or $\mathbb{C}$ and $\mathbb{A}$) to determine which pair, or pairs, of matrices commute.

3. For the following matrix $\mathbb{A}$, determine $\mathbb{A}^{-1}$,

$$\mathbb{A} = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 3 & 0 \\ 2 & 0 & 3 \end{pmatrix} .$$

Explicitly verify that $\mathbb{A}\mathbb{A}^{-1} = \mathbb{A}^{-1}\mathbb{A} = \mathbb{1}$.

4. Determine the eigenvalues and the normalized eigenfunctions of matrix $\mathbb{A}$ of problem 3.

5. Solve the following inhomogeneous algebraic equations by finding the inverse of the matrix of coefficients,

$$x - y + z = 4$$

$$2x + y - z = -1$$

$$3x + 2y + 2z = 5 \ .$$

6. In the theory of spin $\frac{1}{2}$ in quantum physics, we encounter three 2×2 matrices ($Pauli\ spin\ matrices$),

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

Prove the following properties of these matrices:

(a) $\sigma_l \sigma_m = i \sum_{n=1}^{3} \epsilon_{lmn} \sigma_n \quad$ for $l \neq m$.

(b) $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, where $\mathbb{1}$ is the unit, or identity, matrix of size 2×2.

7. Determine the eigenvalues and the normalized eigenvectors of the three Pauli spin matrices of problem 6, separately.

Chapter 8: Vector Analysis

In biomedical physics we encounter several physical quantities. Some of these quantities are characterized by their magnitude only, such as temperature of a sick patient, density of a blood sample, pressure in the lung, etc. Such quantities are called scalars and they are treated as ordinary algebraic numbers. On the other hand, there are physical quantities which are characterized by both magnitude and direction, such as forces on human spine during heavy lifting, velocity of blood flowing in veins, electric and magnetic fields in medical equipment, etc. Such quantities are called vectors. In this chapter we will learn about mathematical properties of vectors.

8.1 VECTORS AND THEIR ADDITION/SUBTRACTION

A vector quantity is represented by a directed line whose length is proportional to the magnitude of the physical quantity and whose direction is the direction of the physical quantity. These directed lines, representing vectors, look like an arrow with one end as the head of the arrow and the other end as the tail of the arrow, as shown in Figure 8.1. Vectors are shown as bold-faced letters like **A** and their magnitudes are shown as ordinary letters

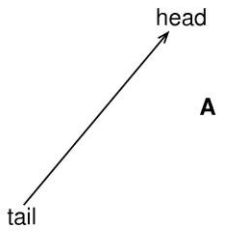


Figure 8.1. A vector, A, with its head and its tail.

like A. We can move a vector, or the directed line representing the vector, around in space and the vector will remain unchanged as long as both the length of the directed line as well as its direction are kept unchanged.

An alternate way of representing a vector is through its components. If the Cartesian coordinates of the head and of the tail of a vector **A** are (x_h, y_h, z_h) and (x_t, y_t, z_t), respectively, then, the components of **A** are

$$A_x = x_h - x_t \ ,$$

$$A_y = y_h - y_t \ ,$$

$$A_z = z_h - z_t \ .$$

The length of the vector is

$$\mathbf{A} = \left[A_x^2 + A_y^2 + A_z^2\right]^{1/2} \quad . \qquad Eq.(8.1a)$$

It is easier to see that if the origin of the coordinate system is chosen at the tail of the vector **A**, then $A_x, A_y, \text{and } A_z$ are merely the projections of **A** onto the three coordinate axes.

Addition and Subtraction of Vectors

Vectors do not add or subtract or multiply like ordinary numbers. First, we will learn the rules for adding and subtracting vectors, which lead to new vectors. Next, we will learn the rules for multiplying vectors, which can lead to either a scalar (scalar product) or a vector (vector product). The vectors can be added using a geometrical method, called head-to-tail method, in which we connect the head of the first vector, **A**, with the tail of the second vector, **B**. A new vector **X** whose head is the free unconnected head of **B** and whose tail is the free unconnected tail of **A** is the sum of vectors **A** and **B**, as shown in the Figure 8.2

$$\mathbf{X} = \mathbf{A} + \mathbf{B} \quad .$$

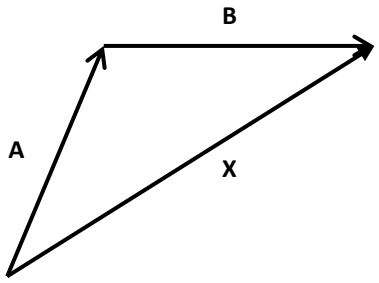


Figure 8.2. Vector **X** is the sum of vectors **A** and **B**.

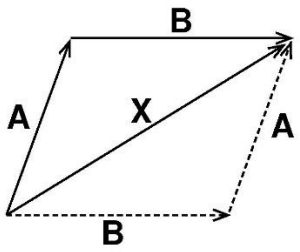


Figure 8.3. Commutative nature of vector addition.

By completing the parallelogram, as in Figure 8.3, it is seen that

$$\mathbf{X} = \mathbf{B} + \mathbf{A}$$

as well. Thus, vector addition is commutative. Addition of more than two vectors proceeds in the same manner. If $\mathbf{Z} = \mathbf{A} + \mathbf{B} + \mathbf{C}$, then from the Figure 8.4

$$\mathbf{X} = \mathbf{A} + \mathbf{B} \ ,$$

$$\mathbf{Y} = \mathbf{B} + \mathbf{C} \ ,$$

and

$$\mathbf{Z} = \mathbf{X} + \mathbf{C} = \mathbf{A} + \mathbf{Y} \ ,$$

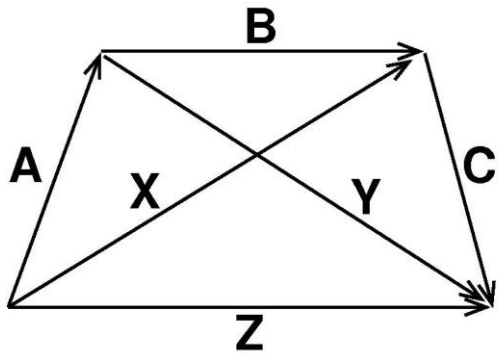


Figure 8.4. Sum of more than two vectors.

that is,

$$\mathbf{Z} = (\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \ .$$

Thus, vector addition is associative as well. Now, let us talk about subtraction of one vector from another. Figure 8.5 shows two vectors, **A** and **B**, of equal length but with opposite directions. According to head-to-tail method of adding vectors, $\mathbf{A} + \mathbf{B} = \mathbf{0}$ or $\mathbf{B} = -\mathbf{A}$.

In other words, $-\mathbf{A}$ (or **B**) is a vector which has same magnitude as **A** but its direction is opposite to that of **A**. Subtraction of one vector from another can now be handled exactly in the same manner as vector addition except the sign (direction) of one of the vectors is reversed. For example,

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}) \ .$$

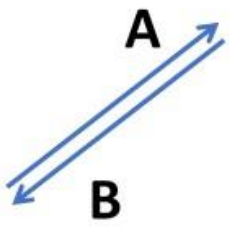


Figure 8.5. Two vectors **A** and **B** of equal length, but opposite directions.

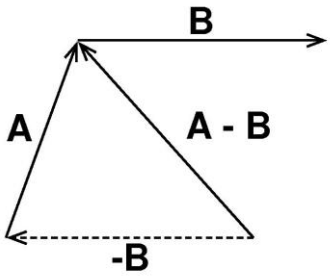


Figure 8.6. Subtraction of two vectors.

Also, graphically the "other" diagonal of the parallelogram in Figure 8.3 gives **A** − **B** as shown in Figure 8.6.

An alternative way of adding vectors is by using the components of the vector. For this purpose, we choose a right-handed Cartesian coordinate system. The meaning of a right-handed coordinate system will be clearer after Eq. (8.6). For this coordinate system, we introduce unit vectors $\mathbf{e_x}, \mathbf{e_y}, \text{and } \mathbf{e_z}$. These three vectors are each of unit length and are directed along the x, y, and z axes, respectively. We note in passing that several other authors use the notation $\mathbf{i}, \mathbf{j}, \text{and } \mathbf{k}$ or $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \text{and } \hat{\mathbf{z}}$ for these Cartesian unit vectors. In our notation, then $A_x\mathbf{e_x}$ is a vector of length A_x along the x-axis, $A_y\mathbf{e_y}$ is a vector of length A_y along the y-axis, and $A_z\mathbf{e_z}$ is a vector of length A_z along the z-axis. Using vector addition by head-to-tail method, as seen in Figure 8.7,

$$\mathbf{A} = A_x\mathbf{e_x} + A_y\mathbf{e_y} + A_z\mathbf{e_z} \ . \qquad Eq.(8.1b)$$

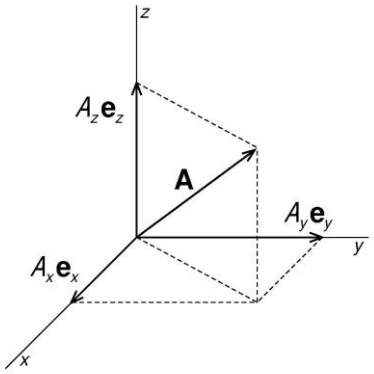


Figure 8.7. Components of a vector **A**.

Now, two vectors **A** and **B** can be added simply by adding their components, that is,

$$\begin{aligned}\mathbf{A} + \mathbf{B} &= \left(\mathbf{e_x}A_x + \mathbf{e_y}A_y + \mathbf{e_z}A_z\right) + \left(\mathbf{e_x}B_x + \mathbf{e_y}B_y + \mathbf{e_z}B_z\right) \\ &= \mathbf{e_x}(A_x + B_x) + \mathbf{e_y}(A_y + B_y) + \mathbf{e_z}(A_z + B_z) \ . \qquad Eq.(8.2a)\end{aligned}$$

Subtraction of one vector from another can now be handled exactly in the same manner as vector addition except sign (direction) of one of the vectors is reversed. For example,

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}) = \mathbf{e_x}(A_x - B_x) + \mathbf{e_y}\big(A_y - B_y\big) + \mathbf{e_z}(A_z - B_z)\ . \qquad Eq.(8.2b)$$

It follows from the discussion of addition and subtraction of vectors that if two vectors are equal to each other, then individually the x, y, and z components of those two vectors are also equal.

8.2 MULTIPLICATION OF VECTORS: SCALAR PRODUCT

Now, we talk about multiplying two vectors. There are two ways to multiply vectors. One way leads to a scalar and the other to a vector. The scalar (or inner or dot) product of two vectors **A** and **B** is a scalar number of magnitude AB $\cos\theta$, where A and B are the magnitudes of the two vectors and θ is the angle between the two vectors. This angle is obtained by joining the tails of both vectors as in Figure 8.8. The scalar product of two vectors is indicated by placing a large dot between the two vectors, **A•B** = AB $\cos\theta$ = **B•A**. Note that the scalar product of two vectors is commutative. The scalar product is also distributive, that is,

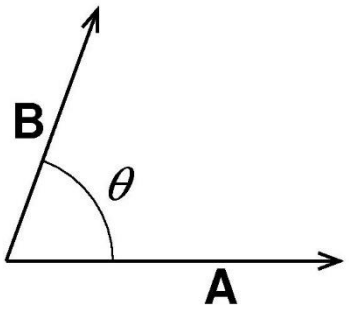


Figure 8.8. Scalar product of two vectors **A** and **B**.

A•(B + C) = A•B + A•C. Also, since unit vectors $\mathbf{e_x}, \mathbf{e_y}, \mathbf{e_z}$ are mutually orthogonal, it follows that.

$$\mathbf{e_x} \bullet \mathbf{e_x} = 1, \qquad \mathbf{e_y} \bullet \mathbf{e_y} = 1, \quad \mathbf{e_z} \bullet \mathbf{e_z} = 1$$

and

$$\mathbf{e_x} \bullet \mathbf{e_y} = \mathbf{e_y} \bullet \mathbf{e_x} = 0, \mathbf{e_x} \bullet \mathbf{e_z} = \mathbf{e_z} \bullet \mathbf{e_x} = 0, \mathbf{e_y} \bullet \mathbf{e_z} = \mathbf{e_z} \bullet \mathbf{e_y} = 0\ .$$

In particular, note

$$\mathbf{A \bullet B} = \left(\mathbf{e_x}A_x + \mathbf{e_y}A_y + \mathbf{e_z}A_z\right) \bullet \left(\mathbf{e_x}B_x + \mathbf{e_y}B_y + \mathbf{e_z}B_z\right)$$

$$= +A_x B_x(\mathbf{e_x} \bullet \mathbf{e_x}) + A_y B_x(0) + A_z B_x(0)$$

$$+A_x B_y(0) + A_y B_y(\mathbf{e_y} \bullet \mathbf{e_y}) + A_z B_y(0)$$

$$+A_x B_z(0) + A_y B_z(0) + A_z B_z(\mathbf{e_z} \bullet \mathbf{e_z})$$

or

$$\mathbf{A} \bullet \mathbf{B} = A_x B_x + A_y B_y + A_z B_z \ . \qquad Eq.(8.3)$$

Also,

$$\mathbf{A} \bullet \mathbf{A} = A_x^2 + A_y^2 + A_z^2$$

so that $\sqrt{\mathbf{A} \bullet \mathbf{A}}$ gives the magnitude of vector **A**.

Now we introduce an alternate, but useful, new notation. We will use $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$ for the three orthogonal unit vectors $\mathbf{e_x}, \mathbf{e_y}, \text{and } \mathbf{e_z}$, and $A_1, A_2, \text{and } A_3$ for the three components $A_x, A_y, \text{and } A_z$ of a vector **A**. In this notation it follows that

$$\mathbf{e_i} \bullet \mathbf{e_j} = \delta_{ij} \ . \qquad Eq.(8.4)$$

Using this identity, we can write the scalar product of two vectors in terms of a Kronecker delta as

$$\mathbf{A} \bullet \mathbf{B} = \left(\sum_i \boldsymbol{e_i} A_i\right) \bullet \left(\sum_j \boldsymbol{e_j} B_j\right) = \sum_{i,j} \delta_{ij} A_i B_j = \sum_i A_i B_i \ .$$

Finally, the law of cosines in trigonometry, which relates the lengths of three sides of a triangle, can be easily obtained using scalar product of two vectors. Three vectors **A, B,** and **C,** form the sides of a triangle as shown in the Figure 8.9. The angle θ between the two vectors **A** and **B** is obtained by joining the tails of both vectors. Using head-to-tail vector addition,

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \ ,$$

or

$$C^2 = \mathbf{C} \bullet \mathbf{C} = (\mathbf{A} + \mathbf{B}) \bullet (\mathbf{A} + \mathbf{B}) = \mathbf{A} \bullet \mathbf{A} + \mathbf{B} \bullet \mathbf{B} + 2\, \mathbf{A} \bullet \mathbf{B} = A^2 + B^2 + 2\, AB\cos\theta$$

or

$$C^2 = A^2 + B^2 - 2\, AB\cos\phi \qquad Eq.(8.5)$$

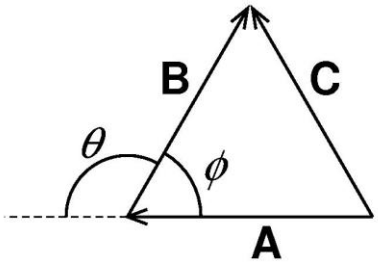


Figure 8.9. Law of cosines.

where ϕ is the angle inside the triangle opposite to side **C**, that is, the angle supplementary to θ. This is the law of cosines in trigonometry, and it is consistent with the Pythagorean theorem for a right-angled triangle (when $\phi = \pi/2$).

8.3 MULTIPLICATION OF VECTORS: VECTOR PRODUCT

Now, we talk about the second way of multiplying two vectors which leads to a vector. The vector (or cross) product of two vectors is indicated by placing a large cross between the two vectors **A** and **B**. It is defined to be a new vector **C**,

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} \ ,$$

such that the magnitude of **C** is $C = ABsin\theta$ and direction of **C** is perpendicular to the plane containing **A** and **B** in such a sense that **A**, **B**, and **C** form a right-handed system. In order to understand the direction of **C** better, let us imagine that the plane containing **A** and **B** is a big wall. We place a screwdriver, with a screw, against the wall such that the screwdriver can be rotated either clockwise or counterclockwise and the screw itself will move either into the wall or out of the wall. In the Figure 8.10 (a), the screwdriver is rotated counterclockwise, as shown by the arrow on angle θ, from **A** to **B,** and the direction of **A** $\times$ **B,** which is along the direction of motion of the screw itself, is coming out of the wall, indicated by symbol $\odot$. In Figure 8.10 (b), the screwdriver is rotated clockwise along the arrow shown on angle θ, from **B** to **A,** and the direction of **B** $\times$ **A**, or the direction of motion of the screw itself, is going into the wall, indicated by symbol $\otimes$. This is referred to as *the right-hand rule*. With this definition, the vector **B** $\times$ **A** has the same magnitude as vector **A** $\times$ **B**; however, direction of **B** $\times$ **A** is opposite to that of **A** $\times$ **B**. Thus,

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \ ,$$

that is, the cross product is anticommuting. In the case of the three unit vectors $\mathbf{e_1}, \mathbf{e_2}$, and $\mathbf{e_3}$,

$$\mathbf{e_1} \times \mathbf{e_1} = 0\ , \qquad \mathbf{e_2} \times \mathbf{e_2} = 0\ , \qquad \mathbf{e_3} \times \mathbf{e_3} = 0\ ,$$

$$\mathbf{e_1} \times \mathbf{e_2} = \mathbf{e_3} = -\mathbf{e_2} \times \mathbf{e_1}, \mathbf{e_2} \times \mathbf{e_3} = \mathbf{e_1} = -\mathbf{e_3} \times \mathbf{e_2}\ , \qquad \mathbf{e_3} \times \mathbf{e_1} = \mathbf{e_2} = -\mathbf{e_1} \times \mathbf{e_3}\ .$$

In a more compact form, we can write all of these nine relationships as

$$\mathbf{e_i} \times \mathbf{e_j} = \sum_{k=1}^{3} \epsilon_{ijk}\, \mathbf{e_k}\,. \qquad Eq.(8.6)$$

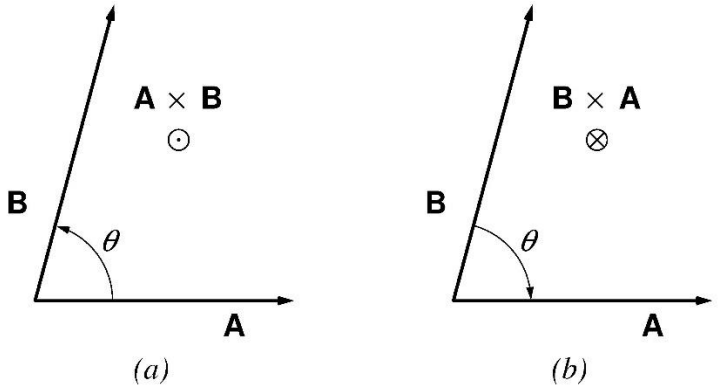


Figure 8.10. Vector product of two vectors **A** and **B**.

The three unit vectors, satisfying Eq. (8.6), describe a right-handed Cartesian coordinate system. Using these relations, the components of $\mathbf{C} = \mathbf{A} \times \mathbf{B}$ can be written as

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} = (\mathbf{e_1}A_1 + \mathbf{e_2}A_2 + \mathbf{e_3}A_3) \times (\mathbf{e_1}B_1 + \mathbf{e_2}B_2 + \mathbf{e_3}B_3)$$

$$= \mathbf{e_3}A_1B_2 - \mathbf{e_2}A_1B_3 - \mathbf{e_3}A_2B_1 + \mathbf{e_1}A_2B_3 + \mathbf{e_2}A_3B_1 - \mathbf{e_1}A_3B_2$$

$$= \mathbf{e_1}(A_2B_3 - A_3B_2) + \mathbf{e_2}(A_3B_1 - A_1B_3) + \mathbf{e_3}(A_1B_2 - A_2B_1)\ .$$

Thus,

$$C_1 = A_2B_3 - A_3B_2$$

$$C_2 = A_3B_1 - A_1B_3$$

$$C_3 = A_1B_2 - A_2B_1$$

or

$$C_i = A_jB_k - A_kB_j$$

with $i, j, \text{and } k$ all different and in cyclic permutation. In terms of Levi-Civita symbol, we can write the components of **C** as

$$C_i = \sum_{j,k} \epsilon_{ijk} A_j B_k \ ,$$

and the vector **C** as

$$\mathbf{C} = \sum_i \mathbf{e_i} C_i \ ,$$

or

$$\mathbf{A} \times \mathbf{B} = \sum_{i,j,k} \epsilon_{ijk} \mathbf{e_i} A_j B_k \ . \qquad Eq.(8.7)$$

We note from this suggestive form that the vector product can be written in the form of a determinant as

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} = \begin{vmatrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix} \ . \qquad Eq.(8.8)$$

The cross product of two vectors has a very nice geometrical interpretation. Consider a parallelogram, in a plane, whose sides are made up by vectors **A** and **B**, as shown in Figure 8.11. Note that the area of this parallelogram is $AB \sin\theta$, which is the magnitude of **A** $\times$ **B**. However, the direction of **A** $\times$ **B** is perpendicular to the plane of the parallelogram. Thus, in general, the area of an element such as the parallelogram defined above can be treated as a vector with its direction perpendicular to the plane containing that element. Note that the direction of area vector, either into or out of the plane, is related to the sense of traversal of the periphery of the area by the right-hand rule. In Figure 8.11, the direction of the area element **A** $\times$ **B**, of magnitude AB$\sin\theta$, is obtained by traversing the periphery by going first over (solid line) vector **A** and then over (solid line) vector **B**, that is, traversing the periphery indicated by the thick counterclockwise arrow corresponding to direction coming out of the plane of parallelogram according to the right-hand rule. Similarly, the direction of the area element **B** $\times$ **A,** also of same magnitude AB$\sin\theta$, is obtained by traversing the periphery of the parallelogram by going first over (dashed line) vector **B** and then over (dashed line) vector **A**, that is, traversing the periphery indicated by the thick clockwise arrow corresponding to direction going into the plane of the parallelogram according to the right-hand rule.

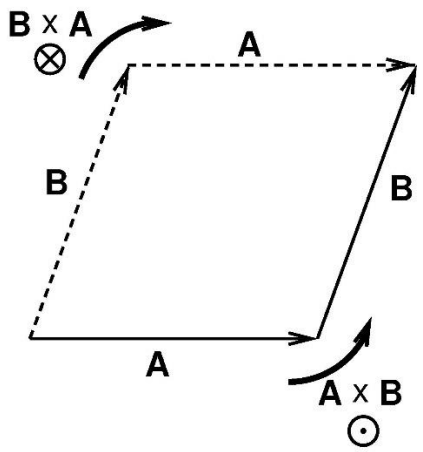


Figure 8.11. An area element as a vector.

8.4 TRIPLE SCALAR PRODUCT

Now let us explore how we can multiply three vectors **A**, **B**, and **C** to obtain either a scalar or a vector. The product **A**•(**B** × **C**) is a scalar and is referred to as the triple scalar product. In terms of components,

$$\mathbf{A}\bullet(\mathbf{B}\times\mathbf{C}) = A_1(\mathbf{B}\times\mathbf{C})_1 + A_2(\mathbf{B}\times\mathbf{C})_2 + A_3(\mathbf{B}\times\mathbf{C})_3$$

$$= A_1(B_2C_3 - B_3C_2) + A_2(B_3C_1 - B_1C_3) + A_3(B_1C_2 - B_2C_1)\ .$$

The expression on the right-hand side can be rewritten in two alternate forms:

$$\text{first form} = B_1(C_2A_3 - C_3A_2) + B_2(C_3A_1 - C_1A_3) + B_3(C_1A_2 - C_2A_1) = \mathbf{B}\bullet(\mathbf{C}\times\mathbf{A})\ ,$$

or

$$\text{second form} = C_1(A_2B_3 - A_3B_2) + C_2(A_3B_1 - A_1B_3) + C_3(A_1B_2 - A_2B_1) = \mathbf{C}\bullet(\mathbf{A}\times\mathbf{B})\ .$$

Thus, the triple scalar product can be written as,

$$\mathbf{A}\bullet(\mathbf{B}\times\mathbf{C}) = \mathbf{B}\bullet(\mathbf{C}\times\mathbf{A}) = \mathbf{C}\bullet(\mathbf{A}\times\mathbf{B}) = -\mathbf{A}\bullet(\mathbf{C}\times\mathbf{B}) = -\mathbf{B}\bullet(\mathbf{A}\times\mathbf{C}) = -\mathbf{C}\bullet(\mathbf{B}\times\mathbf{A})\ . \qquad Eq.(8.9a)$$

In other words, dot and cross can be interchanged in a triple scalar product. The triple scalar product can also be written in the form of a determinant:

$$\mathbf{A}\bullet(\mathbf{B}\times\mathbf{C}) = \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix}\ , \qquad Eq.(8.9b)$$

or, in terms of Levi-Civita symbol,

$$\mathbf{A}\bullet(\mathbf{B}\times\mathbf{C}) = \sum_{i,j,k} \epsilon_{ijk} A_i B_j C_k \quad . \qquad Eq.(8.9c)$$

The triple scalar product has a nice geometrical interpretation. Consider the parallelepiped made up by vectors **A**, **B**, and **C** as shown in Figure 8.12. The area of the base of the parallelepiped is the parallelogram **A** × **B**. The direction of this area element is along $\boldsymbol{n}$, the normal to the base. Then,

$$(\mathbf{A}\times\mathbf{B})\bullet\mathbf{C} = |\mathbf{A}\times\mathbf{B}|(\boldsymbol{n}\bullet\mathbf{C})$$

= (area of the base times height) = volume of the parallelepiped.

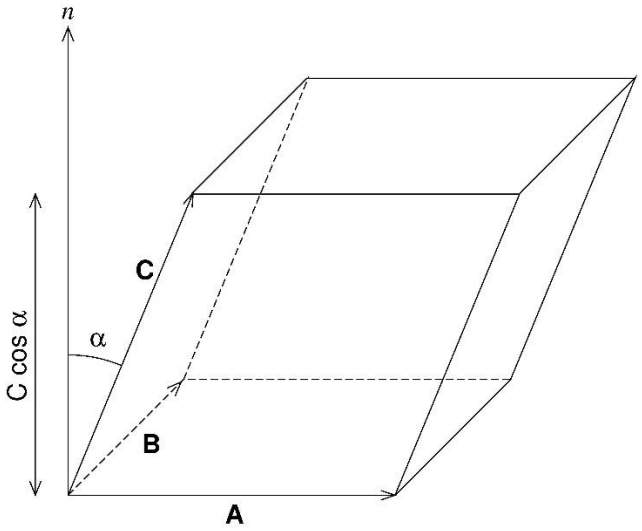


Figure 8.12. A parallelepiped made up by vectors **A**, **B**, and **C**.

8.5 TRIPLE VECTOR PRODUCT

Now we explore the possibility of obtaining a vector by multiplying three vectors **A**, **B**, and **C**. In the triple vector product **A** × (**B** × **C**), the order of multiplying vectors is important. For example,

$$\mathbf{e_1}\times(\mathbf{e_2}\times\mathbf{e_2}) \neq (\mathbf{e_1}\times\mathbf{e_2})\times\mathbf{e_2} \ ,$$

since the left-hand side is zero because the vector product of a vector by itself is zero, but the right-hand side is nonzero, equal to $\mathbf{e_3}\times\mathbf{e_2} = -\mathbf{e_1}$. In general, the triple vector product can be written as

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \mathbf{A}\times\mathbf{V} = \sum_{i,j,k} \epsilon_{ijk}\mathbf{e_i} A_j V_k \qquad Eq.(8.10a)$$

where

$$\mathbf{V} = \mathbf{B} \times \mathbf{C} \quad \text{and} \quad V_k = \sum_{l,m} \epsilon_{klm} B_l C_m \ . \qquad Eq.(8.10b)$$

On substituting for V_k from Eq. (8.10b) into Eq. (8.10a), we get

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \sum_{ijk} \epsilon_{ijk} \mathbf{e_i} A_j \sum_{lm} \epsilon_{klm} B_l C_m \ .$$

To carry out the sum over the repeated index k, we use the epsilon-delta identity (see Appendix C for proof),

$$\sum_{k} \epsilon_{ijk} \epsilon_{klm} = \sum_{k} \epsilon_{kij} \epsilon_{klm} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl} \ ,$$

so that the triple vector product looks like:

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \sum_{ij} \sum_{lm} \mathbf{e_i} A_j B_l C_m \left[\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\right] \ .$$

Because of the Kronecker deltas, the sum over the repeated dummy indices l and m can be carried out easily,

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \sum_{ij} \mathbf{e_i} A_j \left[B_i C_j - B_j C_i\right] = \sum_{i} \mathbf{e_i} B_i \sum_{j} A_j C_j - \sum_{i} \mathbf{e_i} C_i \sum_{j} A_j B_j$$

$$= \mathbf{B}(\mathbf{A} \bullet \mathbf{C}) - \mathbf{C}(\mathbf{A} \bullet \mathbf{B}) \ . \qquad Eq.(8.11)$$

The order of vectors appearing on the right-hand side is remembered by using the mnemonic BAC-CAB. Another form of triple vector product is

$$(\mathbf{A} \times \mathbf{B}) \times \mathbf{C} = -\mathbf{C} \times (\mathbf{A} \times \mathbf{B}) = -\mathbf{A}(\mathbf{B} \bullet \mathbf{C}) + \mathbf{B}(\mathbf{C} \bullet \mathbf{A}) \ , \qquad Eq.(8.12)$$

after using the BAC-CAB rule in the last step. Now, the order of vectors in the two forms of the triple vector product, namely, Eqs. (8.11) and (8.12), can be determined by:

Vector product of three vectors = middle vector times scalar product of the remaining two vectors minus $(-)$ other vector in the parenthesis times scalar product of the remaining two vectors.

PROBLEMS FOR CHAPTER 8

1. Using the concept of scalar, or dot, product of two vectors, determine the angle between vectors $\mathbf{A} = \mathbf{e_x} + 2\mathbf{e_y} + 2\mathbf{e_z}$ and $\mathbf{B} = 2\mathbf{e_x} - 3\mathbf{e_y} + 6\mathbf{e_z}$.

2. Prove that the vectors $\mathbf{A} = 3\mathbf{e_x} + \mathbf{e_y} - 2\mathbf{e_z}$, $\mathbf{B} = -\mathbf{e_x} + 3\mathbf{e_y} + 4\mathbf{e_z}$, and $\mathbf{C} = 4\mathbf{e_x} - 2\mathbf{e_y} - 6\mathbf{e_z}$ can form the sides of a triangle. Find the lengths of the medians of this triangle.

3. Using the concept of vector, or cross, product of two vectors, determine the area of a triangle with vertices at points A(1,2,3), B(2,4,5), and C(1,-8,0). In our notation here, P(x, y, z) refers to the Cartesian coordinates of the point P.

4. Determine a unit vector perpendicular to the plane containing vectors $\mathbf{A} = 2\mathbf{e_x} - 6\mathbf{e_y} - 3\mathbf{e_z}$ and $\mathbf{B} = 4\mathbf{e_x} + 3\mathbf{e_y} - \mathbf{e_z}$.

5. The eight corners of a parallelepiped are labeled as C_1, C_2, C_3, C_4, C_5, C_6, C_7, and C_8. The Cartesian (x, y, z) coordinates of the corners are C_1 (5, 2, 0), C_2 (2, 5, 0), C_3 (4, 5, 0), C_4 (7, 2, 0), C_5 (6, 3, 3), C_6 (3, 6, 3), C_7 (5, 6, 3), and C_8 (8, 3, 3).

   (a) Determine the three vectors **A**, **B**, and **C** that define this parallelepiped.

   (b) Determine the volume of the parallelepiped.

   [Hint: To visualize this parallelepiped, plot and label the first four and the last four corner points in two $x - y$ planes separately. The two $x - y$ planes are $z = 0$ plane and $z = 3$ plane.]

6. Prove the following identity for vector triple product:

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) + \mathbf{B} \times (\mathbf{C} \times \mathbf{A}) + \mathbf{C} \times (\mathbf{A} \times \mathbf{B}) = 0 \ .$$

7. Prove the following two identities:

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

$$(\mathbf{A} \times \mathbf{B}) \times (\mathbf{C} \times \mathbf{D}) = [(\mathbf{A} \times \mathbf{B}) \bullet \mathbf{D}]\,\mathbf{C} - [(\mathbf{A} \times \mathbf{B}) \bullet \mathbf{C}]\,\mathbf{D} = [(\mathbf{C} \times \mathbf{D}) \bullet \mathbf{A}]\,\mathbf{B} - [(\mathbf{C} \times \mathbf{D}) \bullet \mathbf{B}]\,\mathbf{A} \ .$$

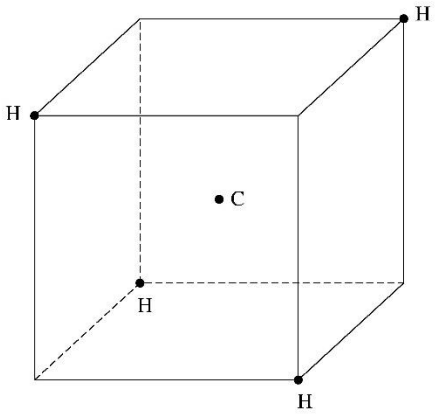
8. If **A** and **B** are any two vectors, then show that

$$(\mathbf{A} \times \mathbf{B}) \bullet (\mathbf{A} \times \mathbf{B}) = A^2 B^2 - (\mathbf{A} \bullet \mathbf{B})^2 \ .$$

9. A rhombus is a parallelogram with all four sides of equal length. Using the concept of scalar (or dot) product of vectors, show that the two diagonals of a rhombus are always perpendicular to each other.

10. **Biomedical Physics Application**. The "life", as we know it, is based on the chemistry of carbon-containing molecules. An understanding of the structure of these molecules is essential to understanding life. One of the simplest hydrocarbon molecules is methane (CH_4), in which four hydrogen atoms are placed symmetrically at the vertices of a tetrahedron around a single carbon atom. Alternatively, the four hydrogen atoms are located at the four corners of a cube, as shown in the figure below, with the carbon atom at the center of the cube. The bond angle is the angle between the lines that join the carbon atom to two of the hydrogen atoms. Using your knowledge of the scalar product of vectors, show that the bond angle in methane molecule is 109.5°.

[Hint: Take the origin of the coordinate system to be at the center of the cube (at carbon atom) of side length *2a* and determine the coordinates of each hydrogen atom.]



11. Consider an arbitrary triangle whose three sides are made up of three vectors **A**, **B** and **C**, then according to vector addition, **A** + **B** + **C** = 0. The three medians of this triangle, namely $\mathbf{M_A}$, $\mathbf{M_B}$, and $\mathbf{M_C}$, are also vectors. The median $\mathbf{M_A}$ has its tail at the vertex opposite **A** and its head at the midpoint of **A**. There is similar construction for other two medians $\mathbf{M_B}$ and $\mathbf{M_C}$. Using vector addition, show that

$$\mathbf{M_A} = -\frac{1}{2}(\mathbf{B} - \mathbf{C}) \ ,$$

$$\mathbf{M_B} = -\frac{1}{2}(\mathbf{C} - \mathbf{A}) \ ,$$

$$\mathbf{M_C} = -\frac{1}{2}(\mathbf{A} - \mathbf{B}) \ .$$

12. The three vertices of a triangle in the $x - y$ plane have Cartesian coordinates $(x_1, y_1), (x_2, y_2), \text{and } (x_3, y_3)$. Using the facts that three medians of a triangle intersect each other at a common point, called a centroid, and at this point each median is trisected, show, using vector addition, that the Cartesian coordinates of the centroid are

$$(x_c, y_c) = \left(\frac{x_1 + x_2 + x_3}{3}, \frac{y_1 + y_2 + y_3}{3}\right) .$$

13. The three vertices of a triangle in the $x - y$ plane have Cartesian coordinates $(x_1, y_1), (x_2, y_2), \text{and } (x_3, y_3)$. Using the cross product of two vectors, show that the area of this triangle is given by the following determinant:

$$\text{area } = \frac{1}{2} \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} .$$

Chapter 9: Curvilinear Coordinates and Multiple Integrals

In this chapter, we will introduce two important curvilinear coordinate systems. These are the spherical coordinate system and the cylindrical coordinate system. They are very useful in carrying out multiple integrals encountered in problems with spherical and cylindrical symmetries.

9.1 CARTESIAN COORDINATES

First, we review the details of the well-known and more common rectangular Cartesian coordinate system. In this system, a point has three coordinates (x, y, z) which are the distances perpendicular to three mutually orthogonal surfaces. All three coordinates (x, y, z), vary independently from $-\infty$ to $+\infty$. Then we define some families of surfaces and families of curves for Cartesian coordinate system. A surface on which the value of the x

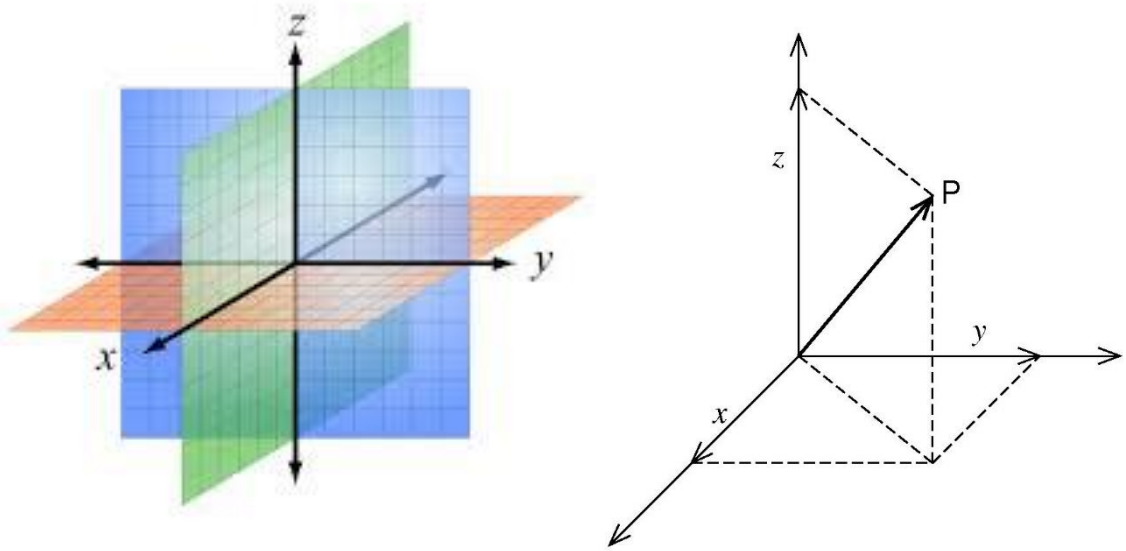


Figure 9.1. A point P in space with Cartesian coordinates x, y, z.

coordinate is constant, while the values of the y and z coordinates can vary, is called the $x =$ constant surface. This surface is a plane. Similarly, $y =$ constant and $z =$ constant surfaces are planes, as shown in Figure 9.1. A curve along which the x coordinate varies, while the y and z coordinates are constant and fixed is called the x–curve (or, x–axis). This curve is a straight line. Similarly, the y–curve and z– curve are straight lines. The unit vectors $\mathbf{e_x}$, $\mathbf{e_y}$, and $\mathbf{e_z}$ are tangentially along the x–curve, y–curve, and z–curve, respectively. Consider a point, P, in space with Cartesian coordinates (x, y, z). The vector, shown in Figure 9.1,

$$\mathbf{r} = x\,\mathbf{e_x} + y\,\mathbf{e_y} + z\,\mathbf{e_z}\ ,$$

locates the position of point P with respect to the origin. A length element **dl** in this coordinate system is

$$\mathbf{dl} = dx\,\mathbf{e}_x + dy\,\mathbf{e}_y + dz\,\mathbf{e}_z\ , \qquad Eq.(9.1a)$$

and the distance between two neighboring points is

$$(dl)^2 = (dx)^2 + (dy)^2 + (dz)^2\ . \qquad Eq.(9.1b)$$

A volume element in Cartesian coordinate system is

$$dV = dxdydz\ . \qquad Eq.(9.1c)$$

An area element in this system is

$$\mathbf{dS} = dxdy\,\mathbf{e}_z + dydz\,\mathbf{e}_x + dzdx\,\mathbf{e}_y\ . \qquad Eq.(9.1d)$$

9.2 SPHERICAL COORDINATES

Again, consider the same point, P, in space located at a position given by the vector $\mathbf{r}$ as shown in Figure 9.2. The spherical coordinates (r, θ, ϕ) of point P are defined as follows. The coordinate r is the length of vector $\mathbf{r}$ and its range is $0 \leq r \leq \infty$. The coordinate θ is the angle between vector $\mathbf{r}$ and the z-axis and its range is $0 \leq \theta \leq \pi$. It is similar to the angle of co-latitude in astronomy. In order to describe the third spherical coordinate, ϕ, we need to focus our attention on two planes, the $y = 0$ plane (or the $x-z$ plane) and the plane containing the vector $\mathbf{r}$ and the z-axis. Then, ϕ is the angle between these two planes. The coordinate ϕ is similar to the angle of longitude in astronomy and its range is $0 \leq \phi \leq 2\pi$. The surface $r =$ constant is a sphere of radius r. The surface

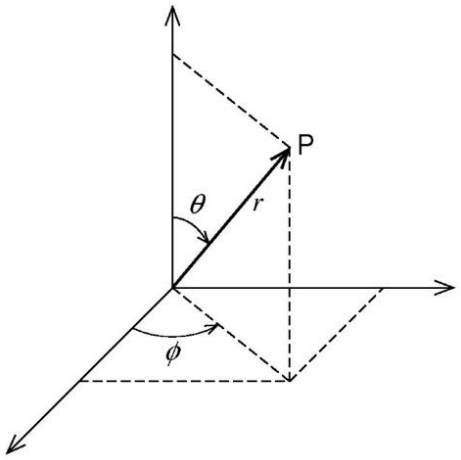


Figure 9.2. The point P of Figure 9.1 shown with its spherical coordinates r, θ, ϕ.

$\theta =$ constant is the curved surface of a cone having vertex at the origin. This cone becomes the $z = 0$ plane if $\theta = \pi/2$, becomes the $+z$-axis if $\theta = 0$, and becomes the $-z$-axis if $\theta = \pi$. The surface $\phi =$ constant is a semi-infinite plane. A curve along which the r coordinate varies, while the θ and ϕ coordinates are constant and fixed, is called the $r-$ curve. This curve is a straight line. Similarly, the $\theta-$ curve is a semicircle and the $\phi-$ curve is a full

circle. The unit vectors $\mathbf{e_r}$, $\mathbf{e_\theta}$, and $\mathbf{e_\phi}$ are tangentially along the r–curve, θ–curve, and ϕ–curve, respectively. The relationships between spherical coordinates (r, θ, ϕ) and the Cartesian coordinates (x, y, z) are

$$\begin{aligned} x &= r \sin\theta \cos\phi \\ y &= r \sin\theta \sin\phi \\ z &= r \cos\theta \ , \end{aligned} \qquad Eq.(9.2a)$$

or the inverse relationships,

$$\begin{aligned} r &= (x^2 + y^2 + z^2)^{1/2} \\ \theta &= \arccos\left(\frac{z}{(x^2+y^2+z^2)^{1/2}}\right) \\ \phi &= \arctan(y/x) \end{aligned} \qquad Eq.(9.2b)$$

Thus,

$$\begin{aligned} dx &= \sin\theta \cos\phi \, dr + r \cos\theta \cos\phi \, d\theta - r \sin\theta \sin\phi \, d\phi \ , \\ dy &= \sin\theta \sin\phi \, dr + r \cos\theta \sin\phi \, d\theta + r \sin\theta \cos\phi \, d\phi \ , \\ dz &= \cos\theta \, dr - r \sin\theta \, d\theta \ . \end{aligned} \qquad Eq.(9.3)$$

The distance between two neighboring points is

$$(dl)^2 = (dx)^2 + (dy)^2 + (dz)^2 \ ,$$

or, on substituting for dx, dy, and dz in terms of dr, $d\theta$, and $d\phi$ from Eq. (9.3),

$$(dl)^2 = (dr)^2 + (r \, d\theta)^2 + (r \sin\theta \, d\phi)^2 \ . \qquad Eq.(9.4a)$$

Thus, we note that the roles of length elements dx, dy, and dz of Cartesian coordinates are taken over, in spherical coordinates, by length elements dr, $rd\theta$, and $r \sin\theta \, d\phi$. Thus, a volume element in spherical coordinates is

$$dV = r^2 \sin\theta \, dr \, d\theta \, d\phi \ , \qquad Eq.(9.4b)$$

and an area element is

$$\mathbf{dS} = r dr d\theta \, \mathbf{e_\phi} + r^2 \sin\theta \, d\theta d\phi \, \mathbf{e_r} + r \sin\theta \, dr d\phi \, \mathbf{e_\theta} \ . \qquad Eq.(9.4c)$$

9.3 CYLINDRICAL COORDINATES

Once again, consider the same point, P, in space with Cartesian coordinates (x, y, z) and spherical coordinates (r, θ, ϕ). The cylindrical coordinates (ρ, ϕ, z) of this point are shown in Figure 9.3 and are defined as follows. The cylindrical z coordinate is same as the Cartesian z coordinate. Thus, the range of the z coordinate is $-\infty \leq z \leq \infty$. The cylindrical ϕ coordinate is same as the spherical ϕ coordinate. So, the range of this coordinate is $0 \leq \phi \leq 2\pi$. The cylindrical ρ coordinate is the projection of the position vector $\mathbf{r}$ onto the Cartesian $x-y$ plane. The range of this coordinate is $0 \leq \rho \leq \infty$. The surface $z =$ constant is a plane perpendicular to the z-axis. The surface $\phi =$ constant is a semi-infinite plane. The surface $\rho =$ constant is a cylinder coaxial with the z-axis. As in Cartesian coordinates, the $z-$curve is a straight line. As in spherical coordinates, the $\phi-$ curve is a full circle. The

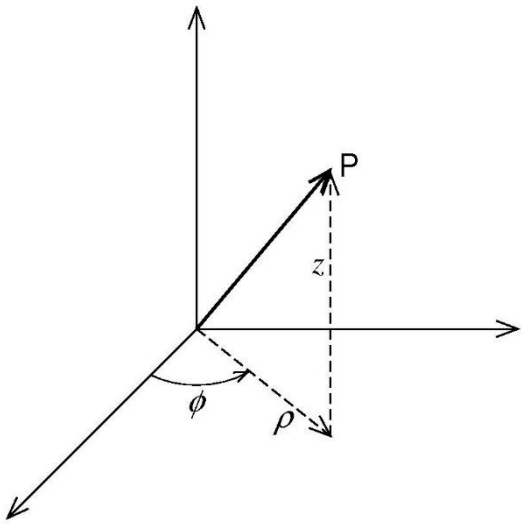


Figure 9.3. The point P of Figures 9.1 and 9.2 shown with its cylindrical coordinates ρ, ϕ, z.

$\rho-$curve is a curve along which the coordinate ρ varies while the other two coordinates ϕ and z are held constant. This curve is a straight line. The unit vectors $\mathbf{e}_\rho$, $\mathbf{e}_\phi$, and $\mathbf{e}_z$ are tangentially along the $\rho-$curve, the $\phi-$curve, and the $z-$ curve, respectively. The relationships between cylindrical coordinates (ρ, ϕ, z) and the Cartesian coordinates (x, y, z) are

$$x = \rho\cos\phi\,, y = \rho\sin\phi \quad \text{and} \quad z = z\ , \qquad Eq.(9.5a)$$

and the inverse relationships are

$$\rho = \sqrt{x^2 + y^2},\ \phi = \arctan(y/x)\ \text{and}\ z = z\ . \qquad Eq.(9.5b)$$

Using these relationships,

$$dx = \cos\phi\ d\rho - \rho\sin\phi\ d\phi\ ,$$

$$dy = \sin\phi \; d\rho + \rho\cos\phi \; d\phi \;, \qquad Eq.(9.6)$$

$$dz = dz \;.$$

Now, the separation between two neighboring points, given by

$$(dl)^2 = (dx)^2 + (dy)^2 + (dz)^2,$$

can be rewritten on substituting for dx, dy, and dz in terms of $d\rho$, $d\phi$, and dz from Eq. (9.6), as

$$(dl)^2 = (d\rho)^2 + (\rho d\phi)^2 + (dz)^2 \;. \qquad Eq.(9.7a)$$

Thus, we note that the length elements dx, dy and dz of Cartesian coordinates are replaced, in cylindrical coordinates, by length elements $d\rho$, $\rho d\phi$ and dz. So, the volume element and the area element in cylindrical coordinates become

$$dV = \rho d\rho d\phi dz \;, \qquad Eq.(9.7b)$$

and,

$$\mathbf{dS} = \rho d\rho d\phi \; \mathbf{e_z} + \rho d\phi dz \; \mathbf{e_\rho} + dz d\rho \; \mathbf{e_\phi} \;, \qquad Eq.(9.7c)$$

respectively.

9.4 PLANE POLAR COORDINATES

We note in passing that because the z coordinate in Cartesian coordinates and in cylindrical coordinates is the same, it is possible to treat the $x-y$ plane as a $\rho-\phi$ plane. This relationship is a depiction of two-dimensional *plane polar coordinates*. In particular, an area element in this plane can be written as

$$dS = dxdy \;\text{ or }\; dS = \rho d\rho \; d\phi \;. \qquad Eq.(9.8)$$

As an example of plane polar coordinates, we attend to an unfinished integral that we encountered in our discussion of Gaussian integrals in Chapter 2. We stated there, without proof, that [see remark after Eq. (2.15)]

$$I_g^0 = \int_0^\infty \exp(-ax^2) \; dx = \frac{1}{2}\sqrt{\frac{\pi}{a}} \;.$$

Now we evaluate this integral. But, first we evaluate the related integral,

$$I = \int_0^\infty \exp(-x^2)\, dx \;.$$

Then, the integral I_g^0 will be obtained from I by replacing x^2 by ax^2. We write the I integral twice, first with dummy variable x and next with dummy variable y, and multiply the two integrals to get

$$I^2 = \int_0^\infty \int_0^\infty \exp[-(x^2 + y^2)]\, dx\, dy \ .$$

Now, think of this integral as an integral over the first quadrant of the $x-y$ plane. Converting to plane polar coordinates in the first quadrant,

$$I^2 = \int_{\rho=0}^{\infty} \int_{\phi=0}^{\pi/2} \exp(-\rho^2)\, \rho d\rho\, d\phi = \frac{\pi}{2} \frac{1}{2} \int_0^\infty \exp(-u)\, du = \frac{\pi}{4} \ ,$$

with $u = \rho^2$. Thus,

$$I = \int_0^\infty \exp(-x^2)\, dx = \frac{\sqrt{\pi}}{2} \ . \qquad Eq.(9.9)$$

Now, to get the Gaussian integral I_g^0 simply replace x^2 by ax^2 in Eq.(9.9). This gives

$$I_g^0 = \int_0^\infty \exp(-ax^2)\ dx = \frac{1}{2}\sqrt{\frac{\pi}{a}} \ .$$

9.5 MULTIPLE INTEGRALS

The curvilinear coordinates, discussed above, are very useful in carrying out integrals which involve areas or volumes of circles, spheres and cylinders. A few examples will impress the reader about the utility of curvilinear coordinates.

Example: Show that the area, A, of a circle of radius R is, $A = \pi R^2$.

Solution:

The statement of this problem is so well-known that it is almost taken as a fact. However, historically speaking, it was considered quite an intellectual feat by Eudoxus in the fifth century BC to realize that the area of a circle is proportional to the square of its radius. At the time of Eudoxus, the formula for finding out the area of a triangle was well-known. So, the area of a circle was determined by inscribing in it regular polygons, which can be broken into several triangles. If A_n is the area of an inscribed polygon of n sides, then $A_{circle} = \lim_{n\to\infty} A_n$.

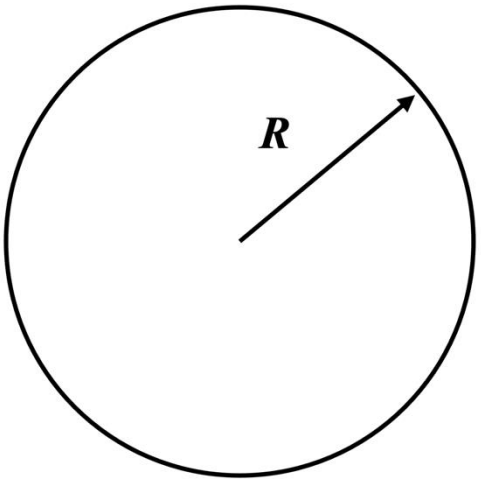


Figure 9.4. A circle of radius R.

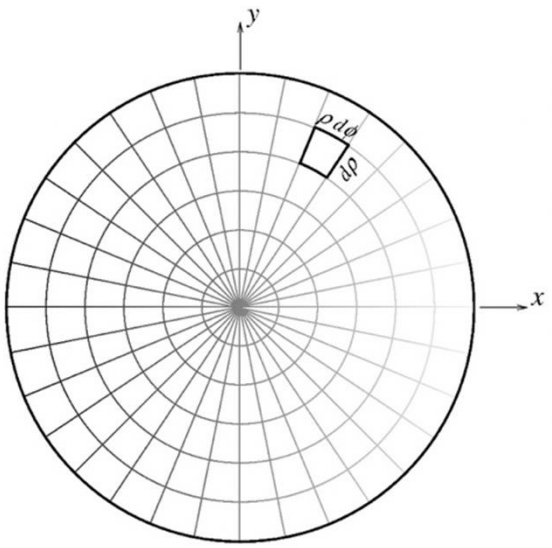


Figure 9.5. Calculating area of a circle.

Now, using multiple integrals, it is very easy to determine the area of a circle. Break the circle into large number of small pieces by drawing various concentric circles and radial spokes inside the circle, as in Figure 9.5. The area of the tiny area element shown in the figure is,

$$dA = (d\rho)(\rho d\phi) = \rho d\rho d\phi \ .$$

Then,

$$A_{circle} = \int_{circle} dA = \int_0^R \rho d\rho \int_0^{2\pi} d\phi = \frac{R^2}{2}(2\pi) = \pi R^2 \ . \qquad Eq.(9.10)$$

Example: Determine volume, V, of a sphere of radius R.

Solution: Since a small volume element in spherical coordinates, $\boldsymbol{dV}$, is given in Eq. (9.4b), the volume of a whole sphere can be determined by summing (or, integrating) over all the small volume elements that make up the sphere. Explicitly,

$$V = \int_{sphere} dV = \int_{r=0}^{R} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} r^2 \sin\theta \, d\theta dr d\phi \ ,$$

or,

$$V = \int_{\phi=0}^{2\pi} d\phi \int_{\theta=0}^{\pi} \sin\theta \, d\theta \int_{r=0}^{R} r^2 dr = (2\pi)(2)\left(\frac{R^3}{3}\right) = \frac{4\pi}{3}R^3 \ . \qquad Eq.(9.11)$$

9.6 SOLIDS OF REVOLUTION

Consider a small strip of length dx located in the $x-y$ plane at $(x, y) \equiv (x, R)$ as in Figure 9.6. If this strip is revolved about the x-axis, it creates a disk of radius R and thickness dx, and, therefore, of volume $V = \pi R^2 dx$. Generalizing this result, if a curve, representing a function $f(x)$ with $a \leq x \leq b$, is revolved about x-axis, as in Figure 9.6, the volume of the solid generated will be

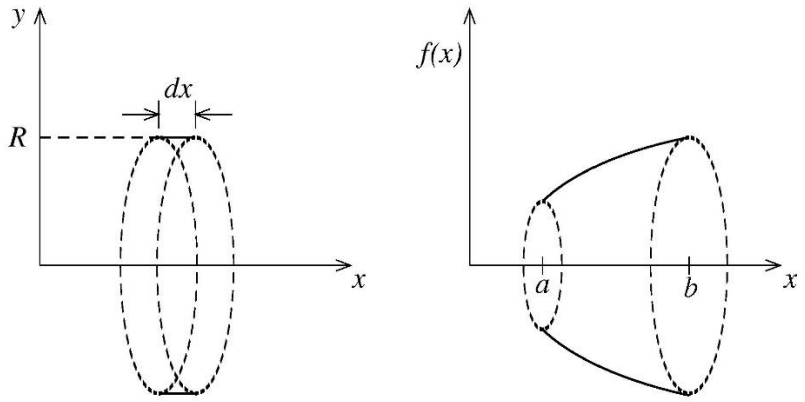


Figure 9.6. Solids of revolution.

$$V = \int_{a}^{b} \pi[f(x)]^2 dx \ . \qquad Eq.(9.12)$$

The solids obtained by revolving a curve about an axis are called the solids of revolution.

Example: **A straight line, given by the equation**

$$\boldsymbol{f(x) = \frac{R}{H}x,}$$

is shown in the Figure 9.7. Revolution of this straight line about x-axis leads to a solid of revolution in the form of a cone of radius R and height H. The vertex of this cone is at the origin and its axis lies along the x-axis. The volume of this cone is

$$V = \int_{0}^{H} \pi\left[\frac{R}{H}x\right]^2 dx = \pi\frac{R^2}{H^2}\int_{0}^{H} x^2 dx = \pi\frac{R^2}{H^2}\frac{H^3}{3} = \frac{\pi}{3}R^2 H \ . \qquad Eq.(9.13)$$

Thus, in general, the volume of a cone of radius R and height H is $\pi R^2 H/3$.

Example: Consider a horizontal line of length H in the $x-y$ plane at $y=R$. Revolution of this line around x-axis generates a solid in the form of a cylinder of radius R and height H. Its volume is

$$\boldsymbol{V = \int_0^H \pi[R]^2 dx = \pi R^2 H} \ .$$

Thus, in general, the volume of a cylinder of radius R and height H is $\pi R^2 H$.

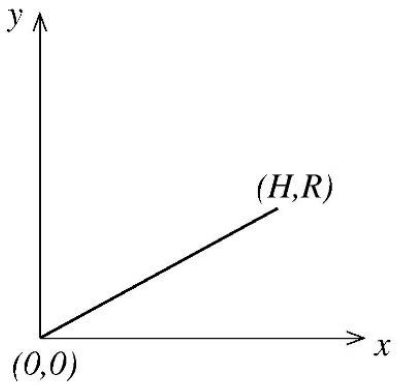


Figure 9.7. A cone generated as a solid of revolution.

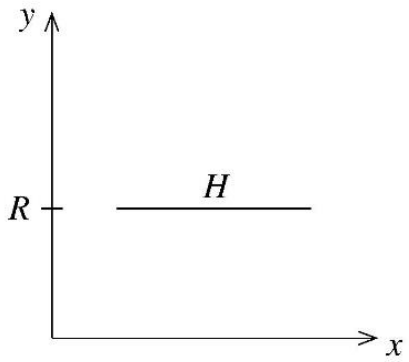


Figure 9.8. A cylinder generated as a solid of revolution.

9.7 CENTER OF MASS

The concept of center of mass is very useful in the study of mechanics of a distribution of masses. The center of mass of the distribution, under the influence of external forces, moves as if all the mass of the distribution were concentrated there and as if all the external forces were applied there.

Consider a mass distribution consisting of n discrete point-like masses, $\Delta m_1, \Delta m_2, \cdots \Delta m_i \cdots \Delta m_n$ located at $\boldsymbol{r_1}, \boldsymbol{r_2} \cdots \boldsymbol{r_i} \cdots \boldsymbol{r_n}$, respectively, as shown in Figure 9.9. The total mass M is the sum of all discrete masses. The center of mass of this distribution is located at

$$\boldsymbol{r}_{cm} = \frac{(\Delta m_1)\boldsymbol{r}_1 + (\Delta m_2)\,\boldsymbol{r}_2 + \ldots + (\Delta m_i)\,\boldsymbol{r}_i + \ldots + (\Delta m_n)\,\boldsymbol{r}_n}{(\Delta m_1) + (\Delta m_2) + \cdots + (\Delta m_n)}$$

$$= \frac{1}{M}\sum_{i=1}^{n} \boldsymbol{r}_i\,\Delta m_i \quad . \qquad Eq.(9.14a)$$

An extended object of total mass M can be imagined as a collection of a large number of point-like masses which are combined together to form a continuous distribution of mass. For such an object, the sum in Eq. (9.14a) can

be replaced by an integral. Thus, for a continuous distribution of mass, M, the center of mass of the body is located at

$$\boldsymbol{r}_{cm} = \frac{1}{M} \int_{all\ body} \boldsymbol{r}\ dm\ , \qquad Eq.(9.14b)$$

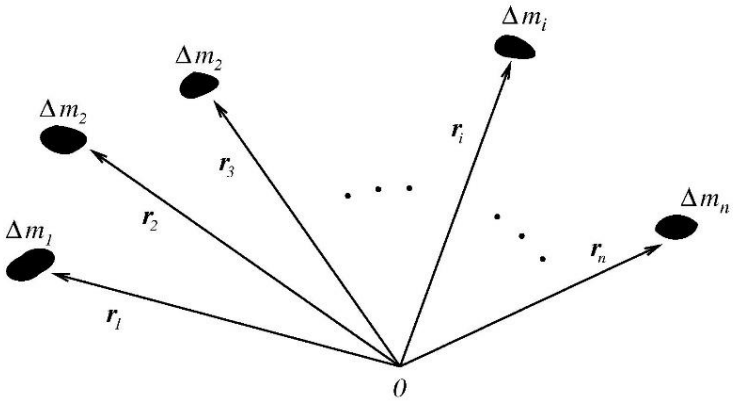


Figure 9.9. The center of mass of a distribution of n discrete point-like masses.

or, in Cartesian coordinates,

$$x_{cm} = \frac{1}{M}\int x\ dm\ ,$$

$$y_{cm} = \frac{1}{M}\int y\ dm\ ,$$

$$z_{cm} = \frac{1}{M}\int z\ dm\ .$$

For a mass distribution with some symmetry, the center of mass is normally at the center of symmetry or on the axis of symmetry. For example, for a full circular disk of radius R, the center of mass is at the center of symmetry, namely, the center of the disk.

Example: Determine the center of mass of a semicircular disk of radius *R* and mass *M*.

Solution:

In this case, the mass per unit area is $\mu = M/\left(\frac{1}{2}\pi R^2\right) = 2M/(\pi R^2)$. Using plane polar coordinates in the plane of the disk:

$$x_{cm} = \frac{1}{M}\int_0^R \int_0^\pi \mu\,(\rho\cos\phi)\,d\rho\;\rho\;d\phi \;= \frac{\mu}{M}\int_0^R \rho^2 d\rho \int_0^\pi \cos\phi\;\,d\phi = 0$$

$$y_{cm} = \frac{1}{M}\int_0^R \int_0^\pi \mu\,(\rho\sin\phi)\,d\rho\;\rho\;d\phi \;= \frac{\mu}{M}\int_0^R \rho^2 d\rho \int_0^\pi \sin\phi\;\,d\phi = \frac{\mu}{M}\frac{R^3}{3}2 = \frac{2}{\pi R^2}\frac{R^3}{3}2 = \frac{4R}{3\pi}\;.$$

Thus, center of mass of a semicircular disk is located on its axis of symmetry at $(0,4R/3\pi)$.

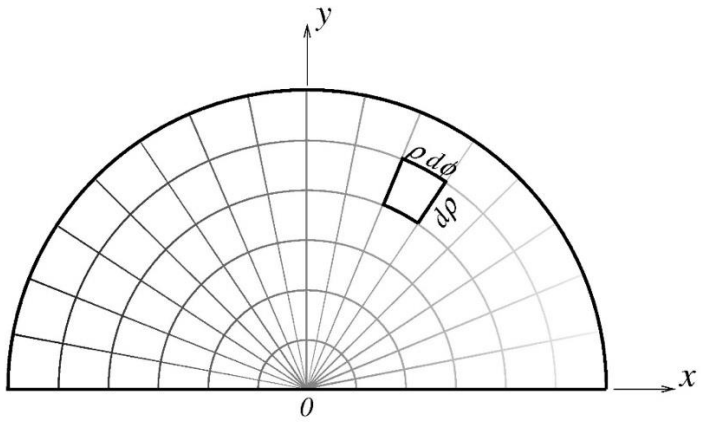


Figure 9.10. Center of mass of a uniform semicircular disk.

9.8 MOMENT OF INERTIA ABOUT AN AXIS

In elementary physics we learn that, during a translational motion, the kinetic energy of a body of mass m moving with a linear velocity v is $K = \frac{1}{2}mv^2$. Consider a person riding a stationary bicycle for exercise. Since the bicycle is stationary, there is no translational kinetic energy associated with the exercise machine. All the work done by pedaling the bicycle is converted into $rotational\ kinetic\ energy$ of the rotating wheel. If the wheel of the exercise bicycle is rotating with a constant angular velocity ω, then its kinetic energy is $K = \frac{1}{2}I\omega^2$, where I is the moment of inertia or the rotational inertia of the wheel. In order to define the moment of inertia of a mass distribution, imagine several discrete point-like masses $\Delta m_1, \Delta m_2, \cdots \Delta m_i \cdots \Delta m_n$ rotating about a certain common axis of rotation, as shown in Figure 9.11. Furthermore, $r_1, r_2 \cdots r_i \cdots r_n$, respectively, are the shortest, perpendicular distances of various masses from the axis of rotation. Then, the moment of inertia of this mass distribution about the axis of rotation is defined as

$$I = \sum_{i=1}^{n} r_i^2\,(\Delta m_i)\;\;. \qquad\qquad Eq.(9.15a)$$

An extended object can be imagined as a collection of a large number of point-like masses which are combined together to form a continuous distribution of mass. For such an object, the sum in Eq. (9.15a) can be replaced by an integral so that

$$I = \int_{all\ body} r^2\, dm \ . \qquad Eq.(9.15b)$$

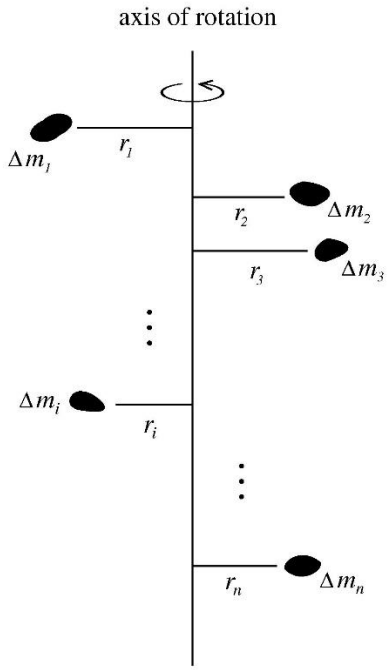


Figure 9.11. Moment of inertia of a distribution of n discrete point-like masses rotating about an axis.

If the mass is distributed uniformly in a volume V with density D_V (mass per unit volume), then

$$I = \int_V r^2\, D_V\, dV \ .$$

If the mass is distributed uniformly in an area A with density D_A (mass per unit area), then

$$I = \int_A r^2 D_A\, dA \ .$$

If the mass is distributed uniformly along a line L with density D_L (mass per unit length), then

$$I = \int_L r^2 D_L \, dL \ .$$

Example: Determine the moment of inertia of a thin spherical shell of mass M and radius R rotating about any diameter.

Solution: The shell has surface area of $4\pi R^2$. If D_A is mass per unit area, $D_A = M/(4\pi R^2)$, then mass of a small patch of area $dA = (Rd\theta)(R\sin\theta\, d\phi)$ is $dm = D_A\, dA$. The perpendicular distance of this patch of area from the axis of rotation is $r = R\sin\theta$. So,

$$I = \int_{shell} [R\sin\theta]^2 dm = \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} [R\sin\theta]^2 D_A\, R^2 \sin\theta \ d\theta\, d\phi$$

or, with $u = \cos\theta$,

$$I = D_A\, R^4 \int_0^{\pi} \sin^3\theta\, d\theta \int_0^{2\pi} d\phi = D_A\, R^4\, 2\pi \int_{-1}^{1} (1-u^2)\, du$$

$$= 2\pi\, D_A\, R^4 \left(\frac{4}{3}\right) = \frac{8\pi}{3}\left(\frac{M}{4\pi R^2}\right) R^4 = \frac{2}{3} MR^2 \ . \qquad Eq.(9.16)$$

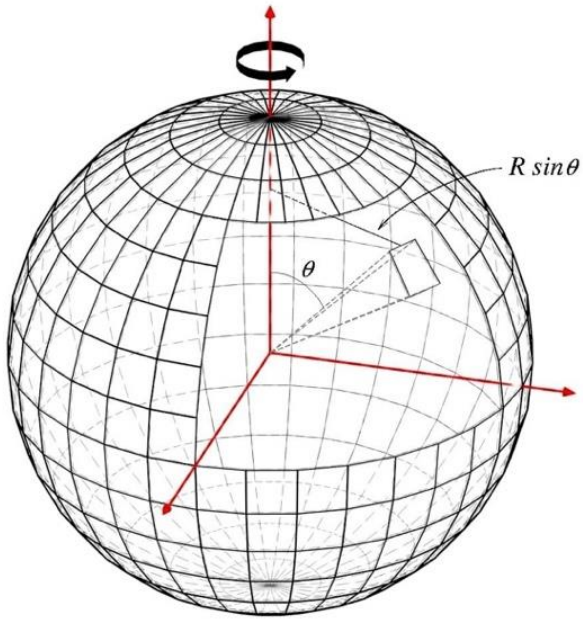


Figure 9.12. Moment of inertia of a thin spherical shell.

Example: A thin rod of length L and mass M is rotating about an axis. The axis is passing through the center of the rod and is perpendicular to the rod. Determine the moment of inertia of this rotating rod.

Solution:

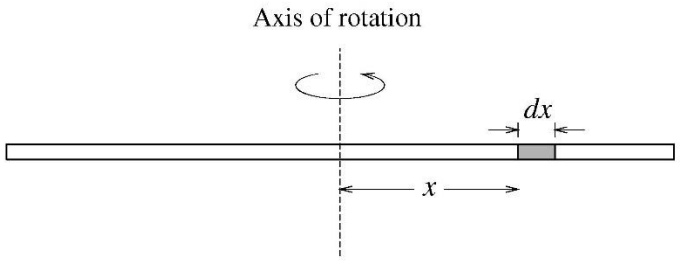


Figure 9.13. Moment of inertia of a rod rotating about an axis.

Since the mass is distributed along the length of the rod, we can define mass per unit length as $D_L = M/L$. Consider a small mass element of length dx located a distance x away from the axis of rotation. The mass of this element is $dm = D_L\, dx$. So, the moment of inertia is

$$I = \int_{rod} x^2 dm = \int_{-\frac{L}{2}}^{\frac{L}{2}} x^2 D_L\, dx = D_L \left.\frac{x^3}{3}\right|_{-\frac{L}{2}}^{\frac{L}{2}} = D_L \left[\frac{L^3}{24} - \left(-\frac{L^3}{24}\right)\right],$$

or,

$$I = \frac{M}{L}\left(\frac{L^3}{12}\right) = \frac{1}{12} ML^2 \ . \qquad Eq.\,(9.17)$$

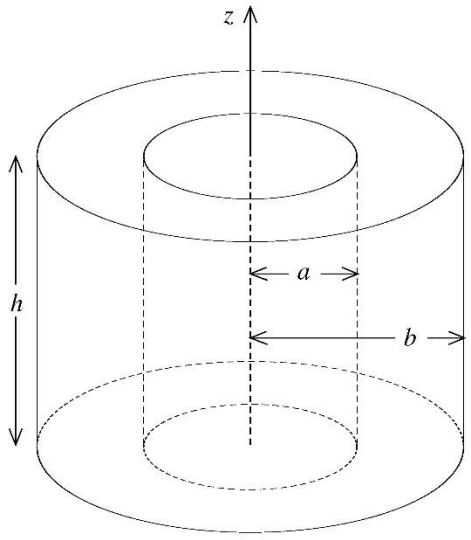
PROBLEMS FOR CHAPTER 9

1. Find the position of the center of mass of a solid cone of radius R and height H.

2. A solid sphere, of radius R, is cut into two identical hemispheres. Find the position of the center of mass of such a hemisphere.

3. Find the moment of inertia of a solid cone of radius R, uniform mass density D, height H, and mass M about its axis of symmetry. Give the answer in terms of M and R.

4. (a) Determine the moment of inertia of a thick-walled right circular cylindrical tube of mass M, inner radius a, outer radius b, and height h, about its central axis of symmetry.



(b) Using the result of part (a), determine the moment of inertia of a solid right circular cylinder of radius R and mass M.

(c) Using the result of part (a), determine the moment of inertia of a right circular cylindrical shell, open at both ends, of radius R and mass M. The thickness of the shell can be considered as almost zero.

5. (a) Determine the moment of inertia of a thick spherical shell, of mass M, about an axis passing through its center. The inner and outer radii of the shell are a and b, respectively.

(b) Using the result of part (a), determine the moment of inertia of a solid sphere of radius R and mass M.

(c) Using the result of part (a), determine the moment of inertia of a thin spherical shell of radius R and mass M. The thickness of the shell can be considered as almost zero.

6. **Biomedical Physics Application**. The human heart is a muscular organ that is responsible for pumping blood throughout the body. Without resorting to an actual surgery, the size of a healthy heart can be estimated by

using a medical imaging technique called CAT scan or simply CT scan (for computed tomography). The scan produces equally spaced cross-sectional views of the heart. Suppose that a CT scan of a human heart shows cross sections spaced 1.0 cm apart. The heart is about 10 cm long and the cross-sectional areas, in square centimeters, are 0, 18, 38, 57, 72, 84, 71, 60, 39, 20, and 0. Using this information estimate the volume of the heart.

7. **Biomedical Physics Application**. The velocity v of blood that flows in a blood vessel with radius R and length L at a distance r from the central axis is

$$v(r) = \frac{P}{4\eta L}(R^2 - r^2) \ ,$$

where P is the pressure difference between the ends of the vessel and η is the viscosity of the blood. The flux (volume of blood flowing per unit time) F of the blood can be calculated by approximating the cross-sectional area of the blood vessel by concentric rings and integrating over all such rings. Show that the flux is given by

$$F = \frac{\pi P R^4}{8\eta L} \ .$$

This is called Poiseuille's Law.

Chapter 10: Vector Calculus

In this chapter we will combine our previous knowledge of calculus and of vectors to learn about vector functions and their derivatives. We will also gain knowledge of a couple of specialized theorems, the Gauss's theorem (or, divergence theorem) and Stoke's theorem, which are relevant for vector functions.

10.1 VECTOR FUNCTIONS

A vector function is a vector whose components are functions of coordinates. As an example, a vector function in $Cartesian\ coordinates$ is

$$\mathbf{A}(x_1, x_2, x_3) = \mathbf{e_1} A_1(x_1, x_2, x_3) + \mathbf{e_2} A_2(x_1, x_2, x_3) + \mathbf{e_3} A_3(x_1, x_2, x_3) \ . \qquad Eq.(10.1)$$

Here $\mathbf{e_1}, \mathbf{e_2}, \text{and } \mathbf{e_3}$ are the three orthogonal unit vectors and $A_1, A_2, \text{and } A_3$ are the three scalar functions of coordinates $x_1, x_2, \text{and } x_3$.

Now, in vector calculus, just as in multivariate calculus, all derivatives are related to the rate of change of a function. For a scalar function $f(x_1, x_2, x_3)$, the three derivatives, $\frac{\partial f}{\partial x_i}$ $(i = 1,2,3)$, represent the rate of variation of the function f with respect to x_i $(i = 1,2,3)$. These three derivatives of f behave like the three components of a vector, which is called the gradient of f. The gradient of f is written as

$$\boldsymbol{\nabla} f = \mathbf{e_1} \frac{\partial f}{\partial x_1} + \mathbf{e_2} \frac{\partial f}{\partial x_2} + \mathbf{e_3} \frac{\partial f}{\partial x_3} \ . \qquad Eq.(10.2)$$

Note that gradient of f is a vector function. The differential operator

$$\boldsymbol{\nabla} = \mathbf{e_1} \frac{\partial}{\partial x_1} + \mathbf{e_2} \frac{\partial}{\partial x_2} + \mathbf{e_3} \frac{\partial}{\partial x_3} = \sum_{i=1}^{3} \mathbf{e_i} \frac{\partial}{\partial x_i}$$

is called the del or $gradient$ operator. Let us now interpret the gradient of a scalar function. Consider the values of a scalar function f at two neighboring points in space. The first point is at $\mathbf{r} = (x_1, x_2, x_3)$ and the value of the function there is $f(x_1, x_2, x_3)$. The neighboring point is at $\mathbf{r} + \mathbf{dr} = (x_1 + dx_1, x_2 + dx_2, x_3 + dx_3)$ and the value of the function there is $f(x_1 + dx_1, x_2 + dx_2, x_3 + dx_3)$, as shown in Figure 10.1. The change in the value of the function, f, on moving from one point to a neighboring point, is

$$df = f(x_1 + dx_1, x_2 + dx_2, x_3 + dx_3) - f(x_1, x_2, x_3) \ ,$$

or, keeping only the first order terms in the Taylor-series expansion of the function,

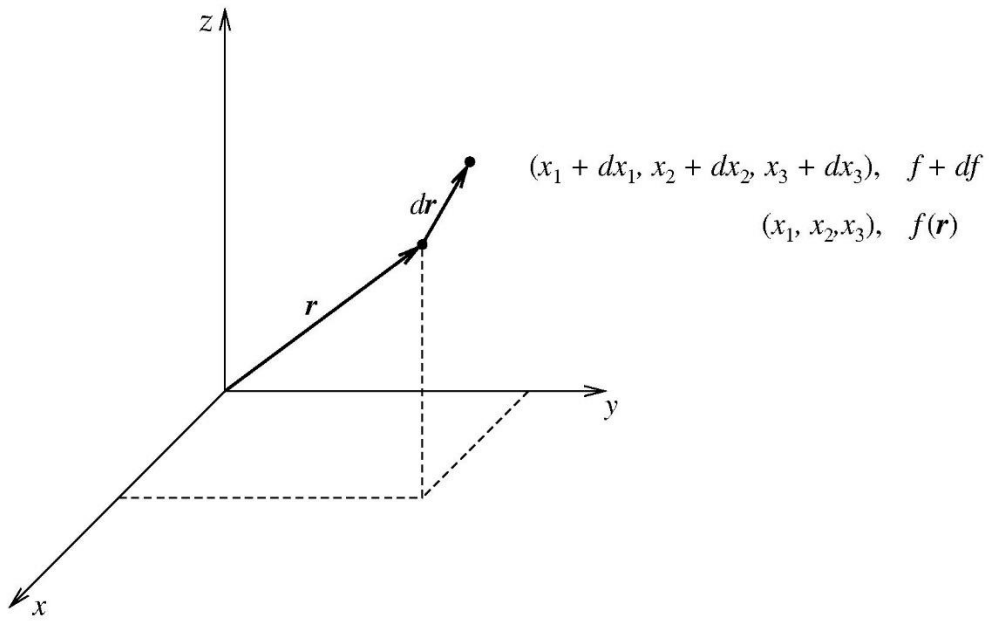


Figure 10.1. Values of a function at two neighboring points.

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \frac{\partial f}{\partial x_3} dx_3 = \boldsymbol{\nabla} f \bullet \mathbf{dr} \ . \qquad Eq.(10.3)$$

This expression leads to the following properties of the gradient of a scalar function f.

First, the change, df, in the value of the function is maximum when $\boldsymbol{\nabla} f$ is parallel to $\mathbf{dr}$. Thus $\boldsymbol{\nabla} f$ is a vector that represents the magnitude and direction of the greatest rate of change of f. Second, if $\mathbf{dr}$ is tangential to the surface $f(\mathbf{r}) = \text{constant}$, then both $\mathbf{r}$ and $\mathbf{r} + \mathbf{dr}$ lie on the surface. So,

$$df = f(\mathbf{r} + \mathbf{dr}) - f(\mathbf{r}) = 0 \ ,$$

or,

$$\boldsymbol{\nabla} f \bullet \mathbf{dr} = 0 \ .$$

Since $\mathbf{dr}$ is tangential to the surface, it means that $\boldsymbol{\nabla} f$ is normal (or, perpendicular) to the surface $f(\mathbf{r}) = \text{constant}$. We note in passing that in electrostatics the electric field, $\boldsymbol{E} = -\boldsymbol{\nabla}\phi$, is normal to an equipotential surface on which the electrostatic potential, ϕ, is constant. This fact is a direct consequence of the second property of the gradient of a function. The rate of variation of the scalar function f along any $arbitrary$ direction $\boldsymbol{n}$ is given by $\boldsymbol{n} \bullet \boldsymbol{\nabla} f$, which is called the $directional\ derivative$.

Divergence and Curl of a Vector Function

Because of the vector nature of the del operator, $\boldsymbol{\nabla}$, we can form the following two combinations for a vector function **A**:

$$\boldsymbol{\nabla}\bullet\,\mathbf{A}=\left(\mathbf{e_1}\frac{\partial}{\partial x_1}+\mathbf{e_2}\frac{\partial}{\partial x_2}+\mathbf{e_3}\frac{\partial}{\partial x_3}\right)\bullet(\mathbf{e_1}A_1+\mathbf{e_2}A_2+\mathbf{e_3}A_3)=\frac{\partial A_1}{\partial x_1}+\frac{\partial A_2}{\partial x_2}+\frac{\partial A_3}{\partial x_3}\qquad Eq.(10.4)$$

which is called the divergence of function **A**, and

$$\begin{aligned}\boldsymbol{\nabla}\times\mathbf{A}&=\left(\mathbf{e_1}\frac{\partial}{\partial x_1}+\mathbf{e_2}\frac{\partial}{\partial x_2}+\mathbf{e_3}\frac{\partial}{\partial x_3}\right)\times(\boldsymbol{e_1}\,A_1+\mathbf{e_2}A_2+\mathbf{e_3}A_3)\\&=\mathbf{e_1}\left(\frac{\partial A_3}{\partial x_2}-\frac{\partial A_2}{\partial x_3}\right)+\mathbf{e_2}\left(\frac{\partial A_1}{\partial x_3}-\frac{\partial A_3}{\partial x_1}\right)+\mathbf{e_3}\left(\frac{\partial A_2}{\partial x_1}-\frac{\partial A_1}{\partial x_2}\right)\qquad Eq.(10.5)\end{aligned}$$

which is called the curl of function **A**. If, for a vector function **A**, $\boldsymbol{\nabla}\bullet\mathbf{A}=0$, then **A** is called a solenoidal vector function. On the other hand, if $\boldsymbol{\nabla}\times\mathbf{A}=0$, then **A** is called an irrotational or conservative vector function. Finally, we define $\boldsymbol{\nabla}\bullet\boldsymbol{\nabla}f$, which is the divergence of the gradient of a scalar function, as the Laplacian of f. It is written as $\nabla^2 f$. Explicitly,

$$\nabla^2 f=\left(\mathbf{e_1}\frac{\partial}{\partial x_1}+\mathbf{e_2}\frac{\partial}{\partial x_2}+\mathbf{e_3}\frac{\partial}{\partial x_3}\right)\bullet\left(\mathbf{e_1}\frac{\partial f}{\partial x_1}+\mathbf{e_2}\frac{\partial f}{\partial x_2}+\mathbf{e_3}\frac{\partial f}{\partial x_3}\right)=\frac{\partial^2 f}{\partial x_1^2}+\frac{\partial^2 f}{\partial x_2^2}+\frac{\partial^2 f}{\partial x_3^2}\ .$$

Or,

$$\nabla^2 f=\sum_{i=1}^{3}\frac{\partial^2 f}{\partial x_i^2}\ .\qquad Eq.(10.6)$$

The Laplacian operator, ∇^2, can also operate on a vector function $\mathbf{A}(x_1,x_2,x_3)$ as,

$$\nabla^2\mathbf{A}=\mathbf{e_1}\nabla^2 A_1+\mathbf{e_2}\nabla^2 A_2+\mathbf{e_3}\nabla^2 A_3\ .\qquad Eq.(10.7)$$

Note carefully that since the Cartesian unit vectors $\mathbf{e_1},\mathbf{e_2},\text{and }\mathbf{e_3}$ are constant in magnitude and direction, they can be taken out of the ∇^2 operator. Finally, we prove an important identity,

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{A})=\boldsymbol{\nabla}(\boldsymbol{\nabla}\bullet\,\mathbf{A})-\nabla^2\mathbf{A}\ .\qquad Eq.(10.8)$$

To prove it, we start with the left-hand side of the identity,

$$\text{Left Hand Side} = \sum_{ijk} \epsilon_{ijk}\ \mathbf{e_i}\ \frac{\partial}{\partial x_j}\{\boldsymbol{\nabla} \times \mathbf{A}\}_k$$

$$= \sum_{ijk} \epsilon_{ijk}\ \mathbf{e_i}\ \frac{\partial}{\partial x_j}\left\{\sum_{lm} \epsilon_{klm} \frac{\partial}{\partial x_l}\ A_m\right\} = \sum_{ij}\sum_{lm} \mathbf{e_i}\ \frac{\partial}{\partial x_j}\ \frac{\partial}{\partial x_l} A_m \sum_{k} \epsilon_{kij}\epsilon_{klm}$$

$$= \sum_{ij}\sum_{lm} \mathbf{e_i} \frac{\partial}{\partial x_j}\ \frac{\partial}{\partial x_l} A_m\{\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\} = \sum_{ij} \mathbf{e_i} \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} A_j - \sum_{ij} \mathbf{e_i} \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_j} A_i$$

$$= \sum_{i} \mathbf{e_i} \frac{\partial}{\partial x_i}\sum_{j} \frac{\partial A_j}{\partial x_j} - \sum_{j} \frac{\partial^2}{\partial x_j^2}\sum_{i} \mathbf{e_i} A_i = \boldsymbol{\nabla}(\boldsymbol{\nabla}\bullet\mathbf{A}) - \nabla^2\mathbf{A}\ .$$

Here we used the epsilon-delta identity of Appendix C. The above vector identity is valid only for Cartesian components of the vector function $\mathbf{A}(x_1, x_2, x_3)$ since the unit vectors $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$ are taken to be constant in magnitude and direction.

In the remainder of this chapter, including the following examples, it will be easier to use (x, y, z) notation instead of (x_1, x_2, x_3) notation. However, for the Cartesian unit vectors we will continue to use $\mathbf{e_1}, \mathbf{e_2}, \text{and } \mathbf{e_3}$.

Example: Consider the scalar function,

$$\boldsymbol{f(x, y, z) = x^2 + y^2 - z^2 + xy - xz - yz}\ .$$

Determine $\boldsymbol{\nabla f}$, unit vector along $\boldsymbol{\nabla f}$, directional derivative along $\mathbf{e_1} + \mathbf{e_2}$, and $\boldsymbol{\nabla^2 f}$, all at $(1, 1, 1)$.

Solution: Starting with gradient of $\boldsymbol{f}$,

$$\boldsymbol{\nabla} f = \mathbf{e_1}\frac{\partial f}{\partial x} + \mathbf{e_2}\frac{\partial f}{\partial y} + \mathbf{e_3}\frac{\partial f}{\partial z}$$

$$= \mathbf{e_1}(2x + y - z) + \mathbf{e_2}\ (2y + x - z) + \mathbf{e_3}(-2z - x - y)\ .$$

$$\boldsymbol{\nabla} f \text{ at } (1,1,1) = \mathbf{e_1}(2) + \mathbf{e_2}(2) + \mathbf{e_3}(-4)\ .$$

$$\text{Unit vector along } \boldsymbol{\nabla} f \text{ at } (1,1,1) = \frac{1}{\sqrt{24}}(2\mathbf{e_1} + 2\mathbf{e_2} - 4\mathbf{e_3}) = \frac{1}{\sqrt{6}}(\mathbf{e_1} + \mathbf{e_2} - 2\mathbf{e_3})\ .$$

In order to determine the directional derivative along $\mathbf{e_1} + \mathbf{e_2}$, we first determine the unit vector along $\mathbf{e_1} + \mathbf{e_2}$, which we will label as $\boldsymbol{n}$. It is

$$\boldsymbol{n} = \frac{1}{\sqrt{2}}(\mathbf{e_1} + \mathbf{e_2}) \ .$$

So, the directional derivative is

$$\boldsymbol{n}\bullet\boldsymbol{\nabla} f = \frac{1}{\sqrt{2}}(\mathbf{e_1} + \mathbf{e_2})\bullet[\mathbf{e_1}(2x + y - z) + \mathbf{e_2}(2y + x - z) + \mathbf{e_3}(-2z - x - y)]$$

$$= \frac{1}{\sqrt{2}}[2x + y - z + 2y + x - z] = \frac{3x + 3y - 2z}{\sqrt{2}} = \frac{4}{\sqrt{2}} \quad \text{at (1,1,1)} \ .$$

Finally,

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 2 + 2 - 2 = 2 \ .$$

Example: Given the vector function

$$\mathbf{A} = \boldsymbol{x^2 y}\,\mathbf{e_1} + \boldsymbol{y^2 z}\,\mathbf{e_2} + \boldsymbol{z^2 x}\,\mathbf{e_3} \ ,$$

determine its divergence, curl, and Laplacian. Then, explicitly verify that

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{A}) = \boldsymbol{\nabla}(\boldsymbol{\nabla}\bullet\mathbf{A}) - \boldsymbol{\nabla^2}\mathbf{A} \ .$$

Solution: The divergence of **A** is

$$\boldsymbol{\nabla}\bullet\mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} = 2xy + 2yz + 2zx \ .$$

The curl of **A** is

$$\boldsymbol{\nabla} \times \mathbf{A} = \begin{vmatrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ x^2 y & y^2 z & z^2 x \end{vmatrix} = \mathbf{e_1}(0 - y^2) + \mathbf{e_2}(0 - z^2) + \mathbf{e_3}(0 - x^2) = -y^2\mathbf{e_1} - z^2\mathbf{e_2} - x^2\mathbf{e_3} \ .$$

The Laplacian of **A** is

$$\nabla^2\mathbf{A} = \mathbf{e_1}(2y) + \mathbf{e_2}(2z) + \mathbf{e_3}(2x) \ .$$

Now we verify the identity,

$$\nabla \times (\nabla \times \mathbf{A}) = \begin{vmatrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ -y^2 & -z^2 & -x^2 \end{vmatrix} = \mathbf{e_1}(2z) + \mathbf{e_2}(2x) + \mathbf{e_3}(2y) = 2[z\mathbf{e_1} + x\mathbf{e_2} + y\mathbf{e_3}]$$

$$\nabla(\nabla \bullet \mathbf{A}) = \mathbf{e_1}(2y + 2z) + \mathbf{e_2}(2x + 2z) + \mathbf{e_3}(2y + 2x) \ .$$

On substituting explicitly,

$$\nabla(\nabla \bullet \mathbf{A}) - \nabla^2 \mathbf{A} = 2z\mathbf{e_1} + 2x\mathbf{e_2} + 2y\mathbf{e_3} = \nabla \times (\nabla \times \mathbf{A}) \ .$$

Example: Find $\nabla f(r)$ where $\mathbf{r} = x\,\mathbf{e_1} + y\,\mathbf{e_2} + z\,\mathbf{e_3}$ with $r^2 = x^2 + y^2 + z^2$. Note $f(r)$ is a function of only the magnitude r of the vector r.

Solution: Starting with gradient of $f(r)$,

$$\nabla f(r) = \mathbf{e_1}\frac{\partial}{\partial x}f(r) + \mathbf{e_2}\frac{\partial}{\partial y}f(r) + \mathbf{e_3}\frac{\partial}{\partial z}f(r)$$

$$= \mathbf{e_1}\frac{\partial r}{\partial x}\frac{df(r)}{dr} + \mathbf{e_2}\frac{\partial r}{\partial y}\frac{df(r)}{dr} + \mathbf{e_3}\frac{\partial r}{\partial z}\frac{df(r)}{dr}$$

Use

$$\frac{\partial r}{\partial x} = \frac{x}{r}, \qquad \frac{\partial r}{\partial y} = \frac{y}{r}, \qquad \frac{\partial r}{\partial z} = \frac{z}{r} \ ,$$

to obtain

$$\nabla f(r) = \frac{x\,\mathbf{e_1} + y\,\mathbf{e_2} + z\,\mathbf{e_3}}{r}\frac{df}{dr} = \frac{\mathbf{r}}{r}\frac{df}{dr} \ .$$

10.2 LINE INTEGRALS

The line integral of a vector function **A** over an arbitrary path joining point 1 to point 2 is written as,

$$I = \int_1^2 \mathbf{A} \bullet \mathbf{dr} \ . \qquad Eq.(10.9)$$

It is evaluated by breaking the path, shown in Figure 10.2, into a large number of small segments of lengths $\mathbf{dr}_1, \mathbf{dr}_2, \cdots \mathbf{dr}_i \cdots \mathbf{dr}_n$. For each tiny segment, we calculate the value of the integrand, namely, $\mathbf{A}_i \bullet \mathbf{dr}_i$ and add all the values. The limiting value of this sum as the number of segments becomes very large or as the segment sizes become very small is the value of the line integral. In other words,

$$I = \int_1^2 \mathbf{A} \bullet \mathbf{dr} = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{A}_i \bullet \mathbf{dr}_i$$

From this definition of the line integrals, we can derive an important result. Recalling from Eq. (10.3) that for any scalar function

$$df = \boldsymbol{\nabla} f \bullet \mathbf{dr} \ ,$$

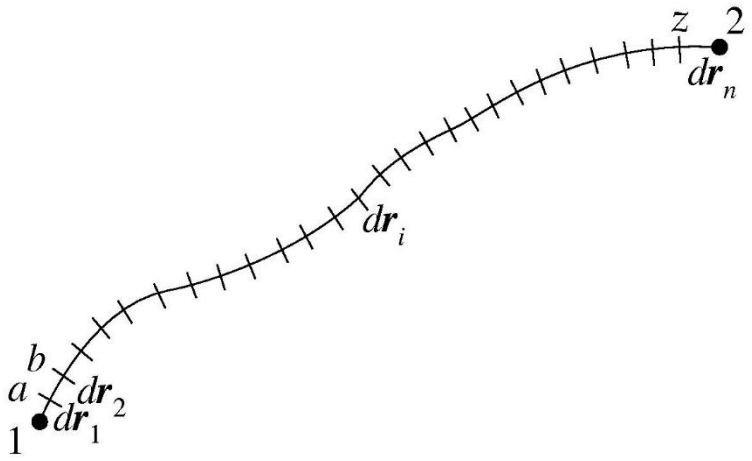


Figure 10.2. The path for a line integral.

we obtain

$$f(a) - f(1) = (\boldsymbol{\nabla} f)_1 \bullet \mathbf{dr}_1 \ ,$$

$$f(b) - f(a) = (\boldsymbol{\nabla} f)_2 \bullet \mathbf{dr}_2 \ ,$$

$$f(c) - f(b) = (\boldsymbol{\nabla} f)_3 \bullet \mathbf{dr}_3 \ ,$$

$$\ldots\ldots\ldots$$

$$f(2) - f(z) = (\boldsymbol{\nabla} f)_n \bullet \mathbf{dr}_n \ .$$

Simply adding all these lines, we get

$$f(2) - f(1) = \sum_{i=1}^{n} (\boldsymbol{\nabla} f)_i \bullet \mathbf{dr}_i \to \int_1^2 \boldsymbol{\nabla} f \bullet \mathbf{dr} \ ,$$

independent of the path joining points 1 and 2. In particular, for a closed path,

$$\oint (\nabla f) \bullet \mathbf{dr} = 0 \ , \qquad Eq.(10.10)$$

for any scalar function $f(x, y, z)$. The symbol of a circle on an integral sign indicates that the path of integration is a closed path.

10.3 SURFACE INTEGRALS

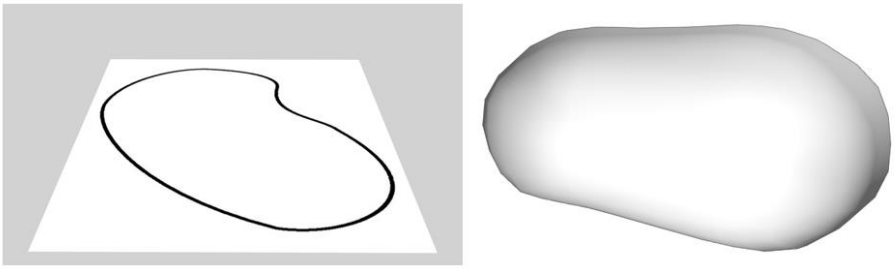


Figure 10.3a. Comparison of an open surface versus a closed surface.

In surface integrals, we encounter two different kinds of surfaces, open surfaces versus closed surfaces, as shown in Figure 10.3a. An open surface has two sides with a boundary C but no enclosed volume. An example of such a surface is a flat or a crumpled sheet of paper. In order to go from one side of an open surface to its other side, one has to cross the boundary C. A closed surface, on the other hand, has an enclosed volume but no boundary. Examples of closed surfaces include the surface of the Earth, the brown surface of an Idaho potato, etc. The surface integral of a vector function **A** over an arbitrary surface, open or closed, is written as

$$J = \int_S \mathbf{A} \bullet \mathbf{dS} = \int_S \mathbf{A} \bullet \boldsymbol{n} \, dS \ . \qquad Eq.(10.11)$$

Here $\boldsymbol{n}$ is the normal to the area element $\mathbf{dS}$. The surface integral is evaluated by breaking the whole surface area, shown in Figure 10.3b, into a large number of small pieces of areas $\Delta\mathbf{S}_1, \Delta\mathbf{S}_2, \cdots \Delta\mathbf{S}_i \cdots \Delta\mathbf{S}_n$. For each small piece, we calculate the value of the integrand, namely, $\mathbf{A}_i \bullet \Delta\mathbf{S}_i$ and add all the values. The limiting value of this sum as the number of pieces becomes very large or as the size of all pieces becomes very small is the value of the surface integral. In other words,

$$J = \int_S \mathbf{A} \bullet \mathbf{dS} = \lim_{n\to\infty} \sum_{i=1}^{n} (\mathbf{A} \bullet \boldsymbol{n})_i \, \Delta S_i \ .$$

In the case of a closed surface, this integral is called $the\ flux\ of\ vector\ function$ ***A*** out of the closed volume V.

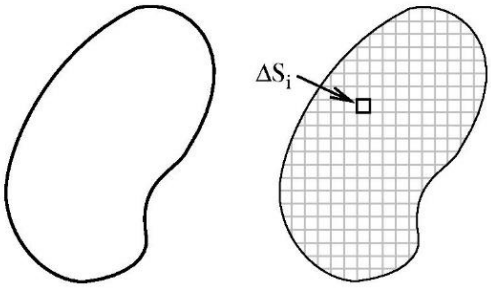


Figure 10.3b. Breaking a surface area into a large number of small area elements.

10.4 VOLUME INTEGRALS

The volume integral of a scalar function $f(x, y, z)$ over an arbitrary volume V is written as,

$$K = \int_V f(x, y, z)\, dV \ . \qquad Eq.(10.12)$$

It is evaluated by breaking the volume, shown in Figure 10.4, into a large number of small elements of volume $\Delta V_1, \Delta V_2, \cdots \Delta V_i \cdots \Delta V_n$. For each little element, we calculate the value of the integrand, namely, $f(x_i, y_i, z_i)\Delta V_i$ and add all the values. The limiting value of this sum as the number of volume elements becomes very large

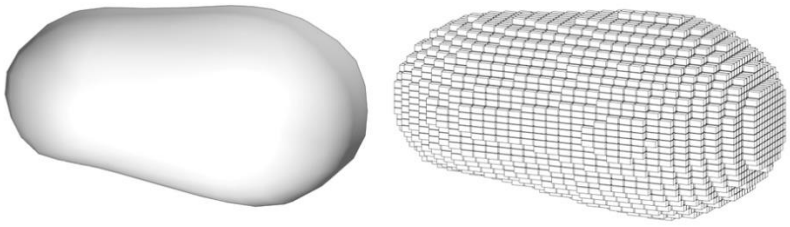


Figure 10.4. Breaking a volume into many small volume elements.

or as the element size becomes very small is the value of the volume integral. In other words,

$$K = \int_V f(x,y,z)\, dV = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i, y_i, z_i)\Delta V_i \ .$$

10.5 FLUX OF A VECTOR FUNCTION: GAUSS'S (OR DIVERGENCE) THEOREM

The flux of a vector function **A** out of a closed volume V has a very interesting property. If we break the volume V into two smaller pieces, then the sum of fluxes of **A** out of two closed smaller volumes is equal to the flux of **A** out of the original large, closed volume V. Let us prove it.

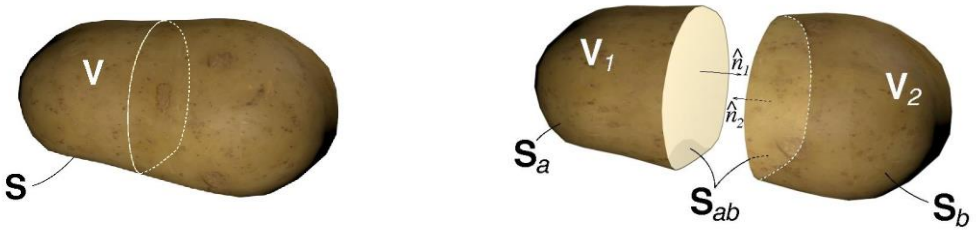


Figure 10.5. Breaking a volume V into two smaller volumes V_1 and V_2.

Consider a closed surface S enclosing a volume V, as in Figure 10.5. For clear visualization, imagine this surface to be the brown surface of an Idaho potato. In order to determine the total flux of a vector function **A** out of volume V, we first consider the flux of **A** through a small surface element $\mathbf{dS}$. It is $\mathbf{A} \bullet \boldsymbol{n}\, dS = \mathbf{A} \bullet \mathbf{dS}$. Total flux of **A** out of volume V is $\int_S \mathbf{A} \bullet \mathbf{dS}$. Now break the volume V into two pieces by cutting the potato with a knife. The two cut pieces have volumes V_1 and V_2 and the corresponding brown surface areas are S_a and S_b, respectively. Clearly, $V = V_1 + V_2$ and $S = S_a + S_b$. The cutting of the potato also exposes the interior white surface area, S_{ab}, that is common to both pieces of the potato. The volume V_1 is enclosed by surface $S_1 = S_a + S_{ab}$ and the volume V_2 is enclosed by surface $S_2 = S_b + S_{ab}$. Total flux of **A** out of volume V_1 is

$$\int_{S_a} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{S_{ab}} \mathbf{A} \bullet \boldsymbol{n}_1\, dS \ ,$$

where $\boldsymbol{n}_1$ is the outward normal on the S_{ab} part of S_1. Total flux of **A** out of volume V_2 is

$$\int_{S_b} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{S_{ab}} \mathbf{A} \bullet \boldsymbol{n}_2\, dS \ ,$$

where $\boldsymbol{n}_2$ is the outward normal on the S_{ab} part of S_2. Now, the sum of total flux of **A** out of volume V_1 and the total flux of **A** out of volume V_2 is

$$\int_{S_a} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{S_{ab}} \mathbf{A} \bullet \boldsymbol{n}_1\, dS + \int_{S_b} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{S_{ab}} \mathbf{A} \bullet \boldsymbol{n}_2\, dS \ .$$

Noting that $\boldsymbol{n}_1 = -\boldsymbol{n}_2$, the sum of fluxes out of volume V_1 and out of volume V_2 becomes,

$$\int_{s_a} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{s_b} \mathbf{A} \bullet \boldsymbol{n}\, dS = \int_{s_a + s_b} \mathbf{A} \bullet \boldsymbol{n}\; dS = \int_{s} \mathbf{A} \bullet \boldsymbol{n}\, dS \;,$$

which is equal to the total flux of **A** out of volume V. We can generalize this result by subdividing a giant volume V into a large number of smaller component volumes, of any size and shape. Then the sum of fluxes of **A** out of smaller component volumes is equal to the flux of **A** out of giant volume V. For convenience, we take each small component volume to be a parallelepiped in shape.

Flux of A Out of a Parallelepiped

Now, we will determine the flux of a vector function **A** out of a parallelepiped of side lengths $(\Delta x, \Delta y, \Delta z)$. Total volume of the parallelepiped is $\Delta V = \Delta x \Delta y \Delta z$ and its surface area ΔS consists of six faces; namely, left face, right face, front face, back face, top face, and bottom face, as shown in Figure 10.6. Total flux of **A** out of the parallelepiped, $\int_{\Delta S} \mathbf{A} \bullet \mathbf{dS}$, is the sum of six terms each representing a flux out of a different face. We will determine the flux out of three pairs of faces, the left-right pair, the front-back pair and the top-bottom pair.

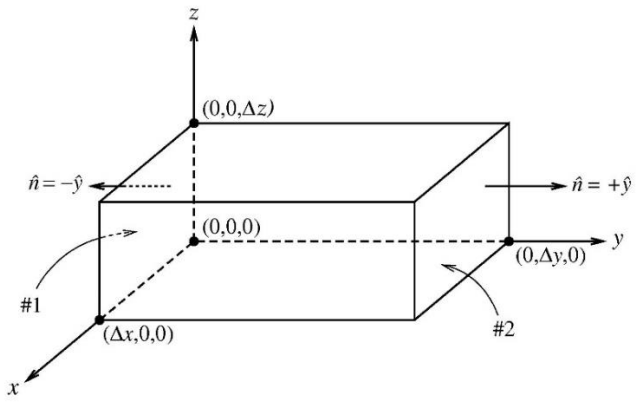


Figure 10.6. Flux of vector function **A** out of a parallelepiped.

Flux out of the left face 1 = $\int_{\#1} \mathbf{A} \bullet \boldsymbol{n}\, dS = -\int_{\#1} A_y \; dx\, dz = -A_y(1)\, \Delta x \Delta z$

Flux out of the right face 2= $\int_{\#2} \mathbf{A} \bullet \boldsymbol{n}\, dS = \int_{\#2} A_y \; dx\, dz = +A_y(2)\, \Delta x \Delta z$

Total flux out of the left-right pair of faces = $\left[A_y(2) - A_y(1)\right]\Delta x \Delta z \cong \left[\left(\frac{\partial A_y}{\partial y}\right) \Delta y\right] \Delta x \Delta z = \frac{\partial A_y}{\partial y} \Delta V$.

Similarly, total flux out of the front-back pair of faces = $\left[\left(\frac{\partial A_x}{\partial x}\right)\,\Delta x\right]\Delta y\Delta z = \frac{\partial A_x}{\partial x}\Delta V$ and total flux out of the top-bottom pair of faces = $\left[\left(\frac{\partial A_z}{\partial z}\right)\,\Delta z\right]\Delta x\Delta y = \frac{\partial A_z}{\partial z}\Delta V$.

Adding the contributions from all six faces, total flux of **A** out of volume ΔV is

$$\int_{\Delta S} \mathbf{A} \bullet \boldsymbol{n}\, dS = (\boldsymbol{\nabla} \bullet \mathbf{A})\Delta V \ .$$

Now, any arbitrary volume V can be broken into a large number of small parallelepipeds as shown in Figure 10.7, so that, in general, for a closed surface S,

Figure 10.7. Breaking an arbitrary volume into a large number of parallelepipeds.

$$\int_S \mathbf{A} \bullet \boldsymbol{n}\ dS = \int_{enclosed\ V} (\boldsymbol{\nabla} \bullet \mathbf{A}) dV \ . \qquad Eq.(10.13)$$

This is known as the Gauss's theorem or the divergence theorem for vector functions.

Example: Consider a vector function:

$$\mathbf{A} = xy\ \mathbf{e_1} + yz\ \mathbf{e_2} + zx\ \mathbf{e_3}\ ,$$

and a parallelepiped with side lengths a, b, c **and volume** $V = abc$. **Show that the divergence theorem is satisfied.**

Solution: The divergence of the vector function is

$$\boldsymbol{\nabla} \bullet \mathbf{A} = y + z + x\ .$$

The integral of this divergence over the volume of the parallelepiped of Figure 10.8 is

$$\int_V \boldsymbol{\nabla} \bullet \mathbf{A}\, dV = \int_V (x + y + z)\, dxdydz\ ,$$

$$\int_V \nabla \bullet \mathbf{A}\, dV = \int_0^a x\, dx \int_0^b dy \int_0^c dz + \int_0^a x \int_0^b y\, dy \int_0^c dz + \int_0^a dx \int_0^b dy \int_0^c z\, dz$$

$$= \frac{a^2}{2} bc + a\frac{b^2}{2} c + ab\frac{c^2}{2} = \frac{V}{2}(a+b+c)\ .$$

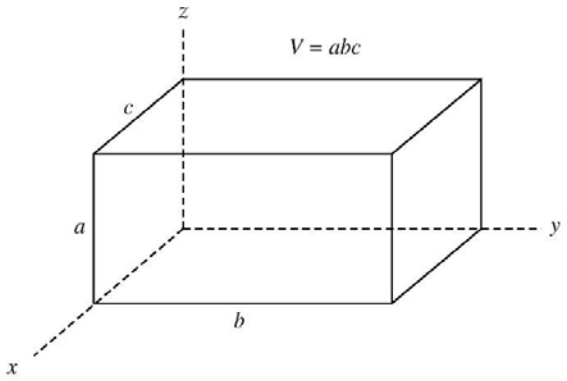


Figure 10.8. A parallelepiped with side lengths $a, b,$ and c.

The surface integral of the divergence theorem over the six faces (left-right pair, front-back pair and the top-bottom pair) of the parallelepiped is

$$\int_S \mathbf{A} \bullet \boldsymbol{n}\, dS = \int_{front} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{back} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{left} \mathbf{A} \bullet \boldsymbol{n}\, dS$$

$$+ \int_{right} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{top} \mathbf{A} \bullet \boldsymbol{n}\, dS + \int_{bottom} \mathbf{A} \bullet \boldsymbol{n}\, dS$$

$$= \int_{front} \mathbf{A} \bullet (\mathbf{e_1}) dydz + \int_{back} \mathbf{A} \bullet (-\mathbf{e_1}) dydz + \int_{left} \mathbf{A} \bullet (-\mathbf{e_2}) dxdz\ .$$

$$+ \int_{right} \mathbf{A} \bullet (\mathbf{e_2})\, dxdz + \int_{top} \mathbf{A} \bullet (\mathbf{e_3})\, dxdy + \int_{bottom} \mathbf{A} \bullet (-\mathbf{e_3})\, dxdy$$

$$= \int_0^b \int_0^c a\, y\, dydz + \int\int (-0) dydz + \int\int (-0) dzdy$$

$$+ \int_0^a \int_0^c b\, z\, dxdz + \int_0^a \int_0^b c\, x\, dxdy + \int\int (-0) dxdy$$

$$= a\frac{b^2}{2} c + ab\frac{c^2}{2} + \frac{a^2}{2} bc = \frac{V}{2}(a+b+c)\ .$$

Thus, divergence theorem:

$$\int_{six\ faces} \mathbf{A} \bullet \boldsymbol{n}\, dS = \int_{parallelepped} (\boldsymbol{\nabla} \bullet \mathbf{A}) dV \ ,$$

is satisfied.

10.6 CIRCULATION OF A VECTOR FUNCTION: STOKE'S THEOREM

Consider an open surface S whose boundary is the closed loop L as in Figure 10.9. We define the $circulation$ of a vector function **A** in closed loop L as the line integral

$$\oint_L \mathbf{A} \bullet \mathbf{dr} \ . \qquad Eq.(10.14)$$

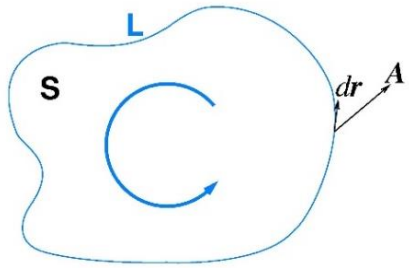


Figure 10.9. Circulation of a vector function **A** in loop L.

Here $\mathbf{dr}$ is everywhere tangential to the loop L. In general, the circulation of a vector function in a loop L has a very interesting property. Imagine that the area enclosed by loop L is broken into two smaller areas, with each area enclosed by a smaller loop. Then, the sum of circulations of a vector function **A** in two smaller loops (calculated in the same sense as in the original loop) is equal to the circulation of **A** in the large original loop L.

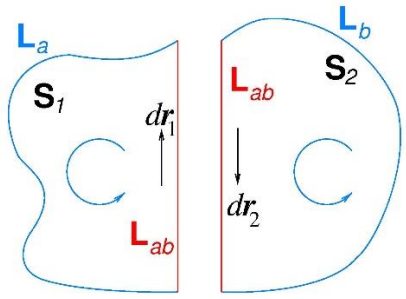


Figure 10.10. Breaking a loop L into two smaller pieces L_a and L_b.

In order to have a clear visualization, imagine that the large loop is made up of a blue color line as in Figure 10.10. This loop encloses an area S. Now, using a pair of scissors, we cut the area along the line L_{ab} to get two smaller areas S_1 and S_2. Also, during this cutting, the blue line is cut into two pieces L_a and L_b. Clearly, $S = S_1 + S_2$ and $L = L_a + L_b$. Area S_1 is enclosed by loop $L_1 = L_a + L_{ab}$ and area S_2 is enclosed by loop $L_2 = L_b + L_{ab}$. Now, circulation of **A** in loop L_1 is

$$\int_{L_a} \mathbf{A} \bullet \mathbf{dr} + \int_{L_{ab}} \mathbf{A} \bullet \mathbf{dr}_1 \ ,$$

where $\mathbf{dr}_1$ is along the line of cut L_{ab}. The circulation of **A** in loop L_2 is

$$\int_{L_b} \mathbf{A} \bullet \mathbf{dr} + \int_{L_{ab}} \mathbf{A} \bullet \mathbf{dr}_2 \ ,$$

where $\mathbf{dr}_2$ is along the line of cut L_{ab}. Thus, the sum of the circulations of **A** in loop L_1 and in loop L_2 becomes

$$\int_{L_a} \mathbf{A} \bullet \mathbf{dr} + \int_{L_{ab}} \mathbf{A} \bullet \mathbf{dr}_1 + \int_{L_b} \mathbf{A} \bullet \mathbf{dr} + \int_{L_{ab}} \mathbf{A} \bullet \mathbf{dr}_2 \ .$$

If the sense of circulations in the two smaller loops is the same, as shown in Figure 9.10, then $\mathbf{dr}_1 = -\mathbf{dr}_2$. So, the sum of circulations in loops L_1 and L_2 becomes

$$\int_{L_a} \mathbf{A} \bullet \mathbf{dr} + \int_{L_b} \mathbf{A} \bullet \mathbf{dr} = \int_{L_a + L_b} \mathbf{A} \bullet \mathbf{dr} = \int_{L} \mathbf{A} \bullet \mathbf{dr} \ ,$$

which is equal to the circulation of **A** in loop L. This result holds even when S_1 and S_2 are not *coplanar.*

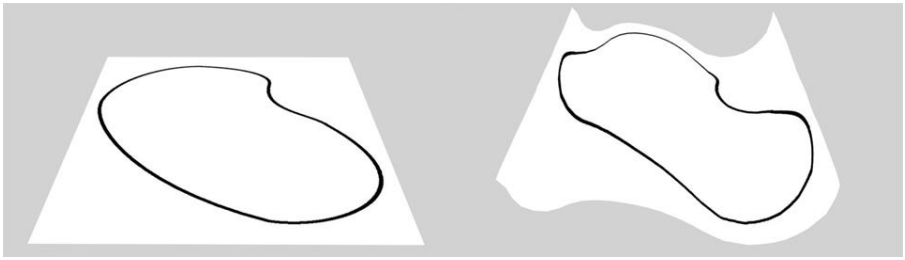


Figure 10.11. Planar versus nonplanar open surfaces.

Generalization is achieved by noting that loop L can enclose an infinite number of open surfaces. These surfaces need not be planar in shape and can be warped like a crumpled sheet of paper, as in Figure 10.11. Any one of these surfaces can be broken into a large number of smaller surface areas of any size and shape. Then the sum of circulations in loops surrounding these smaller surface areas is equal to the circulation in original loop L. For convenience, we take each of the smaller surface areas as a flat rectangle.

Circulation of A in the Loop Surrounding a Flat Rectangle

Now, we will determine the circulation of a vector function **A** in the loop surrounding a flat rectangle, in the $x-y$ plane, of side lengths Δx and Δy, as shown in Figure 10.12. The area of the loop is $\mathbf{dS} = \Delta x \Delta y \, \mathbf{e_3}$. the loop consists of four sides of the rectangle labeled 1, 2, 3, and 4. So, the circulation of **A** has four contributions,

$$\oint_{Rectangle} \mathbf{A} \bullet \mathbf{dr} = \int_1 A_x dx + \int_2 A_y dy - \int_3 A_x dx - \int_4 A_y dy \ ,$$

$$\cong A_x(1)\Delta x + A_y(2)\Delta y - A_x(3)\Delta x - A_y(4)\Delta x \ ,$$

$$= -[A_x(3) - A_x(1)]\Delta x + [A_y(2) - A_y(4)]\Delta y \ ,$$

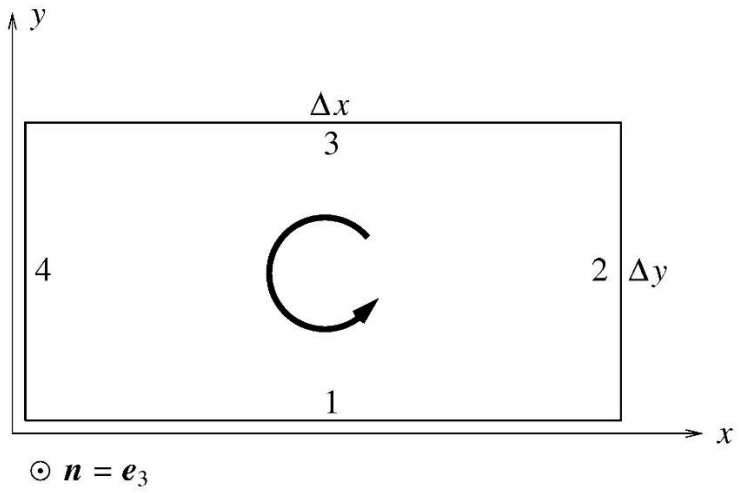


Figure 10.12. A rectangular loop in the $x-y$ plane.

or

$$\oint_{Rectangle} \mathbf{A} \bullet \mathbf{dr} = -\left[\frac{\partial A_x}{\partial y}\,\Delta y\right]\Delta x + \left[\frac{\partial A_y}{\partial x}\,\Delta x\right]\Delta y = (\boldsymbol{\nabla} \times \mathbf{A})_z \, \Delta x \Delta y = (\boldsymbol{\nabla} \times \mathbf{A}) \bullet \boldsymbol{n} \, dS \ .$$

Now, going back to any general open surface, with loop L as its boundary, that is broken into a large number of smaller surface areas in the form of flat rectangles:

$$\oint_L \mathbf{A} \bullet \mathbf{dr} = \int_{Any\ open\ surface\ bounded\ by\ L} (\boldsymbol{\nabla} \times \mathbf{A}) \bullet \boldsymbol{n} \, dS \ . \qquad Eq.(10.15)$$

This is the statement of the Stoke's theorem.

Example: Consider a vector function A $= xy\,\mathbf{e}_1 + yz\,\mathbf{e}_2 + zx\,\mathbf{e}_3$ and a square of side a in the $y-z$ plane. Show that Stoke's theorem is satisfied by the vector function A and the square loop.

Solution:

For this vector function, the curl is

$$\nabla \times \mathbf{A} = \begin{vmatrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ xy & yz & zx \end{vmatrix}$$

$$= \mathbf{e_1}\left[\frac{\partial}{\partial y}(zx) - \frac{\partial}{\partial z}(yz)\right] + \mathbf{e_2}\left[\frac{\partial}{\partial z}(xy) - \frac{\partial}{\partial x}(zx)\right] + \mathbf{e_3}\left[\frac{\partial}{\partial x}(yz) - \frac{\partial}{\partial y}(xy)\right],$$

or

$$\nabla \times \mathbf{A} = -\mathbf{e_1}y - \mathbf{e_2}z - \mathbf{e_3}x \ .$$

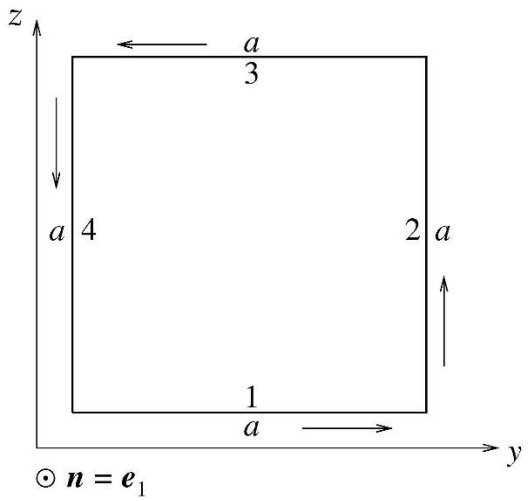


Figure 10.13. A square loop in the $y-z$ plane.

Since the square lies in the $y-z$ plane and we will traverse it in the counterclockwise sense, $\boldsymbol{n} = \mathbf{e_1}$. So,

$$\int_S (\nabla \times \mathbf{A}) \bullet \boldsymbol{n}\, dS = \int_0^a \int_0^a (-\mathbf{e_1}y - \mathbf{e_2}z - \mathbf{e_3}x) \bullet \mathbf{e_1}\, dy\, dz$$

$$= -\int_0^a y\, dy \int_0^a dz = -\frac{a^2}{2} \cdot a = -\frac{a^3}{2} \ .$$

Now, four sides of the square are labeled with numbers 1, 2, 3, and 4. Circulation of **A** in the square is

$$\int_{square} \mathbf{A} \bullet \mathbf{dr} = \int_{1+2+3+4} [xy\,dx + yz\,dy + zx\,dz] = 0 + \int_{1+2+3+4} y\,z\,dy + 0\ ,$$

since $x = 0$ everywhere on the square in the $y-z$ plane. So, circulation becomes

$$\int_{square} \mathbf{A} \bullet \mathbf{dr} = \int_1 (z = 0)\,ydy + \int_2 yz\,(dy = 0) + \int_3 (z = a)\,ydy + \int_4 yz\,(dy = 0)$$

$$= 0 + 0 + a\int_a^0 ydy + 0 = -\frac{a^3}{2}\ .$$

Thus, Stoke's theorem,

$$\oint_{Square\ loop\ L} \mathbf{A} \bullet \mathbf{dr} = \int_{Square\ surface\ bounded\ by\ L} (\boldsymbol{\nabla} \times \mathbf{A}) \bullet \boldsymbol{n}\ dS\ .$$

is satisfied.

PROBLEMS FOR CHAPTER 10

1. Determine gradient ($\boldsymbol{\nabla}$) and Laplacian (∇^2) of r and $1/r$ where $r = \sqrt{x^2 + y^2 + z^2}$.

2. Determine the divergence and curl of the vector function $\boldsymbol{F} = \frac{yz}{x}\,\mathbf{e_x} + \frac{zx}{y}\,\mathbf{e_y} + \frac{xy}{z}\,\mathbf{e_z}$.

3. If $\mathbf{r} = x\,\mathbf{e_x} + y\,\mathbf{e_y} + z\,\mathbf{e_z}$ is the position vector and $r = \sqrt{x^2 + y^2 + z^2}$ is its magnitude, then determine,

   (a) $\boldsymbol{\nabla}\bullet(r\,\boldsymbol{r})$,

   (b) $\boldsymbol{\nabla} \times (r\,\boldsymbol{r})$

   (c) $\boldsymbol{\nabla}^2(r^2)$.

4. Consider a vector function **A** = $\ln(xyz)\,\mathbf{e_x} + \ln(yz)\,\mathbf{e_y} + \ln(z)\,\mathbf{e_z}$. For this vector determine

   (a) $\boldsymbol{\nabla} \bullet \mathbf{A}$,

   (b) $\boldsymbol{\nabla} \times \mathbf{A}$,

   (c) $\nabla^2\mathbf{A}$,

   (d) $\boldsymbol{\nabla}\,(\boldsymbol{\nabla} \bullet \mathbf{A})$,

   (e) $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{A})$.

   (f) Using these results, verify $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{A}) = \boldsymbol{\nabla}\,(\boldsymbol{\nabla} \bullet \mathbf{A}) - \boldsymbol{\nabla}^2\,\mathbf{A}$.

5. Consider a vector function **A** = $(x^2 - y^2)\,\mathbf{e_x} + xyz\,\mathbf{e_y} - (x + y + z)\,\mathbf{e_z}$ and a cube bounded by the planes $x = 0, x = 1, y = 0, y = 1, z = 0$ and $z = 1$.

   (a) Determine the volume integral $\int_V \boldsymbol{\nabla} \bullet \mathbf{A}\,\mathrm{dV}$, where V is the volume of the cube.

   (b) Next determine the surface integral $\int_S \mathbf{A} \bullet \mathbf{n}\,dS$, where S is the surface area of the cube. Compare with the result of part (a)

6. Consider a vector function **A** = $(2x - y)\,\mathbf{e_x} + yz^2\,\mathbf{e_y} + y^2z\,\mathbf{e_z}$. Also, S is the flat surface area of a rectangle bounded by the lines $x = \pm 1$ and $y = \pm 2$ and C is its (rectangular) boundary in the $x - y$ plane.

   (a) Determine the line integral $\int_C \mathbf{A} \cdot \mathbf{dr}$

   (b) Next determine the surface integral $\int_S (\boldsymbol{\nabla} \times \mathbf{A}) \bullet \mathbf{n}\,dS$. Compare with the result of part (a).

7. **Biomedical Physics Application** Entomologists have known that fruit flies are attracted to sugary substances and feed on overripe fruit as well as on spilled sodas. The fruit flies approach the fruit along the direction in

which the smell of the decaying fruit is strongest. The concentration of the smell of the rotten fruit at a point (x, y, z), due to some overripe fruit at the origin (0, 0, 0), is,

$$C(x, y, z) = C_o \exp(-2\,x^2 - y^2 + z^2)\ .$$

Here x, y and z are in meters. If a fruit fly detects the presence of rotten food while it is at the point $(0.6\ m, 0.8\ m, 0.9\ m)$, determine the unit vector along the initial direction in which the fruit fly approaches the food.

Chapter 11: First Order Differential Equations

In our previous discussion of calculus, we were given a known function $F(x)$ and were asked to find either the derivative or the integral of this function. In some other situations, we are given the first, second or higher order derivative of an unknown function and are asked to figure out the function. That brings us to the domain of differential equations. Our aim in this chapter is to provide merely a flavor of differential equations, so we will confine our attention only to first order ordinary differential equations. The prolific topics related to differential equations, ordinary as well as partial, are covered in books which are entirely devoted to this subject.

11.1 A FIRST ORDER DIFFERENTIAL EQUATION

Consider the case where an unknown function $y(x)$, which is yet to be determined, satisfies,

$$\frac{dy}{dx} = F(x) \quad \text{or} \quad \frac{dy}{dx} = G(y) \quad \text{or} \quad \frac{dy}{dx} = H(x,y) \ , \qquad Eq.(11.1)$$

where the functional form of $F(x)$ or $G(y)$ or $H(x,y)$ is known. Such equations are called differential equations. The *order* of a differential equation is the order of highest derivative appearing in the equation. An $ordinary$ differential equation involves derivatives of a function of a single independent variable. All equations shown in Eq. (11.1) are first order ordinary differential equations. A $partial$ differential equation, on the other hand, involves derivative of a function of multiple variables. The wave equation, that we encountered in Chapter 1, serves as an example of a partial differential equation since it contains the following derivatives,

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2} \ .$$

The wave equation is a second order equation. When a differential equation for the function $y(x)$ is integrated, a constant of integration, C, will appear naturally as a part of integration. This constant will need to be evaluated using the constraints on the function y at some fixed value of variable x. Such constraining relationships are called the boundary conditions. If the variable in the differential equation is time, t, the boundary conditions are referred to as the initial conditions.

Several first order differential equations can be solved by direct integration. The following examples will illustrate this point.

Example: Solve the differential equation

$$\frac{dy}{dx} = \frac{x}{\sqrt{1-x^2}} \qquad |x| \le 1 \ , \qquad Eq.(11.2)$$

for the function $y(x)$ with the boundary condition $y(1) = 0$.

Solution: On integrating both sides of the differential equation (11.2) with respect to variable x, we get

$$y = \int \frac{x dx}{\sqrt{1-x^2}} + C \ .$$

To do the integral, substitute $1 - x^2 = u^2$, or $x\, dx = -u\, du$,

$$\int \frac{x dx}{\sqrt{1-x^2}} = -\int \frac{u\, du}{u} = -u = -\sqrt{1-x^2} \ .$$

So, $y = -\sqrt{1-x^2} + C$ is the solution of the differential equation and the boundary condition gives $C = 0$.

Example: Solve the differential equation for $y(x)$,

$$\frac{dy}{dx} = 1 + y^2 \ , \qquad Eq.(11.3)$$

with boundary condition $y(0) = 0$.

Solution: Since we have to solve for function y in terms of variable x, we separate out the x and y dependent parts in Eq. (11.3) and then integrate to get

$$\int \frac{dy}{1+y^2} = \int dx + C \ .$$

Or

$$\arctan y = x + C \ .$$

Or $y = \tan(x + C)$ is the solution of the differential equation. The boundary condition implies $C = 0$.

In some problems $\frac{dy}{dx}$ can be written in a separable form as

$$\frac{d}{dx} = g(x)h(y) \ .$$

Then,

$$\int \frac{dy}{h(y)} = \int g(x)dx + C \ .$$

Example: Solve the following differential equation for $y(x)$,

$$\frac{dy}{dx} = \frac{9\exp x}{y^2} \ , \qquad Eq.(11.4)$$

with $y(0) = 3$.

Solution: Write Eq. (11.4) in separable form and then integrate to get

$$\int y^2 dy = \int 9\exp x \, dx + \ C \ .$$

Or,

$$\frac{y^3}{3} = 9\exp x + C \ .$$

On substituting boundary condition, $C = 0$. Thus, the function $y(x)$ is,

$$y = 3\exp\left(\frac{x}{3}\right) .$$

11.2 INTEGRATING FACTOR

Some first order differential equations are of the form

$$a\frac{dy}{dx} + by = F(x) \ , \qquad Eq.(11.5a)$$

where a and b are constants. These equations are solved by introducing an $integrating\ factor$. The integrating factor for equations of this type is

$$\exp\left(\frac{b}{a}x\right) .$$

First, we rewrite Eq. (11.5a) after dividing by constant a,

$$\frac{dy}{dx} + \frac{b}{a}y = \frac{1}{a}F(x) \ . \qquad Eq.(11.5b)$$

On multiplying both sides of Eq. (11.5b) by the integrating factor, the left-hand side becomes an exact differential,

$$\frac{d}{dx}\left[y\exp\left(\frac{b}{a}x\right)\right]=\frac{\exp\left(\frac{b}{a}x\right)}{a}F(x)\ .$$

Integrate both sides of this equation with respect to x to obtain the solution

$$y\exp\left(\frac{b}{a}x\right)=\frac{1}{a}\int\exp\left(\frac{b}{a}x\right)\,F(x)dx+C\ , \qquad Eq.(11.6)$$

where C, the constant of integration, will be determined by the boundary conditions.

Example: $An\ R-L\ circuit$. The circuit consisting of an inductor of inductance L and a resistor of resistance R is connected to a source of alternating current of frequency ω and voltage $V_0\cos(\omega t)$. Using Kirchhoff's rule in physics, the current $I(t)$ in the circuit satisfies the first order differential equation

$$L\frac{dI}{dt}+RI=V_0\cos(\omega t)\ . \qquad Eq.(11.7a)$$

Determine the current $I(t)$ as a function of time t.

Solution: In this case the differential Eq. (11.7a) is of the same form as Eq. (11.5a). So, the integrating factor is $\exp\left(\frac{R}{L}t\right)$. On multiplying the differential Eq. (11.7a) by the integrating factor and dividing by L, it becomes

$$\frac{d}{dt}\left[\exp\left(\frac{R}{L}t\right)\,I(t)\right]=\frac{V_0}{L}\exp\left(\frac{R}{L}t\right)\cos(\omega t)\ .$$

On integrating both sides of this equation with respect to t using the standard integral,

$$\int\exp(at)\cos(bt)\ dt\ =\frac{\exp(at)}{[a^2+b^2]}\ [\mathrm{a}\cos(bt)+b\sin(bt)]\ ,$$

we obtain

$$\exp\left(\frac{R}{L}t\right)I(t)=\frac{V_0}{L}\frac{\exp\left(\frac{Rt}{L}\right)}{\left[\left(\frac{R}{L}\right)^2+\omega^2\right]}\left[\frac{R}{L}\cos(\omega t)+\omega\sin(\omega t)\right]+C\ .$$

The constant of integration C is determined from the initial condition. If, for example, at $t=0, I=0$, then

$$C=-\frac{V_0}{L}\frac{R}{L}\frac{1}{\left[\left(\frac{R}{L}\right)^2+\omega^2\right]}\ ,$$

so that, finally,

$$I(t) = \frac{V_0}{L}\frac{1}{\left[\left(\frac{R}{L}\right)^2 + \omega^2\right]}\left[\frac{R}{L}\cos(\omega t) + \omega\sin(\omega t)\right] - \frac{V_0}{L}\frac{R}{L}\frac{\exp\left(-\frac{Rt}{L}\right)}{\left[\left(\frac{R}{L}\right)^2 + \omega^2\right]} \quad . \qquad Eq.(11.7b)$$

Example: *Molecular Isomerization*. Some molecules with the same number of identical atoms can exist in several stable configurations or arrangements of their component atoms. Such configurations are called different isomers of the same molecule. For example, the molecule of propane-based alcohol, commonly called propanol, has two common isomers. The first isomer, 1-propanol, a primary alcohol, and the second isomer, 2-propanol, a secondary alcohol, have structural formulas as shown in the Figure 11.1.

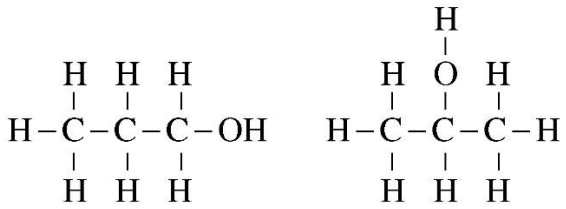


Figure 11.1. Two isomers of propanol.

The propanol molecule is of significant importance in biological physics. Determine the ratio of populations of two isomers of the propanol molecule in equilibrium.

Solution: A typical molecule isomerizes (or flips) between two configurations M_1 and M_2. The probability that the molecule in configuration M_2 flips into M_1 in time Δt is proportional to Δt—the longer the time period Δt, the higher the probability. Thus, the flipping probability is $k_+\Delta t$. The constant of proportionality k_+, which represents the probability per unit time, is called the reaction rate. Stated differently, the rate of change of a single molecule from configuration M_2 to configuration M_1 is k_+. If there are N_2 molecules in configuration M_2 then rate of change of all molecules from M_2 to M_1 is k_+N_2. Similarly, probability that the molecule in configuration M_1 flips into M_2 in time Δt is $k_-\Delta t$, where k_- is the reaction rate for the reverse isomerization. The complete isomerization process can be expressed as

$$M_2 \underset{k_-}{\overset{k_+}{\rightleftharpoons}} M_1$$

Note that in this isomerization process a single molecule of first configuration is converted into a single molecule of second configuration. Thus, the total number of molecules, $N_{tot} = N_1 + N_2$, stays the same at all times.

Assume that in the beginning, at $t = 0, N_2 = N_{tot}$ and $N_1 = 0$, which is the initial condition. The rates of change of N_1, the number of molecules with configuration M_1, and of N_2, the number of molecules with configuration M_2, are

$$\frac{dN_2}{dt} = -k_+ N_2 + k_- N_1 = -(k_+ + k_-)N_2 + k_- N_{tot} \ ,$$

$$\frac{dN_1}{dt} = +k_+ N_2 - k_- N_1 = -(k_+ + k_-)N_1 + k_+ N_{tot} = -\frac{dN_2}{dt} \ .$$

Thus, the first-order differential equation satisfied by $N_2(t)$ is

$$\frac{dN_2}{dt} + (k_+ + k_-)N_2 = k_- N_{tot} \ , \qquad Eq.(11.8a)$$

which is of the same form as Eq. (11.5b). The integrating factor of this equation is $\exp[(k_+ + k_-)t]$. Therefore,

$$\frac{d}{dt}\{N_2 \exp[(k_+ + k_-)t]\} = k_- N_{tot} \exp[(k_+ + k_-)t] \ ,$$

which on integration gives

$$N_2(t) = \frac{k_- N_{tot}}{(k_+ + k_-)} + C_2 \exp[-(k_+ + k_-)t] \ . \qquad Eq.(11.8b)$$

Here C_2 is the constant of integration. Similarly,

$$N_1(t) = \frac{k_+ N_{tot}}{(k_+ + k_-)} + C_1 \exp[-(k_+ + k_-)t] \ , \qquad Eq.(11.8c)$$

where C_1 is the constant of integration. Using the initial condition, we get

$$C_1 = -C_2 = -\frac{k_+ N_{tot}}{(k_+ + k_-)} \ .$$

Thus

$$N_2(t) = \frac{k_- N_{tot}}{(k_+ + k_-)} + \frac{k_+ N_{tot}}{(k_+ + k_-)} \exp[-(k_+ + k_-)t]$$

and

$$N_1(t) = \frac{k_+ N_{tot}}{(k_+ + k_-)}[1 - \exp[-(k_+ + k_-)t]] \ .$$

Note $N_1(t) + N_2(t) = N_{tot}$ at all times. Also, $N_2(0) = N_{tot}$, $N_1(0) = 0$. Finally, after a sufficiently long time $[t \to \infty]$, when equilibrium is attained,

$$N_2(\infty) = \frac{k_- N_{tot}}{(k_+ + k_-)} \ , \qquad N_1(\infty) = \frac{k_+ N_{tot}}{(k_+ + k_-)} \ .$$

Thus, at equilibrium

$$N_{1,eq} = \frac{k_+}{k_+ + k_-} N_{tot}, \qquad N_{2,eq=} \frac{k_-}{k_+ + k_-} N_{tot} \text{ and } \frac{dN_2}{dt} = 0 \text{ as well as } \frac{dN_1}{dt} = 0 \ .$$

Also, note that at equilibrium,

$$\frac{N_{1,eq}}{N_{2,eq}} = \frac{k_+}{k_-} \ ,$$

that is, the ratio of population of two isomers is the same as the ratio of two rates.

Example: $Newton's\ Law\ of\ Cooling$. It is a common experience that if a glass of water at room temperature is placed outdoors on a hot day in summer, the temperature of water starts rising quickly. Similarly, if the same glass of water at room temperature is placed outdoors during a cold wintry night when the temperature has plunged due to the polar vortex, the temperature of water in the glass starts falling rapidly. Newton formulated the Law of Cooling by stating that the rate of change of temperature of a body, at any time t, is proportional to the difference between the constant temperature of the body's surroundings, T_s, and the temperature of the body, $T(t)$, at time t. Mathematically,

$$\frac{dT(t)}{dt} = C\,[T_s - T(t)] \ ,$$

where C $(C > 0)$ is the constant of proportionality. This is a first-order differential equation whose solution provides the temperature of a body, versus the constant temperature of its surroundings, as a function of time. Determine an explicit expression for the temperature of the body, *T*(*t*), at any time *t*.

Solution: The differential equation to be solved is

$$\frac{dT(t)}{dt} + C\,T(t) = C\,T_s \ . \qquad Eq.(11.9a)$$

If we define a new function, $u(t)$, as $u(t) = T(t) - T_s$, then

$$\frac{du(t)}{dt} = -C\,u(t) \quad \text{or} \quad \frac{du(t)}{u} = -C\,dt \ . \qquad Eq.(11.9b)$$

A straightforward integration gives

$$\ln u(t) = -C\ t + k\ ,$$

where k is the constant of integration. This equation can be rewritten as

$$u(t) = T(t) - T_s = \exp(-C\ t + k)\ ,$$

and can be rearranged as

$$T(t) = T_s + \exp(-C\ t + k)\ .$$

The constant of integration, k, can be obtained by setting $t = 0$, to get

$$\exp(k) = T(0) - T_s\ ,$$

so that, finally,

$$T(t) = T_s + [T(0) - T_s]\exp(-C\ t)\ . \qquad Eq.(11.9c)$$

Note that at a much later time, that is, $t \to \infty$, $T(\infty) = T_s$ and the temperature of the body becomes equal to the temperature of its surroundings, as expected.

11.3 APPLICATIONS OF NUCLEAR PHYSICS

Without any doubt, nuclear physics has made tremendous contributions to modern medicine. Radiology and nuclear medicine involve the use of radioactive nuclei to diagnose, evaluate, and treat various diseases. The nucleus of any atom contains protons and neutrons. A single nuclear species, having unique values of the number of protons and number of neutrons, is called a nuclide. About 80% of the nuclides occurring on the Earth are stable nuclides; that is, they do not disintegrate, or decay, into other nuclides. On the other hand, about 20% of the nuclides are unstable, or radioactive, nuclides which can decay into other stable or unstable nuclides. In a sample of radioactive material, the number of radioactive nuclei continues to decrease with time as some of them disintegrate into other nuclides. In a radioactive transformation, the disintegrating nuclides are called $parent$ nuclides and they decay into so-called $daughter$ nuclides. Generally, the rate at which the parent nuclides decay varies from nuclide to nuclide. If $N(t)$ is the number of radioactive nuclides in a sample, then the number dN of nuclides that will undergo disintegration in time dt will be proportional both to dt as well as to $N(t)$. Saying this differently, the longer the time period dt, the more decays occur, and the larger the number N of nuclides available to decay, the more decays occur. Thus,

$$dN = -\lambda\ N\ dt\ ,$$

where the negative sign signifies a decrease in the number N. The constant of proportionality λ is called the decay constant and it is different for different nuclides. Therefore,

$$\frac{dN}{dt} + \lambda\, N = 0 \ . \qquad Eq.(11.10)$$

Example: $Radioactivity$. In a sample of radioactive material, the number of nuclides at $t = 0$ is $N(0)$. Determine $N(t)$, the number of radioactive nuclides later at time t. Also, determine the time at which the number of radioactive nuclides has decreased to half of the number at $t = 0$. This time is called the half-life, $T_{1/2}$, of the radioactive sample.

Solution: The first order differential equation satisfied by $N(t)$ is

$$\frac{dN}{dt} = -\lambda\, N \ ,$$

which can be rewritten as

$$\frac{dN}{N} = -\lambda\, dt \ .$$

A simple integration of this equation gives

$$\ln N(t) = -\lambda\, t + C \ .$$

Here C is the constant of integration. Its value can be obtained using the initial condition that at $t = 0, N = N(0)$. So, $C = \ln N(0)$. Therefore,

$$\ln N = -\lambda\, t + \ln N(0) \ ,$$

or

$$N(t) = N(0)\exp(-\lambda\, t) \ . \qquad Eq.(11.11)$$

To determine the half-life, note that at $t = T_{1/2}$, we have $N = N(0)/2$, so that

$$\frac{1}{2} = \exp\left(-\lambda\, T_{1/2}\right) \ ,$$

or

$$T_{1/2} = \frac{ln\, 2}{\lambda} \ . \qquad Eq.(11.12)$$

Thus, starting from the original number $N(0)$, the number of remaining radioactive nuclides after each successive time interval of $T_{1/2}$ are $\frac{N(0)}{2}, \frac{N(0)}{4}, \frac{N(0)}{8}$, … and so on.

Example: $Successive\ Radioactivity$. A radioactive nuclide A decays into a daughter nuclide B with a decay constant of λ_A. The daughter nuclide B itself is radioactive and it decays into a stable nuclide C with a decay constant of λ_B. Schematically, the nuclear reaction can be expressed as

$$A \xrightarrow{\lambda_A} B \xrightarrow{\lambda_B} C \ .$$

If at the beginning, that is, at $t = 0$, the number of nuclides is, $N_A(0) = N_0$ and $N_B(0) = 0$, determine the number of nuclides later at time t.

Solution: Following the previous example, since the parent A decays only into the daughter B, the first order differential equation for $N_A(t)$ and its solution are

$$\frac{dN_A}{dt} = -\lambda_A N_A \ , \qquad Eq.(11.13a)$$

and

$$N_A(t) = N_0 \exp(-\lambda_A t) \ . \qquad Eq.(11.13b)$$

The daughter B is produced by decay of parent A and is lost by its own decay into C. So, the first order differential equation for $N_B(t)$ is

$$\frac{dN_B(t)}{dt} = \lambda_A N_A - \lambda_B N_B \ . \qquad Eq.(11.14)$$

The first term on the right-hand side represents the rate of production of B and the second term represents the rate of decay of B. Substituting $N_A(t)$ from Eq. (11.13b) into the right-hand side of Eq. (11.14), we get

$$\frac{dN_B(t)}{dt} + \lambda_B N_B = \lambda_A N_0 \exp(-\lambda_A t) \ . \qquad Eq.(11.15)$$

This differential equation is of the same form as Eq. (11.5b). In this case, the integrating factor is $\exp(\lambda_B t)$ and the solution of Eq. (11.15) is same as in Eq. (11.6), namely,

$$N_B(t) \ \exp(\lambda_B t) = \lambda_A N_0 \int \exp[(\lambda_B - \lambda_A)t] \ dt + C = \frac{\lambda_A N_0}{\lambda_B - \lambda_A} \exp[(\lambda_B - \lambda_A)t] + C \ ,$$

where C is the constant of integration. Using the initial condition for $N_B(t)$ at $t = 0$,

$$C = -\frac{\lambda_A N_0}{\lambda_B - \lambda_A} ,$$

so that

$$N_B(t) = \frac{\lambda_A N_0}{\lambda_B - \lambda_A} [\exp(-\lambda_A t) - \exp(-\lambda_B t)] . \qquad Eq. (11.16)$$

Eq. (11.16) provides the number of daughter nuclides as a function of time. On comparing the half-life of parent A with the half-life of daughter B, we arrive at two conclusions. First, if the half-life of the parent is much larger than the half-life of the daughter, it implies, using Eq. (11.12), that λ_A is much smaller than λ_B. Then, $\lambda_B - \lambda_A \approx \lambda_B$ and $\exp(-\lambda_A t)$ is much larger than $\exp(-\lambda_B t)$, so that

$$N_B(t) = \frac{\lambda_A N_0}{\lambda_B} \exp(-\lambda_A t) .$$

Using Eq. (11.13b), it means that

$$\lambda_A N_A(t) = \lambda_B N_B(t) \qquad \text{or} \qquad \frac{dN_A}{dt} = \frac{dN_B}{dt} . \qquad Eq. (11.17)$$

So, after a sufficiently long time, both the parent and daughter nuclides disintegrate at the same rate. Second, we conclude that if the half-life of the parent is much shorter than the half-life of the daughter, no equilibrium can be reached since the parent nuclide will disintegrate quickly, leaving behind a much longer-lived daughter nuclide.

PROBLEMS FOR CHAPTER 11

1. Determine the function $u(t)$ satisfying the differential equation

$$\frac{du}{dt} = \frac{t^2 + 1}{\cos u}$$

and $u(0) = \pi/2$.

2. Determine the function $y(x)$ satisfying

$$\frac{dy}{dx} = \frac{1}{2}\frac{y}{x}$$

with $y(4) = 1$.

3. On a calculus exam, students were asked to use the product rule to determine the derivative of a function $F(x) = f(x)\, g(x)$. A student forgot the product rule but mistakenly assumed that product rule was,

$$\frac{dF}{dx} = \frac{df(x)}{dx}\frac{dg(x)}{dx} .$$

Luckily, the student got the correct answer.

If one of the functions in the product was, $g(x) = \exp(3x)$, then set up a a first-order differential equation for the other function $f(x)$.

Solve the differential equation to obtain the second function $f(x)$.

4. **Biomedical Physics Application**. A glucose solution is administered intravenously into the bloodstream at a constant rate *r*. As the glucose is added, it is converted into other substances and removed from the bloodstream at a rate that is proportional to the concentration at that time. Thus, a model for the concentration *C*(*t*) of the glucose solution in the bloodstream is

$$\frac{dC}{dt} = r - kC$$

where *k* is a positive constant.

(a) Determine the concentration at any time *t* if the initial concentration at *t* = 0 is C_o.

(b) Determine the value of the concentration at sufficiently long time ($t \to \infty$) after the administration begins. How will this value change if C_o is doubled?

5. **Biomedical Physics Application**. A device called a pacemaker can be implanted inside the human body to monitor the heart activity. The pacemaker is essentially a capacitor, with capacitance C, that stores some charge

at a certain voltage V. If the heart skips a beat, the pacemaker releases the stored energy through a lead wire of resistance R to get the heart back to beating normally. In a pacemaker circuit, the charge, $Q(t)$, on the plates of the capacitor varies with time as

$$R\frac{dQ}{dt} + \frac{Q}{C} = V\exp(-\lambda t)$$

where $\lambda = 2/(RC)$. If at $t = 0, Q(t = 0) = CV$, solve the above differential equation to obtain $Q(t)$ as a function of time, t.

6. **Biomedical Physics Application**. On a clear wintry night, at midnight, when the outdoor temperature is $0°C$, the body of an old man was discovered on a park bench. The temperature of the body at the time of discovery was determined to be $28°C$. The body temperature of a normal healthy adult is $37°C$. It took an hour to take the body from the park to the nearby hospital. At 1:00 AM, the pathologist doing the autopsy finds the temperature of the body to be $24°C$ and declares the cause of death to be hypothermia. Based on this information, determine the time of death for the old man.

7. **Biomedical Physics Application**. In medical research, radioactive materials are helpful in identifying and treating diseases in a noninvasive manner. In a chain of successive radioactivity, a radioactive nuclide A decays into a daughter nuclide B and the daughter nuclide B decays into a stable nuclide C. The decay constants (or half-lives) of both the parent nuclide and the daughter nuclide are equal. If at $t = 0$ the number of nuclides is $N_A(0) = N_0$ and $N_B(0) = 0$, then determine the time when the population of the daughter nuclide, B, is largest. Express this time in terms of the half-life of either nuclide.

Chapter 12: Diffusion Equation

Let us start our discussion of diffusion through some everyday observations. Imagine placing a sugar cube in a glass of water. After a few hours, you notice that sugar cube is no longer there; it has become part of the water and now the water tastes sweet. The sugar cube has dissolved due to diffusion. This process is similar to the process of gas exchange in multicellular organisms. Other examples of diffusion include noticing the smell of spray perfume or cigarette smoke even when the person using perfume or smoking is far away from the observer.

12.1 FICK'S LAW

The process of diffusion refers to the movement of a substance, or more accurately its atoms and molecules, from a region of higher concentration to a region of lower concentration. Diffusion is important in biology in transporting biomolecules from one location to another. Thermal energy can spontaneously move the molecules, and this spontaneous motion is governed by the diffusion equation. Diffusion plays a role in cellular transport and membrane permeability, and in determining conformation of certain large biomolecules, along with other applications. The rate of enzymatic reaction in many cases is also diffusion-limited as it determines the time it will take the reactant molecules to come closer and interact with each other.

To begin with, let us consider the flow of diffusing particles back and forth in one dimension (say, along the x-axis) across a surface S during a time t. The number of particles flowing across is directly proportional to time t and to area S. We define net flux, J_x, as the number of particles flowing per unit area per unit time. It was empirically noted by Fick that the net flux is proportional to the rate at which the concentration of particles varies with distance (that is, the gradient of the concentration). In other words, Fick's law states

$$J_x = -D\frac{\partial C}{\partial x} \ ,$$

where C is the concentration (that is, number of particles per unit volume). The negative sign is indicative of the direction of the flow of particles. The direction of flux, J_x, is the direction of increasing x and of decreasing concentration, C. The constant of proportionality D is called the diffusion constant and it has dimensions of $[\mathrm{L}^2/T]$. Similarly, the flow of particles along y and z directions satisfy

$$J_y = -D\frac{\partial C}{\partial y} \ ,$$

$$J_z = -D\frac{\partial C}{\partial z} \ .$$

So, in three dimensions, we can write Fick's law as

$$\boldsymbol{J} = -D\,\boldsymbol{\nabla}\,C \ . \qquad Eq.(12.1)$$

Equation of Continuity

In order to study the diffusion of particles in three-dimensional space, let us break the whole space into tiny boxes. Over time, as particles in the whole space diffuse from one box to another, the total number of particles in the whole space remains fixed. In other words,

$$N = \int_{whole\ space} C(x,y,z,t)dxdydz \qquad Eq.(12.2)$$

is constant, independent of time.

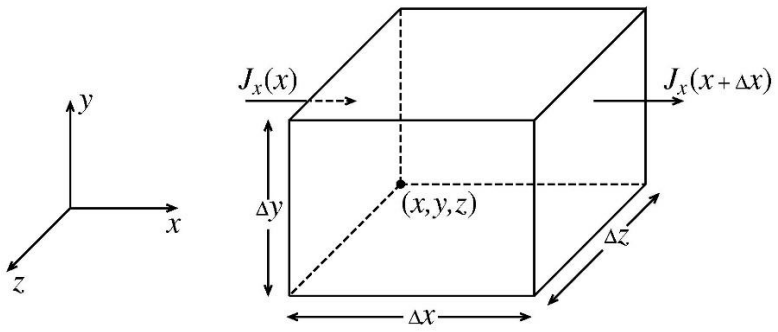


Figure 12.1. Particles diffusing into and out of a box.

Let us consider flow of particles through a box of sides $\Delta x, \Delta y$ and Δz between times t and $t + \Delta t$. The volume of this box is $\Delta V = \Delta x \Delta y \Delta z$. The concentration of particles in the box at time t is $C(x,y,z,t)$ and the number of particles in the box at time t is $n(t) = C(x,y,z,t)\Delta V$. Similarly, the concentration of particles in the box at a later time $t + \Delta t$ is $C(x,y,z,t+\Delta t)$ and the number of particles in the box at later time $t + \Delta t$ is $n(t+\Delta t) = C(x,y,z,t+\Delta t)\Delta V$. Now using the definition of flux as the particles flowing per unit area per unit time, we get

$$[J_x(x+\Delta x) - J_x(x)]\Delta t\Delta y\Delta z + [J_y(y+\Delta y) - J_y(y)]\Delta t\Delta x\Delta z + [J_z(z+\Delta z) - J_z(z)]\Delta t\Delta x\Delta y$$
$$= -\{n(t+\Delta t) - n(t)\} \ .$$

The negative sign on the right-hand side signifies that if more particles diffuse out of the box than diffuse into the box, then the number of particles in the box at the later time will be less than at the earlier time. Similarly, if more particles diffuse into the box than out of the box, then the number of particles in the box at the later time will be more than at the earlier time. Now, keeping only first-order terms in the difference of fluxes, we get

$$\left[\left(\frac{\partial J_x}{\partial x}\right)\Delta x\right]\Delta t\Delta y\Delta z + \left[\left(\frac{\partial J_y}{\partial y}\right)\Delta y\right]\Delta t\Delta x\Delta z + \left[\left(\frac{\partial J_z}{\partial z}\right)\Delta z\right]\Delta t\Delta x\Delta y = -\{C(x,y,z,t+\Delta t) - C(x,y,z,t)\}\Delta V \ .$$

On dividing both sides by Δt and ΔV, we get

$$\frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z} = -\frac{C(x,y,z,t+\Delta t) - C(x,y,z,t)}{\Delta t} = -\frac{\partial C}{\partial t} \ ,$$

or

$$\boldsymbol{\nabla \cdot J} = -\frac{\partial C}{\partial t} \ . \qquad Eq.(12.3)$$

This is the equation of continuity. Physically, it represents the conservation of total number of particles, namely N, independent of time, as in Eq (12.2).

12.2 DIFFUSION EQUATION IN THREE DIMENSIONS

Combining the equation of continuity with Fick's law gives the diffusion equation,

$$D\,\boldsymbol{\nabla^2} C = D\left[\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2}\right] = \frac{\partial C}{\partial t} \ . \qquad Eq.(12.4)$$

This is a time-dependent second order partial differential equation. Solving such an equation is beyond the scope of this book. The solution of the diffusion equation provides the time dependence of the concentration of the particles at any point in space.

Solution of Diffusion Equation

The solution of the diffusion equation, after solving the second order partial differential equation, is

$$C(x,y,z,t) = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3}\exp\left[-\frac{x^2+y^2+z^2}{4Dt}\right] \ . \qquad Eq.(12.5a)$$

Let us verify that this indeed is the solution of the diffusion equation. The first and second derivatives of C with respect to x are

$$\frac{\partial C}{\partial x} = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3}\exp\left[-\frac{x^2+y^2+z^2}{4Dt}\right]\left(-\frac{x}{2Dt}\right) \ ,$$

$$\frac{\partial^2 C}{\partial x^2} = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3}\left\{\exp\left[-\frac{x^2+y^2+z^2}{4Dt}\right]\left(-\frac{x}{2Dt}\right)^2 + \exp[\cdots]\left(-\frac{1}{2Dt}\right)\right\} \ ,$$

or, with $r^2 = x^2 + y^2 + z^2$,

$$\frac{\partial^2 C}{\partial x^2} = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right]\left\{\frac{x^2}{(2Dt)^2} - \frac{1}{2Dt}\right\} .$$

Similarly, the second derivatives with respect to variables y and z are

$$\frac{\partial^2 C}{\partial y^2} = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right]\left\{\frac{y^2}{(2Dt)^2} - \frac{1}{2Dt}\right\} ,$$

and

$$\frac{\partial^2 C}{\partial z^2} = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right]\left\{\frac{z^2}{(2Dt)^2} - \frac{1}{2Dt}\right\} .$$

Thus, combining all the derivatives together,

$$\boldsymbol{\nabla^2} C = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right]\left\{\frac{r^2}{(2Dt)^2} - \frac{3}{2Dt}\right\} .$$

Also, the derivative of C with respect to time, t, is

$$\frac{\partial C}{\partial t} = \exp\left[-\frac{r^2}{4Dt}\right] \frac{N}{\left(\sqrt{4\pi D}\right)^3} \left\{\frac{1}{\left(\sqrt{t}\right)^3}\frac{r^2}{4Dt^2} - \frac{3}{2}\frac{1}{\left(\sqrt{t}\right)^5}\right\}$$

$$= \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right]\left\{\frac{r^2}{4Dt^2} - \frac{3}{2t}\right\} .$$

On placing all these derivatives in the diffusion equation, Eq. (12.4), we note that the equation is indeed satisfied. Notice that the concentration $C(x, y, z, t)$ of diffusing particles, Eq. (12.5a), is a spherically symmetric function, namely, it can be rewritten as

$$C(r, t) = \frac{N}{\left(\sqrt{4\pi Dt}\right)^3} \exp\left[-\frac{r^2}{4Dt}\right] . \qquad Eq. (12.5b)$$

Here r is the distance from the point of release, or the source point, of diffusing particles. So, starting at the source, the diffusing particles spread out in a radial manner if there are no extenuating conditions. As time t increases, the spreading out of the particles increases the spatial extent of diffusion. However, the total number of particles in the whole space stays the same, namely, N. Explicitly,

$$I = \int\limits_{whole\ space} C(x,y,z,t)\, dxdydz = \int\limits_{whole\ space} C(r,t)\, 4\pi r^2 dr = \frac{4\,\pi\, N}{\left(\sqrt{4\pi Dt}\right)^3} \int\limits_0^\infty \exp\left[-\frac{r^2}{4Dt}\right] r^2 dr \ .$$

Make a change of variables from r to $u = r/\sqrt{4Dt}$, so that

$$I = \frac{4\,\pi\, N}{\left(\sqrt{4\pi Dt}\right)^3} (4Dt)^{\frac{3}{2}} \int\limits_{-\infty}^{\infty} \exp\left(-u^2\right) u^2\, du = \frac{4\,\pi\, N}{\pi^{\frac{3}{2}}} \frac{\sqrt{\pi}}{4} = N \ ,$$

independent of time, t. In the last step we used the integral I_g^2 [see, Eq. (2.16)]. The quantity $\sqrt{4Dt}$ is called the diffusion length, L_D.

12.3 DIFFUSION EQUATION IN ONE DIMENSION

For diffusion of particles in one dimension only (say, along the x-axis), the function describing the concentration of particles (that is, particles per unit length) at location x, at time t, is

$$C(x,t) = \frac{N}{\sqrt{\pi}\, L_D} \exp\left[-\frac{x^2}{L_D^2}\right] = C(0,t) \exp\left[-\frac{x^2}{L_D^2}\right] \ . \qquad Eq.(12.6)$$

This is a bell-shaped Gaussian function that we encountered previously in Figure 4.5 and its related discussion. The Gaussian function of Eq. (12.6) has its maximum, central, value of $C_{max}(t) = C(0,t) = \frac{N}{\sqrt{\pi}\, L_D}$. Just as in the three-dimensional case, the total number of diffusing particles N does not change with time for the one-dimensional case of Gaussian concentration. Explicitly,

$$\int\limits_{-\infty}^{\infty} C(x,t)\, dx = \frac{N}{\sqrt{\pi}\, L_D} \int\limits_{-\infty}^{\infty} \exp\left[-\frac{x^2}{L_D^2}\right] dx \ ,$$

or, on changing the variables from x to $u = x/L_D$,

$$\int\limits_{-\infty}^{\infty} C(x,t)\, dx = \frac{N}{\sqrt{\pi}} \int\limits_{-\infty}^{\infty} \exp[-u^2]\ d = \frac{N}{\sqrt{\pi}} \sqrt{\pi} = N \ ,$$

independent of time, t. The maximum value of the concentration function C_{max} is a function of time and it decreases with time. So, to keep N fixed the concentration spreads out in space. The extent of values of x, namely Δx, over which the function $C(x,t)$ is appreciable is given, at any time, by the full width of the function at half of its maximum value (FWHM). Thus,

$$\Delta x = FWHM = 2\sqrt{\ln 2}\ L_D \ . \qquad Eq.(12.7)$$

Note L_D increases with time; that is, the extent to which certain particles diffuse in time t is proportional to $\sqrt{t}$. Also, the time dependence of the concentration

$$C(x,t) = \frac{N}{\sqrt{\pi}\, L_D} \exp\left[-\frac{x^2}{L_D^2}\right],$$

comes through the time dependence of the diffusion length $L_D = \sqrt{4Dt}$. For $t \to 0$, the concentration of the diffusing particles is represented by a Dirac Delta function; that is, $C(x,0) = N\,\delta(x)$, using $Eq.(I.5b)$.

12.4 FINAL COMMENTS

We note in passing that the diffusion equation described here is same as the heat equation in thermal physics. That is not surprising since the heat equation describes diffusion of heat in a solid. In the thermal case, the concentration $C(\boldsymbol{r},t)$ is replaced by temperature, $T(\boldsymbol{r},t)$, at any point; thermal energy diffuses from a region of higher temperature to a region of lower temperature.

It is worth pointing out that the process of diffusion is also very important in industrial settings, such as in metallurgy. For example, to convert ordinary iron into steel we must add carbon to it, a process called carburizing, which hardens the surface of iron. Parts of iron in a high temperature furnace are placed in contact with carbon-rich gases and carbon diffuses into the parts, following the diffusion equation, turning iron into hardened steel. More generally, the process of diffusion coating, in which some elements are diffused onto the surface of metals to improve their hardness and corrosion properties, has been used in nonstick Teflon coatings of kitchen pots and pans.

PROBLEMS FOR CHAPTER 12

1. Consider a diffusing gas whose source is located at $r = 0$ and its diffusion constant D is $4 \times 10^{-4}\ cm^2/s$. At $t = 0$, the source spews out 1000 particles localized in the air. Plot the diffusion length L_D as a function of time t. What is the rate of growth of the horizontal area $(A = \pi L_D^2)$ filled by the diffusing particles?

2. Again, consider the diffusing gas of 1000 particles and the diffusion constant of $D = 4 \times 10^{-4}\ cm^2/s$. For diffusion in one dimension and the source located at $x = 0$, plot the concentration $C(x, t)$ as a function of x for $t = 1\ s, 4\ s$ and $9\ s$. What is the area under the $C(x, t)$ versus t curve for each value of time t?

3. **Biomedical Physics Application**. A truck carrying hydrogen fluoride, a toxic gas, gets into an accident on the road, causing leakage of about 5,000 molecules of the gas. If the diffusion constant of hydrogen fluoride is $1.7\ X\ 10^{-5}\ cm^2/s$, determine the concentrations of the toxic molecules at the site of the accident 6, 12, 18, and 24 hours after the accident.

4. **Biomedical Physics Application**. A researcher in a virology laboratory is working with a deadly virus in a sealed test tube. The laboratory is in the shape of a cube of side length $12\ m$. The researcher is at the center of the floor of the laboratory when the test tube is accidently dropped, spilling the deadly virus. As a result, the laboratory will need to be sealed and later clinically cleaned. If the diffusion constant of the virus is $9\ X\ 10^{-3}\ m^2/s$, how long will it take the virus to spread out to cover the whole floor of the laboratory?

Chapter 13: Probability Distribution Functions

In various branches of science, we often deal with $physical\ variables$, which are measurable and controllable quantities such as momentum and kinetic energy of a particle or rate of a chemical reaction, etc. The values of physical variables are determined by some physical or chemical property such as mass of the moving particle, temperature and concentration of chemicals, and others. On the other hand, a variable whose value is determined by some random phenomenon is called a $random\ variable$. Examples of random variables are the number of babies delivered in a hospital per day or the periods of rotation of various planets around the Sun, and so on. These quantities are measurable but not controllable. Typically, the outcome of an experiment is an event or number which could belong to either a set of discrete possibilities or to a continuous range of possible values. Flipping a coin or rolling a die are examples of experiments in which the outcome is discrete. On the other hand, observing the change in outdoor temperature in a day from morning to evening or measuring the speed of an accelerating car starting from rest leads to continuous outcomes. A probability distribution function is a mathematical way of associating the different outcomes in an experiment with the probability with which they occur. In this chapter we will learn about $discrete$ as well as $continuous$ probability distribution functions.

13.1 A DISCRETE PROBABILITY DISTRIBUTION FUNCTION

A pharmaceutical company assembles a group of twenty volunteers to test its new vaccine. Before starting the testing, the pharmaceutical company needs to get the vital signs of all volunteers. These include the height, weight, body temperature, and pulse rate of each volunteer. The heights of individual persons are, in inches, 60, 61, 62, 62, 64, 64, 64, 65, 65, 65, 65, 65, 65, 66, 66, 66, 68, 68, 69, and 70. To get the average or mean height, we simply add all the heights and divide by the number of persons, as

$$\bar{h} = \frac{60 + 61 + 62 + \cdots + 70}{20} = \frac{1300}{20} = 65 \text{ inches.}$$

[Note in passing that a variable with a bar on it is used to indicate the average or mean value of that variable.] Alternatively, we could organize the persons who have the same height into groups and write

$$\bar{h} = \frac{60 + 61 + 2(62) + 3(64) + 6(65) + 3(66) + 2(68) + 69 + 70}{20} = 65 \text{ inches.}$$

If there is a very large number of persons, say N, whose average height we need to find, we will first note that N_1 persons have height h_1, N_2 have height h_2, N_3 have height h_3… and N_n have height h_n, such that $N_1 + N_2 + \cdots + N_n = N$. Then,

$$\bar{h} = \frac{N_1 h_1 + N_2 h_2 + N_3 h_3 + \cdots + N_n h_n}{N} = f_1 h_1 + f_2 h_2 + f_3 h_3 + \cdots + f_n h_n \ , \quad Eq.(13.1a)$$

where,

$$f_i = \frac{N_i}{N} \text{ is the fraction of persons with height } h_i \ . \qquad Eq.(13.1b)$$

Thus,

$$\bar{h} = \sum_{i=1}^{n} f_i h_i \ , \qquad (Eq.13.1c)$$

where n is the number of different height groups. If instead of height, we are measuring a different vital sign, say, temperature of the volunteer, T, then,

$$\bar{T} = \sum_{i=1}^{n} f_i T_i \ ,$$

or pulse rate, P,

$$\bar{P} = \sum_{i=1}^{n} f_i P_i \ ,$$

or weight, w,

$$\bar{w} = \sum_{i=1}^{n} f_i w_i \ .$$

The number of groups n is different for different variables. In all these cases, the variable [height, temperature, weight, etc.] is a random variable and has discrete values (since the number of possible values is, at most, N). Also, note $\sum_{i=1}^{n} f_i$=1 in all cases, since all fractions of the sample (which, in this example, is the number of volunteers) must add up to 1. This is referred to as the $normalization\ condition$. Note that these fractions are measures of probabilities. For example, assume that in this cohort consisting of 10 males and 10 females, 14 volunteers are given the actual vaccine and the remaining six volunteers are given a placebo (a harmless medicine with no physiological effects). Males are given 4 placebos and 6 vaccines while the females are given 2 placebos and 8 vaccines. Then we can sort all the volunteers into four groups as shown in the following table.

	With placebo	With vaccine	Total
Male	4	6	10
Female	2	8	10
Total	6	14	20

The outcome of any measurement on this group is an event or a number which is discrete. Now, if a person is randomly chosen from this group, the probability of choosing a person who was given the actual vaccine will be 14/20 and probability of choosing a person given a placebo will be 6/20. The probability of choosing a male volunteer will be 10/20 and probability of choosing a female volunteer will also be 10/20. In general, if we are dealing with two or more events, then we need to find the ***joint probability*** that these events will occur simultaneously. If $P(X)$ is the probability of occurrence of event X and $P(Y)$ is the probability of occurrence of event Y, then joint probability of occurrence of both events X and Y simultaneously, if they are independent events, is

$$P(X \text{ and } Y) = P(X) \cdot P(Y) \ . \qquad Eq.(13.2a)$$

On the other hand, if the two events are not independent and their outcomes are related by some conditions, then we talk about ***conditional probability***. By definition, $P(X|Y)$ is the probability of occurrence of event X, given that event Y has already occurred. Similarly, $P(Y|X)$ is the probability of occurrence of event Y, given that event X has already occurred. The rules for finding the conditional probabilities are

$$P(X|Y) = \frac{P(X \text{ and } Y)}{P(Y)} \ \text{ and } \ P(Y|X) = \frac{P(X \text{ and } Y)}{P(X)} \ . \qquad Eq.(13.2b)$$

If X and Y are independent events, it means that occurrence of event X has no effect on the occurrence of event Y and vice versa and, therefore, $P(X|Y) = P(X)$ and $P(Y|X) = P(Y)$. Then, Eqs. (13.2b) reduce to Eq. (13.2a) for independent events. In case of conditional probabilities, it follows from Eq. (13.2b) that

$$P(X)\,P(Y|X) \ = \ P(Y)\,P(X|Y) \ , \qquad Eq.(13.2c)$$

which is called **Bayes' theorem** in statistics. This theorem provides the probability of an event, based on prior circumstances that might affect the event. Going back to the example of cohort with 20 volunteers, assume that event X is choosing a person with the vaccine and event Y is choosing a female person. Then $P(X)$ is $\frac{14}{20}$ and $P(Y)$ is $\frac{10}{20}$. Also from the table, $P(X \ and \ Y)$ is $\frac{8}{20}$. Since $P(X)$ multiplied by $P(Y)$ is not equal to $P(X \ and \ Y)$, it follows that events X and Y are not independent and, therefore, it is necessary to calculate the conditional probability. The probability of choosing a person with a vaccine, given that a female person has already been chosen, is

$P(X|Y) = \frac{8/20}{10/20} = 0.80$. Similarly, the probability of choosing a female person, given that a vaccinated person has already been chosen, is $P(Y|X) = \frac{8/20}{14/20} = \frac{4}{7} = 0.57$. It is easy to verify that Bayes' theorem is satisfied in this case.

If we have the average value $\bar{x}$ of a random variable x in a sample size of N, it does not provide any information about the spread of the values of x. To measure the spread of values, we define the $variance$ and $standard\ deviation$ of x. To quantify the spread of values of the variable, we measure the deviation of each data point from the mean value, namely, $x_i - \bar{x}$. The value of this deviation, for any individual data point, can be either positive or negative, and the simple sum of all of these deviations will be zero. That is so because, by definition of the average,

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \quad \text{or} \quad \sum_{i=1}^{N} x_i = N\,\bar{x} \quad \text{or} \quad \sum_{i=1}^{N} (x_i - \bar{x}) = 0 \ .$$

So, to define a variance and a standard deviation, σ, we first find the average value of the $squares$ of deviations over all data points. This is called variance and its square root is the standard deviation. Explicitly,

$$\text{variance}, \sigma^2 = \sum_{i=1}^{n} f_i (x_i - \bar{x})^2 = \sum_{i=1}^{n} f_i x_i^2 - 2\,\bar{x} \sum_{i=1}^{n} f_i x_i + [\bar{x}]^2 \sum_{i=1}^{n} f_i$$
$$= \overline{x^2} - 2\,\bar{x}\,\bar{x} + [\bar{x}]^2 = \overline{x^2} - [\bar{x}]^2 \ , \qquad Eq.(13.3a)$$

and

$$\text{standard deviation}, \sigma = \sqrt{\overline{x^2} - [\bar{x}]^2} \ . \qquad Eq.(13.3b)$$

Roughly speaking, variance is an overall measure of how spread out the x_i values are from their mean value $\bar{x}$, while standard deviation is used when considering how an individual data point differs from the mean. In particular, the standard deviation is used when determining the statistical significance of an experimental result.

13.2 BINOMIAL DISTRIBUTION FUNCTION

Consider an experiment that has only two possible outcomes, which are mutually exclusive and can be thought of as a success or a failure. Examples of such experiments are tossing a coin or rolling a die. In case of coin-tossing, if getting a head is a success, then the probability of success is ½ and probability of failure is also ½. While in rolling a die, if getting a 4 is a success, then probability of success is 1/6 and probability of failure is 5/6. The binomial distribution provides the probability of observing n successes in N independent trials if the probability of success in a single trial is p. The success probability p is the same for all trials. The binomial distribution is

$$P(n,N,p) = \frac{N!}{n!\,(N-n)!} p^n (1-p)^{N-n} \ . \qquad Eq.(13.4)$$

Example: In the maternity ward of a hospital, young moms can deliver either a boy or a girl with equal probability. On a particular day, six pregnant women are admitted in the hospital. What is the probability that only one boy will be delivered on this day? Or, two boys will be delivered? Or, three boys will be delivered? Or, four boys will be delivered? Or, five boys will be delivered? Finally, what is the probability that the six babies delivered on this day will be either all boys or all girls?

Solution: On this particular day six experiments are being done in the maternity ward and each experiment has two possible outcomes, either a boy or a girl. In this case $N = 6$. Assuming that delivering a boy is a success for the experiment (no sexism intended!), then $p = ½$. Using the binomial distribution, the probability of delivering only one boy (that is, $n = 1$) is

$$P(1,6,0.5) = \frac{6!}{1!\,(6-1)!}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{6-1} = \frac{6}{64} = \frac{3}{32} \ .$$

The probability of delivering two boys (that is, $n = 2$) is

$$P(2,6,0.5) = \frac{6!}{2!\,(6-2)!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{6-2} = \frac{15}{64} \ .$$

Similarly, probabilities of delivering three or four or five boys are

$$P(3,6,0.5) = \frac{20}{64} = \frac{5}{16}, \quad P(4,6,0.5) = \frac{15}{64} \quad \text{and} \quad P(5,6,0.5) = \frac{6}{64} = \frac{3}{32} \ .$$

Finally, the probabilities of delivering all six boys $(n = 6)$ or all six girls $(n = 0)$ are

$$P(6,6,0.5) = \frac{1}{64} \quad \text{and} \quad P(0,6,0.5) = \frac{1}{64} \ .$$

Note that the sum of probabilities of all possibilities adds up to 1, that is

$$\sum_{n=0}^{6} P(n,6,0.5) = 1 \ ,$$

which is the normalization condition.

13.3 POISSON DISTRIBUTION FUNCTION

Let us, once again, consider an experiment which has two possible outcomes, but now we do not know what the probability of success, p, is in a single trial. If we do this experiment for a large number of times ($N \to \infty$), then we can find an average rate of success. For example, if we toss a coin millions of times, we will infer that average rate of getting a head is ½. So, instead of dealing with p, we deal with average rate of success, namely, $Np \equiv \lambda$. So, we replace p by λ/N and take the limit $N \to \infty$ in the binomial distribution to get

$$P(n,\lambda) = \lim_{N\to\infty} \frac{N!}{n!\,(N-n)!}\left(\frac{\lambda}{N}\right)^n \left(1-\frac{\lambda}{N}\right)^{N-n} .$$

To evaluate this limiting expression, we look at it part-by-part. For example,

$$\lim_{N\to\infty} \frac{N!}{(N-n)!}\frac{1}{N^n} = \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{N^n} = \lim_{N\to\infty} \left(\frac{N}{N}\right)\left(\frac{N-1}{N}\right)\cdots\left(\frac{N-n+1}{N}\right) = 1,$$

and

$$\lim_{N\to\infty} \left(1-\frac{\lambda}{N}\right)^{-n} = 1 .$$

Also, in precalculus, Euler's number, e, is introduced as the limit of $(1+1/N)^N$ as N approaches infinity. The value of this constant is $e = 2.71828$. So,

$$\lim_{N\to\infty} \left(1-\frac{\lambda}{N}\right)^{N} = e^{-\lambda} .$$

Thus,

$$P(n,\lambda) = \frac{\lambda^n}{n!}\exp(-\lambda) \ , \qquad Eq.(13.5)$$

which is known as the Poisson distribution. The Poisson distribution is appropriate if the average rate at which certain identical events occur during a time period is known. Furthermore, all events occur independently; that is, occurrence of one event does not affect the probability of occurrence of the next event. The Poisson distribution provides the probability that n events occur during a certain time period where $n = 0, 1, 2, \cdots$. Note in passing,

$$\sum_{n=0}^{\infty} P(n,\lambda) = \exp(+\lambda)\exp(-\lambda) = 1 \ ,$$

which is the normalization condition.

Example: In several parts of the world, sightings of tornadoes are a common occurrence. According to the historical records of the National Weather Service of the USA, in the state of Michigan the average number of tornadoes is 15 per year, with the largest monthly average of three tornadoes occurring in June. With this information, what is the probability that there will be exactly one or two or three or four tornadoes in Michigan next year in June?

Solution: Since the average rate of tornadoes in June is three, $\lambda = 3$. The probability of n tornadoes in June is

$$P(n,3) = \frac{3^n}{n!}\exp(-3) \ .$$

Thus, probabilities of having one or two or three or four tornadoes next June are

$$P(1,3) = \frac{3^1}{1!}\exp(-3) = 0.149 \ ,$$

$$P(2,3) = \frac{3^2}{2!}\exp(-3) = 0.224 \ ,$$

$$P(3,3) = \frac{3^3}{3!}\exp(-3) = 0.224 \ ,$$

$$P(4,3) = \frac{3^4}{4!}\exp(-3) = 0.168 \ .$$

13.4 MAXWELL-BOLTZMANN (CONTINUOUS) DISTRIBUTION FUNCTION

Now suppose we wish to find the average speed of molecules in a room. It will be extremely difficult to do since the number of molecules is immeasurably large and difficult to count. Even if we only wish to find the fraction of molecules with a particular speed v_i, their number will be enormously large. In this case, we break the whole possible range of speeds into small groups or intervals Δv and concentrate on fraction of molecules with speeds between v_i and $v_i + \Delta v$. Then, the fraction of molecules in this interval is $f(v_i)\Delta v$. The function $f(v_i)$ is called a distribution function or probability density function (PDF). In this case,

$$\bar{v} = \sum_{i=1}^{n} v_i f(v_i)\Delta v \ ,$$

and, if $\Delta v \to 0$, the variable v becomes a continuous random variable,

$$\bar{v} = \int_{\text{whole range of } v} v f(v)\, dv \ . \qquad Eq.(13.6a)$$

Also, since all fractions must add up to 1,

$$\int_{\text{whole range of } v} f(v)\, dv = 1 \ . \qquad Eq.(13.6b)$$

This is the normalization condition for the distribution function. This is equivalent to the notion that we can be 100% certain that an observation will fall into the range of all possible outcomes.

The velocity distribution of molecules of mass m in a gas at temperature T is the Maxwell-Boltzmann distribution. It is a continuous distribution function given by

$$f(v) = 4\pi v^2 \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right) \ , \qquad Eq.(13.7)$$

where k is the Boltzmann constant. We note

$$\int_0^\infty f(v)dv = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \int_0^\infty \exp\left(-\frac{mv^2}{2T}\right) v^2\, dv \ .$$

To evaluate this integral, we make a change of variables. Let $u^2 = \frac{m}{2kT} v^2$ or $u = \sqrt{\frac{m}{2kT}}\, v$ and $du = \sqrt{\frac{m}{2kT}} dv$.

Then,

$$\int_0^\infty f(v)dv = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \left(\frac{2kT}{m}\right)^{3/2} \int_0^\infty \exp(-u^2)\, u^2 du = 4\pi \frac{1}{\pi^{3/2}} \frac{\sqrt{\pi}}{4} = 1 \ , \qquad Eq.(13.8a)$$

as expected, since this is the normalization condition. Here we used the integral I_g^2 from Eq. (2.13). We can also find average values of v and of v^2 as

$$\bar{v} = \int_0^\infty v f(v)dv = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \int_0^\infty \exp\left(-\frac{mv^2}{2kT}\right) v^3 dv \ .$$

Again, we make a change of variables. Let $t = \frac{mv^2}{2kT}$ and $dt = \frac{m}{2kT} 2v dv$, so that

$$\bar{v} = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \frac{1}{2} \left(\frac{2kT}{m}\right)^2 \int_0^\infty \exp(-t)\, t\, dt = 2\left(\frac{2kT}{m\pi}\right)^{1/2} \int_0^\infty \exp(-t)\, t\, dt = \left(\frac{8kT}{\pi m}\right)^{1/2} \ , \qquad Eq.(13.8b)$$

and

$$\overline{v^2} = \int_0^\infty v^2 f(v) dv = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \int_0^\infty \exp\left(-\frac{mv^2}{2kT}\right) v^4 dv \ .$$

With the change of variables, $u^2 = \frac{mv^2}{2kT}$,

$$\overline{v^2} = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} \left(\frac{2kT}{m}\right)^{5/2} \int_0^\infty \exp(-u^2)\, u^4 du = \frac{4\pi}{\pi^{3/2}} \frac{2kT}{m} \frac{3}{8}\sqrt{\pi} = \frac{3kT}{m} \ . \qquad Eq.(13.8c)$$

Note in passing that $\frac{1}{2} m\overline{v^2} = \frac{3kT}{2}$, that is, average kinetic energy per molecule in a gas depends only on the temperature of the gas.

13.5 NORMAL OR GAUSSIAN (CONTINUOUS) DISTRIBUTION FUNCTION

Another useful continuous probability distribution function is the Gaussian, or normal, distribution function. A random variable x has normal distribution if $-\infty \leq x \leq \infty$ and the corresponding probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \ . \qquad Eq.(13.9)$$

By definition, $f(x)\, dx$ is the probability that value of variable x lies between x and $x + dx$. First, we verify the normalization condition, namely, $\int_{-\infty}^{\infty} f(x)\, dx = 1$. The integral is

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \ .$$

To evaluate this integral, we make a change of variable from x to u using $u = \frac{x-\mu}{\sigma}$, or $x = \sigma\, u + \mu$, and $dx = \sigma\, du$,

$$\int_{-\infty}^{\infty} f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{u^2}{2}\right] du \ = \frac{1}{\sqrt{2\pi}} \ \sqrt{2\pi} = 1 \ . \qquad Eq.(13.10a)$$

Here we used the integral I_{gg}^0 from Eq. (2.14). Next, we find average values of x and of x^2 using the same change of variable from x to u,

$$\bar{x} \ = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (u\sigma + \mu) \exp\left[-\frac{u^2}{2}\right] du \; = 0 + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{u^2}{2}\right] du \; = \mu \;\; , \qquad Eq.(13.10b)$$

and

$$\overline{x^2} \; = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (u^2\sigma^2 + 2\mu\sigma u \; + \mu^2) \exp\left[-\frac{u^2}{2}\right] du \; = \frac{1}{\sqrt{2\pi}} \{\sigma^2\sqrt{2\pi} + 0 + \mu^2\sqrt{2\pi}\}$$

or,

$$\overline{x^2} \; = \sigma^2 + \mu^2 \;\; . \qquad Eq.(13.10c)$$

Then,

$$\text{variance} = \; \overline{x^2} - [\bar{x}]^2 = (\sigma^2 + \mu^2) - \mu^2 = \sigma^2$$

and,

$$\text{standard deviation} = \sigma \;\; . \qquad Eq.(13.11)$$

So, now we can interpret the normalized Gaussian distribution function. If a variable x is normally distributed, then $f(x)\, dx$ is the probability that x lies between x and $x + dx$. The function $f(x)$ has a mean value of μ and standard deviation of σ. If we change variable from x to $u = \frac{x-\mu}{\sigma}$, then, for the normal distribution,

$$f(x)\, dx = F(u)\, du \;\; ,$$

where

$$F(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \;\; .$$

We note that $F(u)$ is the same Gaussian function that we encountered in chapters 4 and 12. It is worth observing that $F(u)$ is a symmetric function of u, that is, $F(-u) = F(u)$. Also, $F(u)$ is a normalized function, that is, $\int_{-\infty}^{\infty} F(u)\, du = 1$. The quantity $F(u)\, du$ is the probability that u lies between u and $u + du$. The probability that u lies between a and b is

$$\int_a^b F(u)\,du = \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{u^2}{2}\right)du \ .$$

For given a and b, the numerical value of this integral is obtained from the tabulated values of the Gaussian function, $\Phi(x)$, defined as

$$\Phi(x) = \int_{-\infty}^{x} F(u)\,du = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp\left(-\frac{u^2}{2}\right)du \ . \qquad Eq.(13.12)$$

The tabulated numerical values of $\Phi(x)$, as a function of x, are given in Appendix D, in which the shaded area under the Gaussian curve is the value of $\Phi(x)$. Note $\Phi(\infty) = 1, \Phi(0) = \frac{1}{2}, \Phi(-\infty) = 0$. Also, due to the symmetric nature of $F(u)$,

$$\Phi(-x) = \int_{-\infty}^{-x} F(u)\,du = -\int_{\infty}^{x} F(-v)\,dv = \int_{x}^{\infty} F(v)\,dv = \int_{-\infty}^{\infty} F(v)\,dv - \int_{-\infty}^{x} F(v)\,dv$$

$$\Phi(-x) = 1 - \Phi(x) \ . \qquad Eq.(13.13)$$

This relationship allows us to find Φ for negative values of x in terms of Φ for positive values of x. So, in the standard table of Appendix D, the function $\Phi(x)$ is tabulated for positive values of x only. Also, if $b > a$, then

$$\int_a^b F(u)\,du = \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{u^2}{2}\right)du$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{b} \exp\left(-\frac{u^2}{2}\right)du - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} \exp\left(-\frac{u^2}{2}\right)du = \Phi(b) - \Phi(a) \ . \qquad Eq.(13.14)$$

If the range of value of x is from x_L to x_H $(x_L \le x \le x_H)$ instead of from $-\infty$ to ∞, then

$$f(x) = N\frac{1}{\sigma\sqrt{2\pi}}\begin{cases}\exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right) & \text{for} \quad x_L \le x \le x_H \\ 0 & \text{otherwise}\end{cases} \ . \qquad Eq.(13.15a)$$

This is referred to as the $truncated\ normal\ distribution$. The scaling constant N is determined by the fact that $\int_{-\infty}^{\infty} f(x)dx = 1$, or,

$$N\frac{1}{\sigma\sqrt{2\pi}}\int_{x_L}^{x_H} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)dx = 1 \ .$$

Making a change of variable from x to $u = \frac{x-\mu}{\sigma}$,

$$1 = N\frac{1}{\sqrt{2\pi}}\int_{u_L}^{u_H} \exp\left(-\frac{u^2}{2}\right)\mathrm{du} = N[\Phi(u_H) - \Phi(u_L)] \ ,$$

or

$$N = \frac{1}{[\Phi(u_H) - \Phi(u_L)]} \ . \qquad Eq.(13.15b)$$

Example: The grade point averages (GPA) of a large population of students at a university are normally distributed with a mean of 2.4 and a standard deviation of 0.8. The students with GPAs higher than 3.3 receive honors credit while students with GPAs less than 2.0 do not receive the degree.

What percentage of students receive honors credit?

What percentage of students can never graduate?

Solution: Let x be GPA of students, $0 \le x \le 4$.

$$x_H = 4.0, u_H = \frac{x_H - \mu}{\sigma} = \frac{4 - 2.4}{0.8} = 2.0 \ ,$$

$$x_L = 0.0, u_L = \frac{x_L - \mu}{\sigma} = \frac{0 - 2.4}{0.8} = -3.0 \ ,$$

$$N = \frac{1}{\Phi(2.0) - \Phi(-3.0)} = \frac{1}{\Phi(2) - [1 - \Phi(3.0)]} = \frac{1}{\Phi(2) + \Phi(3.0) - 1}$$

$$= \frac{1}{0.9772 + 0.9987 - 1} = \frac{1}{1.9759 - 1} = \frac{1}{0.9759} \ .$$

(a)

$$\text{Percentage of honors students} = \int_{3.3}^{4.0} f(x)dx = N\frac{1}{\sigma\sqrt{2\pi}}\int_{3.3}^{4.0} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx \ .$$

Let $u = \frac{x-\mu}{\sigma}$, then lower limit of $u = \frac{3.3-2.4}{0.8} = \frac{9}{8}$, upper limit of $u = \frac{4.0-2.4}{0.8} = 2$.

$$\text{Percentage} = N\frac{1}{\sqrt{2\pi}}\int_{9/8}^{2} \exp\left[-\frac{u^2}{2}\right]du$$

$$= \frac{\Phi(2) - \Phi\left(\frac{9}{8}\right)}{\Phi(2) - \Phi(-3)} = \frac{0.9772 - 0.8686}{0.9759} = \frac{0.1086}{0.9759} = 0.11 \quad \text{or} \quad 11\% \ .$$

(b)

$$\text{Percentage of failing students} = \int_0^{2.0} f(x)dx = N\frac{1}{\sigma\sqrt{2\pi}}\int_0^{2.0} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx \ .$$

Let $u = \frac{x-\mu}{\sigma}$, lower limit $= \frac{0-2.4}{0.8} = -3$, upper limit $= \frac{2.0-2.4}{0.8} = -0.5$.

$$\text{Percentage } = N\frac{1}{\sqrt{2\pi}}\int_{-3}^{-0.5} \exp\left[-\frac{u^2}{2}\right]du$$

$$= N[\Phi(-0.5) - \Phi(-3)] = N[1 - \Phi(0.5) - 1 + \Phi(3)] = N[\Phi(3) - \Phi(0.5)]$$

$$= \frac{0.9987 - 0.6915}{0.9759} = \frac{0.3072}{0.9759} = 0.31 \quad \text{or} \quad 31\% \ .$$

Example: The U.S. Centers for Disease Control and Prevention has reported several cases of Salmonella outbreak linked to guinea pigs. A veterinary research physician sets up a trap to catch a large number of guinea pigs. If the mean width of guinea pig's shoulder is $\mu = 3.8$ inches with a standard deviation of 0.6 inch, and if the door of the trap is 5 inches wide, what percentage of guinea pigs will pass through the door?

Solution: Let x be the shoulder width of a guinea pig, so that $0 \leq x \leq \infty$. The percentage of guinea pigs with shoulder length less than 5 inches is

$$\text{Percentage} = \frac{\frac{1}{\sigma\sqrt{2\pi}}\int_0^{5.0} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx}{\frac{1}{\sigma\sqrt{2\pi}}\int_0^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx} \ .$$

Make a change of variables from x to $u = \frac{x-\mu}{\sigma}$.

For $x = \infty, u = \infty$. For $x = 0, u = \frac{0-3.8}{0.6} = -\frac{19}{3}$. For $x = 5.0, u = \frac{5-3.8}{0.6} = 2.0$.

$$\text{Percentage} = \frac{\frac{1}{\sqrt{2\pi}}\int_{-19/3}^{2.0} \exp\left[-\frac{u^2}{2}\right]du}{\frac{1}{\sqrt{2\pi}}\int_{-19/3}^{\infty} \exp\left[-\frac{u^2}{2}\right]du} = \frac{\Phi(2) - \Phi(-\frac{19}{3})}{\Phi(\infty) - \Phi(-\frac{19}{3})}$$

$$= \frac{\Phi(2) - [1 - \Phi(6.33)]}{\Phi(\infty) - [1 - \Phi(6.33)]} = \frac{\Phi(2) + \Phi(6.33) - 1}{\Phi(\infty) + \Phi(6.33) - 1} = \frac{0.9772 + 1 - 1}{1 + 1 - 1} = 0.9772 \ ,$$

or 97.7% guinea pigs will fit through the door.

13.6 A FINAL REMARK

As a final remark, it is worth pointing out that concepts related to probability distribution functions are central to quantum physics. The initial formulation of quantum physics using matrix methods led to the understanding of eigenvalues and eigenfunctions of matrices as representing physical observables. The gist of quantum physics is to talk about *possibilities* and *probabilities* in the measurement of physical properties, such as energy, momentum, angular momentum, and so on, at the scale of atomic and subatomic particles. The eigenvalues represent the possibilities and the eigenfunctions (or, their modulus square) represent the corresponding probability distribution functions.

PROBLEMS FOR CHAPTER 13

1. To explain the conduction properties of metals, Paul Drude provided a simple model of a metal as a reservoir of free-electron gas. Even though Drude's model was successful in explaining the classical Ohm's law, it failed to account for the quantum behavior of electrons. The Exclusion Principle in quantum physics asserts that no more than two electrons can occupy the same atomic orbital. Classically, particles of a gas at absolute zero temperature will all have zero kinetic energy. In quantum physics, all electrons in an electron gas cannot be in ground state since that would contradict the Exclusion Principle. At absolute zero temperature, the fraction of electrons in a metal with energies between E and $E + dE$ is given by

$$f(E)\,dE = \begin{cases} C\sqrt{E}\,dE & \text{for } 0 \le E \le E_F \\ 0 & \text{for } E > E_F \end{cases} ,$$

where E_F is a constant (called the Fermi energy).

Determine C if this energy distribution is normalized to 1.

Find the average energy of electrons in terms of E_F.

2. A one-dimensional classical harmonic oscillator of mass m and frequency ω is oscillating along the x–axis with its equilibrium point at $x = 0$. The turning points of the classical oscillator at $x = \pm x_c$ refer to the values of x at which the oscillator momentarily comes to rest and reverses its direction of motion. In other words, x_c is the largest displacement or the amplitude of the classical oscillator. The total energy E of the classical oscillator is conserved and its value at the classical turning points is $E = \frac{1}{2}m\omega^2 x_c^2$. In quantum physics, the one-dimensional harmonic oscillator can be found anywhere along the x–axis from $x = -\infty$ to $x = +\infty$. The probability of finding the quantum oscillator between x and $x + dx$ is

$$P(x)\,dx = C\exp(-x^2/x_c^2)\;dx \;.$$

Determine C if this probability distribution function is normalized to 1.

Find the probability of locating the quantum oscillator outside the classical turning points.

3. The simplest of all atoms is the hydrogen atom consisting of a proton and an electron. The proton is in the nucleus and the electron can be anywhere in the space surrounding the nucleus, with its position given by a distribution function. The probability of finding the electron at a distance between r and $r + dr$ from the nucleus is

$$P(r)\,dr = C\,r^2\exp(-2r/a_o)\,dr \;,$$

where a_o is a constant (called the Bohr radius). Here, r can take any value between 0 and ∞.

Determine C if this probability distribution function is normalized to 1.

Using this probability distribution function, determine the average value $\bar{r}$ of r.

4. **Biomedical Physics Application.** The gestation period of human pregnancies has a normal distribution with a mean μ of 268 days and standard deviation σ of 16 days.

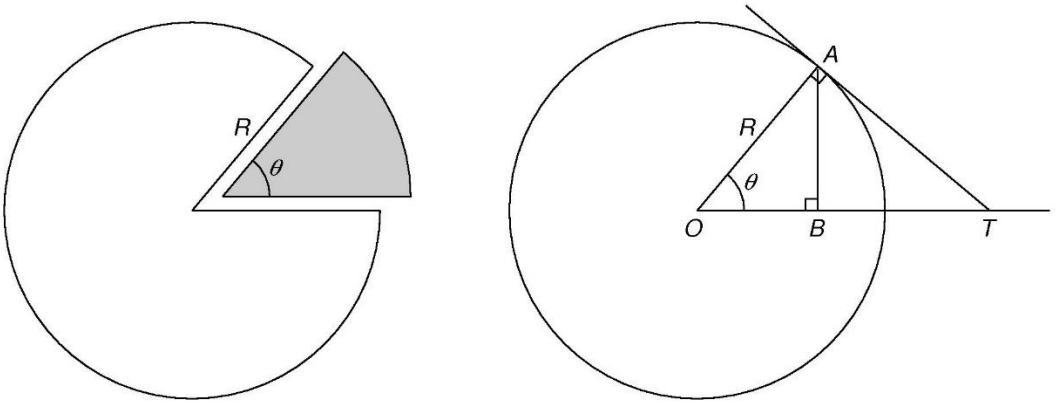
What percentage of pregnancies last between 260 and 284 days?

The colloquial term “preemies” refers to infants who have preterm birth with gestational age of less than 37 weeks, or 259 days. What percentage of infants are preemies?

Post-term babies are born after 284 days of pregnancy. What percentage of pregnancies leads to post-term babies?

Appendix A: Limiting Value of $\sin\theta/\theta$ as $\theta \to 0$.

Using simple facts from geometry and trigonometry, we can determine the limiting value of $\sin\theta/\theta$ as θ approaches 0.



Consider a circular pie, of radius R, which is cut into slices. One slice of angular size ϴ (shown in the left figure) is of area A_s. Using the concept of proportionalities, ratio of area of slice (A_s) and area of full pie (πR^2) is same as the ratio of angular size of slice (ϴ) and angular size of full circular pie, namely, 2π. Thus, $\frac{A_s}{\pi R^2} = \frac{\theta}{2\pi}$ or $A_s = \frac{\theta R^2}{2}$. In the adjoining right figure, we show the same slice of angular size ϴ and label one of its corners as A. From this corner A we drop a perpendicular line that meets the opposite side of the slice at point B. Also, at corner A, we draw a tangent to the circular pie (of center O) that meets the extension of line OB at point T. From this second figure, we note

$$\text{Area of triangle } OAB \leq \text{Area of slice} \leq \text{Area of triangle } OAT .$$

Or,

$$\frac{1}{2}(OB)(AB) \leq A_s \leq \frac{1}{2}(OA)(AT)$$

$$\frac{1}{2}(R\cos\theta)(R\sin\theta) \leq \frac{\theta R^2}{2} \leq \frac{1}{2}(R)(R\tan\theta)$$

$$\cos\theta \leq \frac{\theta}{\sin\theta} \leq \frac{1}{\cos\theta}$$

Since $\cos\theta \to 1$ as $\theta \to 0$, it follows that

$$\lim_{\theta \to 0} \frac{\sin(\theta)}{\theta} = 1 \quad . \qquad Eq.(A.1)$$

In addition, we note that for small values of θ, $\sin\theta \approx \theta$.

Appendix B: Mnemonic for Maxwell's Relations of Thermodynamics

Laws of thermodynamics play fundamental roles in physics, biology, chemistry, chemical engineering, mechanical engineering, and other scientific disciplines. In their mathematical descriptions, these laws relate four physical variables, pressure (P), volume (V), temperature (T) and entropy (S). Derivatives of these variables are related via four Maxwell's relations of thermodynamics, which are

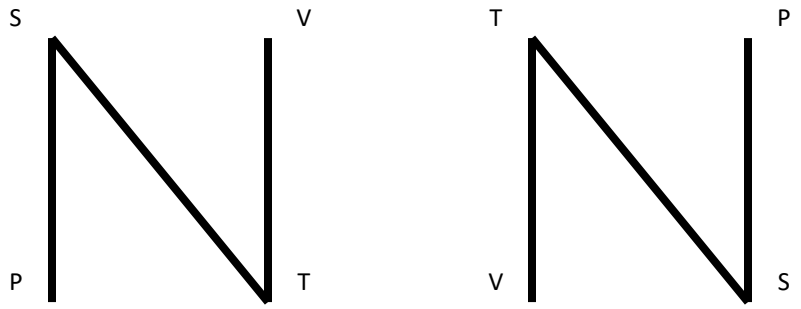
$$+\left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P \ ,$$

$$+\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial P}{\partial S}\right)_V \ ,$$

$$+\left(\frac{\partial P}{\partial T}\right)_V = +\left(\frac{\partial S}{\partial V}\right)_T \ ,$$

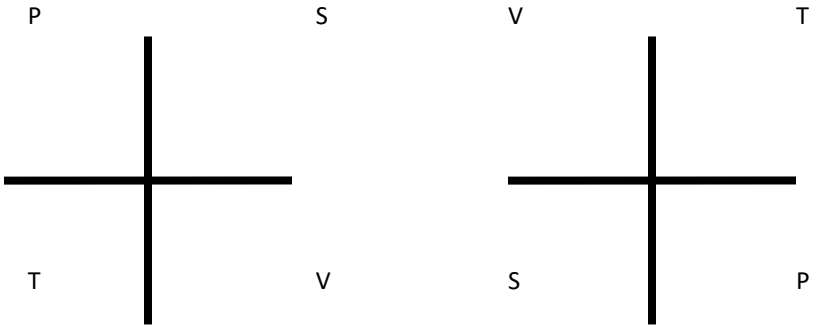
$$+\left(\frac{\partial V}{\partial S}\right)_P = +\left(\frac{\partial T}{\partial P}\right)_S \ .$$

This appendix presents a simple mnemonic device for recalling these relationships without any complications for remembering the correct signs. Note that two of these relations have a positive (+) sign on the right-hand side and the other two have a negative (N) sign on the right-hand side. In each relationship, the variable with respect to which a partial derivative is taken on one side is kept fixed on the other side. The first two relations with the negative sign on the right-hand side can be memorized as follows. Recalling the letter N for the negative relations, wrap the four letters PSTV for the thermodynamic variables, *in alphabetical order*, along this N in two possible ways as shown here:



In these figures, the placement of letters is in the same location/way as in the first two Maxwell's relations, which contain the negative sign on the right-hand side.

To get the mnemonic for the other two Maxwell relations—namely, the ones with the positive sign on the right-hand side—rotate these figures by $\pi/2$ (does not matter whether the rotation is clockwise or counterclockwise) and replace N (negative) by + (positive), to get



In these figures, the placement of letters is in the same location/way as in the last two Maxwell's relations, which contain the positive sign on the right-hand side.

Appendix C: Proof of the Epsilon-Delta Identity

In the introduction of Levi-Civita symbol, ϵ_{ijk}, in the Interlude chapter, a relationship between ϵ_{ijk} and Kronecker delta δ_{ij} was presented as

$$\sum_i \epsilon_{ijk}\epsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl} \ .$$

This epsilon-delta identity can be proved by starting with a determinant of order 3 whose element, a_{ij}, located at the intersection of i^{th} row and j^{th} column, is δ_{ij}. This determinant is

$$\mathbb{A} = \begin{vmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \\ \delta_{31} & \delta_{32} & \delta_{33} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1 \ .$$

In the Interlude chapter, the value of Levi-Civita symbol, ϵ_{ijk}, was given as $+1$ (or -1) if an even (or odd) permutation of (ijk) gives (123). Thus, on replacing the row numbers (123) of $\mathbb{A}$ by (ijk), the value of the determinant will become ϵ_{ijk} instead of 1, that is

$$\begin{vmatrix} \delta_{i1} & \delta_{i2} & \delta_{i3} \\ \delta_{j1} & \delta_{j2} & \delta_{j3} \\ \delta_{k1} & \delta_{k2} & \delta_{k3} \end{vmatrix} = \epsilon_{ijk} \ .$$

Similarly, on replacing the column numbers (123) by (nlm), the value of the determinant will become

$$\begin{vmatrix} \delta_{in} & \delta_{il} & \delta_{im} \\ \delta_{jn} & \delta_{jl} & \delta_{jm} \\ \delta_{kn} & \delta_{kl} & \delta_{km} \end{vmatrix} = \epsilon_{ijk}\epsilon_{nlm} \ .$$

Now, replace n by i, and expand the determinant across first column explicitly, to get

$$\epsilon_{ijk}\epsilon_{ilm} = \delta_{ii}\big(\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}\big) - \delta_{ji}(\delta_{il}\delta_{km} - \delta_{im}\delta_{kl}) + \delta_{ki}\big(\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}\big) \ .$$

Finally, sum over i and use $\sum_{i=1}^{3} \delta_{ii} = 3$ to obtain

$$\sum_i \epsilon_{ijk}\epsilon_{ilm} = 3\big(\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}\big) - \big(\delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}\big) + \big(\delta_{jm}\delta_{kl} - \delta_{jl}\delta_{km}\big) \ ,$$

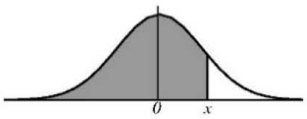
or,

$$\textstyle\sum_i \epsilon_{ijk}\epsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl} \ .$$

Appendix D: Areas under the Standard Normal Curve

The table below provides the values of the Gaussian function, $\Phi(x)$, as a function of *x*.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp(-u^2/2)\,du$$



x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5238	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6330	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7156	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7641	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 13.1. Areas under the standard curve from $-\infty$ to *x*.