Mathematical Methods for Physicists

A concise introduction

TAI L. CHOW
California State University

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http://www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

First published 2000

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times New Roman 10/13pt

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Chow, Tai L.
Mathematical physics : mathematical methods for scientists and engineers / Tai L. Chow.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-65227-8 (hc.) – ISBN 0-521-65544-7 (pbk.)
1. Mathematical physics. I. Title

QC20 .C57 2000
530.15 21–dc21 99–044592

ISBN 0 521 65227 8 hardback
ISBN 0 521 65544 7 paperback

Contents

1

Vector and tensor analysis

Vectors and scalars

Vector methods have become standard tools for the physicists. In this chapter we discuss the properties of the vectors and vector fields that occur in classical physics. We will do so in a way, and in a notation, that leads to the formation of abstract linear vector spaces in Chapter 5.

A physical quantity that is completely specified, in appropriate units, by a single number (called its magnitude) such as volume, mass, and temperature is called a scalar. Scalar quantities are treated as ordinary real numbers. They obey all the regular rules of algebraic addition, subtraction, multiplication, division, and so on.

There are also physical quantities which require a magnitude and a direction for their complete specification. These are called vectors *if* their combination with each other is commutative (that is the order of addition may be changed without affecting the result). Thus not all quantities possessing magnitude and direction are vectors. Angular displacement, for example, may be characterised by magnitude and direction but is not a vector, for the addition of two or more angular displacements is not, in general, commutative (Fig. 1.1).

In print, we shall denote vectors by boldface letters (such as **A**) and use ordinary italic letters (such as A) for their magnitudes; in writing, vectors are usually represented by a letter with an arrow above it such as $\vec{A}$. A given vector **A** (or $\vec{A}$) can be written as

$$\mathbf{A} = A\hat{A}, \tag{1.1}$$

where A is the magnitude of vector **A** and so it has unit and dimension, and $\hat{A}$ is a dimensionless unit vector with a unity magnitude having the direction of **A**. Thus $\hat{A} = \mathbf{A}/A$.

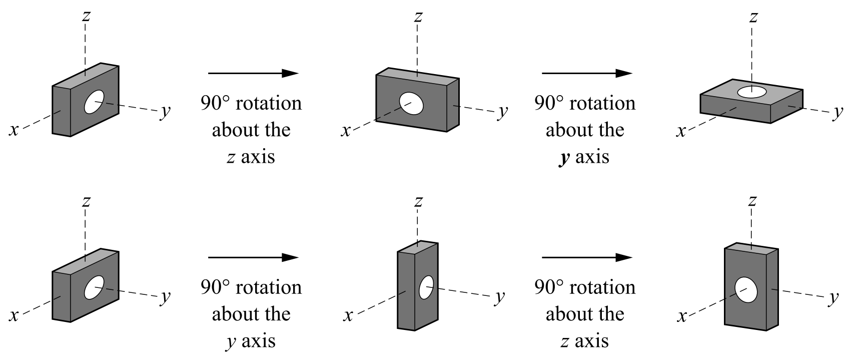


Figure 1.1. Rotation of a parallelpiped about coordinate axes.

A vector quantity may be represented graphically by an arrow-tipped line segment. The length of the arrow represents the magnitude of the vector, and the direction of the arrow is that of the vector, as shown in Fig. 1.2. Alternatively, a vector can be specified by its components (projections along the coordinate axes) and the unit vectors along the coordinate axes (Fig. 1.3):

$$\mathbf{A} = A_1\hat{e}_1 + A_2\hat{e}_2 + A\hat{e}_3 = \sum_{i=1}^{3} A_i\hat{e}_i, \tag{1.2}$$

where $\hat{e}_i$ $(i = 1, 2, 3)$ are unit vectors along the rectangular axes x_i $(x_1 = x, x_2 = y, x_3 = z)$; they are normally written as $\hat{i}, \hat{j}, \hat{k}$ in general physics textbooks. The component triplet (A_1, A_2, A_3) is also often used as an alternate designation for vector $\mathbf{A}$:

$$\mathbf{A} = (A_1, A_2, A_3). \tag{1.2a}$$

This algebraic notation of a vector can be extended (or generalized) to spaces of dimension greater than three, where an ordered n-tuple of real numbers, $(A_1, A_2, \ldots, A_n)$, represents a vector. Even though we cannot construct physical vectors for $n > 3$, we can retain the geometrical language for these n-dimensional generalizations. Such abstract "vectors" will be the subject of Chapter 5.

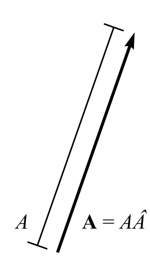


Figure 1.2. Graphical representation of vector **A**.

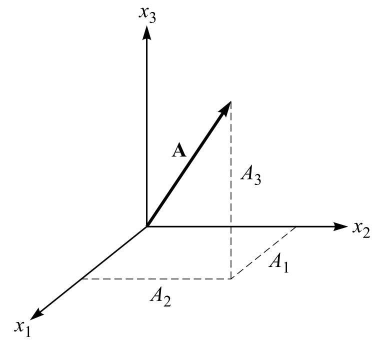


Figure 1.3. A vector **A** in Cartesian coordinates.

Direction angles and direction cosines

We can express the unit vector $\hat{A}$ in terms of the unit coordinate vectors $\hat{e}_i$. From Eq. (1.2), $\mathbf{A} = A_1\hat{e}_1 + A_2\hat{e}_2 + A\hat{e}_3$, we have

$$\mathbf{A} = A\left(\frac{A_1}{A}\hat{e}_1 + \frac{A_2}{A}\hat{e}_2 + \frac{A_3}{A}\hat{e}_3\right) = A\hat{A}.$$

Now $A_1/A = \cos\alpha, A_2/A = \cos\beta$, and $A_3/A = \cos\gamma$ are the direction cosines of the vector **A**, and α, β, and γ are the direction angles (Fig. 1.4). Thus we can write

$$\mathbf{A} = A(\cos\alpha\hat{e}_1 + \cos\beta\hat{e}_2 + \cos\gamma\hat{e}_3) = A\hat{A};$$

it follows that

$$\hat{A} = (\cos\alpha\hat{e}_1 + \cos\beta\hat{e}_2 + \cos\gamma\hat{e}_3) = (\cos\alpha, \cos\beta, \cos\gamma). \qquad (1.3)$$

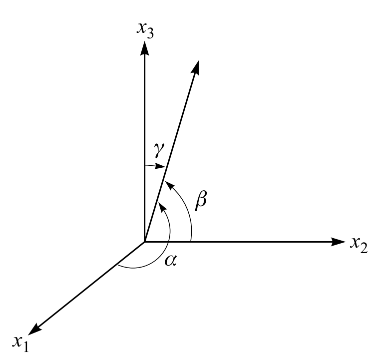


Figure 1.4. Direction angles of vector **A**.

Vector algebra

Equality of vectors

Two vectors, say **A** and **B**, are equal if, and only if, their respective components are equal:

$$\mathbf{A} = \mathbf{B} \quad \text{or} \quad (A_1, A_2, A_3) = (B_1, B_2, B_3)$$

is equivalent to the three equations

$$A_1 = B_1, A_2 = B_2, A_3 = B_3.$$

Geometrically, equal vectors are parallel and have the same length, but do not necessarily have the same position.

Vector addition

The addition of two vectors is defined by the equation

$$\mathbf{A} + \mathbf{B} = (A_1, A_2, A_3) + (B_1, B_2, B_3) = (A_1 + B_1, A_2 + B_2, A_3 + B_3).$$

That is, the sum of two vectors is a vector whose components are sums of the components of the two given vectors.

We can add two non-parallel vectors by graphical method as shown in Fig. 1.5. To add vector **B** to vector **A**, shift **B** parallel to itself until its tail is at the head of **A**. The vector sum **A** + **B** is a vector **C** drawn from the tail of **A** to the head of **B**. The order in which the vectors are added does not affect the result.

Multiplication by a scalar

If c is scalar then

$$c\mathbf{A} = (cA_1, cA_2, cA_3).$$

Geometrically, the vector $c\mathbf{A}$ is parallel to **A** and is c times the length of **A**. When $c = -1$, the vector $-\mathbf{A}$ is one whose direction is the reverse of that of **A**, but both

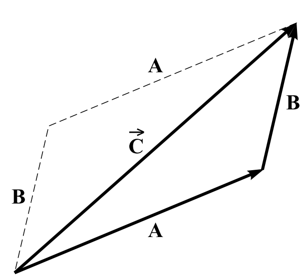


Figure 1.5. Addition of two vectors.

have the same length. Thus, subtraction of vector **B** from vector **A** is equivalent to adding $-\mathbf{B}$ to **A**:

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

We see that vector addition has the following properties:

(a) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ (commutativity);
(b) $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$ (associativity);
(c) $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$;
(d) $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$.

We now turn to vector multiplication. Note that division by a vector is not defined: expressions such as $k/\mathbf{A}$ or $\mathbf{B}/\mathbf{A}$ are meaningless.

There are several ways of multiplying two vectors, each of which has a special meaning; two types are defined.

The scalar product

The scalar (dot or inner) product of two vectors **A** and **B** is a real number defined (in geometrical language) as the product of their magnitude and the cosine of the (smaller) angle between them (Figure 1.6):

$$\mathbf{A} \cdot \mathbf{B} \equiv AB\cos\theta \qquad (0 \le \theta \le \pi). \tag{1.4}$$

It is clear from the definition (1.4) that the scalar product is commutative:

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}, \tag{1.5}$$

and the product of a vector with itself gives the square of the dot product of the vector:

$$\mathbf{A} \cdot \mathbf{A} = A^2. \tag{1.6}$$

If $\mathbf{A} \cdot \mathbf{B} = 0$ and neither **A** nor **B** is a null (zero) vector, then **A** is perpendicular to **B**.

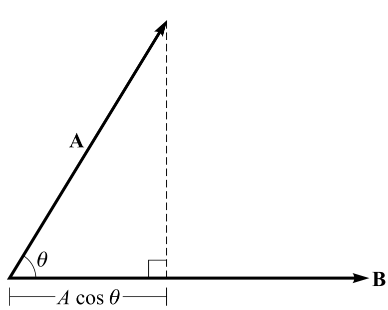


Figure 1.6. The scalar product of two vectors.

We can get a simple geometric interpretation of the dot product from an inspection of Fig. 1.6:

$$(B\cos\theta)A = \text{projection of } \mathbf{B} \text{ onto } \mathbf{A} \text{ multiplied by the magnitude of } \mathbf{A},$$

$$(A\cos\theta)B = \text{projection of } \mathbf{A} \text{ onto } \mathbf{B} \text{ multiplied by the magnitude of } \mathbf{B}.$$

If only the components of **A** and **B** are known, then it would not be practical to calculate $\mathbf{A}\cdot\mathbf{B}$ from definition (1.4). But, in this case, we can calculate $\mathbf{A}\cdot\mathbf{B}$ in terms of the components:

$$\mathbf{A}\cdot\mathbf{B} = (A_1\hat{e}_1 + A_2\hat{e}_2 + A_3\hat{e}_3)\cdot(B_1\hat{e}_1 + B_2\hat{e}_2 + B_3\hat{e}_3); \tag{1.7}$$

the right hand side has nine terms, all involving the product $\hat{e}_i\cdot\hat{e}_j$. Fortunately, the angle between each pair of unit vectors is 90°, and from (1.4) and (1.6) we find that

$$\hat{e}_i\cdot\hat{e}_j = \delta_{ij}, \qquad i,j = 1,2,3, \tag{1.8}$$

where δ_{ij} is the Kronecker delta symbol

$$\delta_{ij} = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j. \end{cases} \tag{1.9}$$

After we use (1.8) to simplify the resulting nine terms on the right-side of (7), we obtain

$$\mathbf{A}\cdot\mathbf{B} = A_1B_1 + A_2B_2 + A_3B_3 = \sum_{i=1}^{3} A_iB_i. \tag{1.10}$$

The law of cosines for plane triangles can be easily proved with the application of the scalar product: refer to Fig. 1.7, where **C** is the resultant vector of **A** and **B**. Taking the dot product of **C** with itself, we obtain

$$\begin{aligned} C^2 &= \mathbf{C}\cdot\mathbf{C} = (\mathbf{A}+\mathbf{B})\cdot(\mathbf{A}+\mathbf{B}) \\ &= A^2 + B^2 + 2\mathbf{A}\cdot\mathbf{B} = A^2 + B^2 + 2AB\cos\theta, \end{aligned}$$

which is the law of cosines.

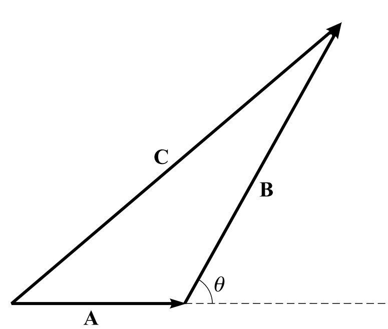


Figure 1.7. Law of cosines.

A simple application of the scalar product in physics is the work W done by a constant force $\mathbf{F}$: $W = \mathbf{F} \cdot \mathbf{r}$, where $\mathbf{r}$ is the displacement vector of the object moved by $\mathbf{F}$.

The vector (cross or outer) product

The vector product of two vectors $\mathbf{A}$ and $\mathbf{B}$ is a vector and is written as

$$\mathbf{C} = \mathbf{A} \times \mathbf{B}. \tag{1.11}$$

As shown in Fig. 1.8, the two vectors $\mathbf{A}$ and $\mathbf{B}$ form two sides of a parallelogram. We define $\mathbf{C}$ to be perpendicular to the plane of this parallelogram with its magnitude equal to the area of the parallelogram. And we choose the direction of $\mathbf{C}$ along the thumb of the right hand when the fingers rotate from $\mathbf{A}$ to $\mathbf{B}$ (angle of rotation less than 180°).

$$\mathbf{C} = \mathbf{A} \times \mathbf{B} = AB \sin\theta \hat{e}_C \qquad (0 \le \theta \le \pi). \tag{1.12}$$

From the definition of the vector product and following the right hand rule, we can see immediately that

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A}. \tag{1.13}$$

Hence the vector product is not commutative. If $\mathbf{A}$ and $\mathbf{B}$ are parallel, then it follows from Eq. (1.12) that

$$\mathbf{A} \times \mathbf{B} = 0. \tag{1.14}$$

In particular

$$\mathbf{A} \times \mathbf{A} = 0. \tag{1.14a}$$

In vector components, we have

$$\mathbf{A} \times \mathbf{B} = (A_1 \hat{e}_1 + A_2 \hat{e}_2 + A_3 \hat{e}_3) \times (B_1 \hat{e}_1 + B_2 \hat{e}_2 + B_3 \hat{e}_3). \tag{1.15}$$

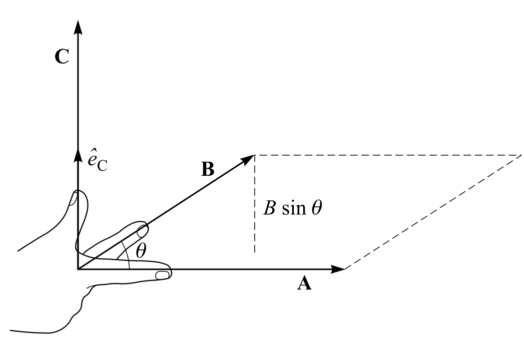


Figure 1.8. The right hand rule for vector product.

Using the following relations

$$\hat{e}_i \times \hat{e}_i = 0, \ i = 1, 2, 3,$$
$$\hat{e}_1 \times \hat{e}_2 = \hat{e}_3, \ \hat{e}_2 \times \hat{e}_3 = \hat{e}_1, \ \hat{e}_3 \times \hat{e}_1 = \hat{e}_2, \tag{1.16}$$

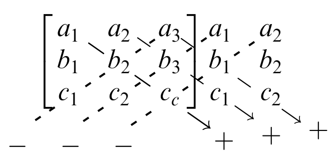
Eq. (1.15) becomes

$$\mathbf{A} \times \mathbf{B} = (A_2B_3 - A_3B_2)\hat{e}_1 + (A_3B_1 - A_1B_3)\hat{e}_2 + (A_1B_2 - A_2B_1)\hat{e}_3. \tag{1.15a}$$

This can be written as an easily remembered determinant of third order:

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \hat{e}_1 & \hat{e}_2 & \hat{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix}. \tag{1.17}$$

The expansion of a determinant of third order can be obtained by diagonal multiplication by repeating on the right the first two columns of the determinant and adding the signed products of the elements on the various diagonals in the resulting array:



The non-commutativity of the vector product of two vectors now appears as a consequence of the fact that interchanging two rows of a determinant changes its sign, and the vanishing of the vector product of two vectors in the same direction appears as a consequence of the fact that a determinant vanishes if one of its rows is a multiple of another.

The determinant is a basic tool used in physics and engineering. The reader is assumed to be familiar with this subject. Those who are in need of review should read Appendix II.

The vector resulting from the vector product of two vectors is called an axial vector, while ordinary vectors are sometimes called polar vectors. Thus, in Eq. (1.11), **C** is a pseudovector, while **A** and **B** are axial vectors. On an inversion of coordinates, polar vectors change sign but an axial vector does not change sign.

A simple application of the vector product in physics is the torque $\boldsymbol{\tau}$ of a force **F** about a point O: $\boldsymbol{\tau} = \mathbf{F} \times \mathbf{r}$, where **r** is the vector from O to the initial point of the force **F** (Fig. 1.9).

We can write the nine equations implied by Eq. (1.16) in terms of permutation symbols ε_{ijk}:

$$\hat{e}_i \times \hat{e}_j = \varepsilon_{ijk}\hat{e}_k, \tag{1.16a}$$

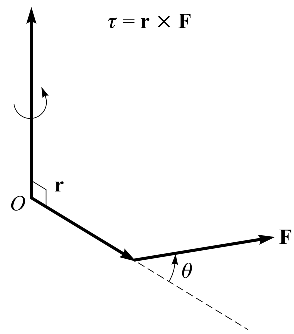


Figure 1.9. The torque of a force about a point O.

where ε_{ijk} is defined by

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } (i, j, k) \text{ is an even permutation of } (1, 2, 3), \\ -1 & \text{if } (i, j, k) \text{ is an odd permutation of } (1, 2, 3), \\ 0 & \text{otherwise (for example, if 2 or more indices are equal).} \end{cases} \tag{1.18}$$

It follows immediately that

$$\varepsilon_{ijk} = \varepsilon_{kij} = \varepsilon_{jki} = -\varepsilon_{jik} = -\varepsilon_{kji} = -\varepsilon_{ikj}.$$

There is a very useful identity relating the ε_{ijk} and the Kronecker delta symbol:

$$\sum_{k=1}^{3} \varepsilon_{mnk}\varepsilon_{ijk} = \delta_{mi}\delta_{nj} - \delta_{mj}\delta_{ni}, \tag{1.19}$$

$$\sum_{j,k} \varepsilon_{mjk}\varepsilon_{njk} = 2\delta_{mn}, \quad \sum_{i,j,k} \varepsilon_{ijk}^2 = 6. \tag{1.19a}$$

Using permutation symbols, we can now write the vector product $\mathbf{A} \times \mathbf{B}$ as

$$\mathbf{A} \times \mathbf{B} = \left(\sum_{i=1}^{3} A_i \hat{e}_i \right) \times \left(\sum_{j=1}^{3} B_j \hat{e}_j \right) = \sum_{i,j}^{3} A_i B_j \left(\hat{e}_i \times \hat{e}_j \right) = \sum_{i,j,k}^{3} \left(A_i B_j \varepsilon_{ijk} \right) \hat{e}_k.$$

Thus the kth component of $\mathbf{A} \times \mathbf{B}$ is

$$(\mathbf{A} \times \mathbf{B})_k = \sum_{i,j} A_i B_j \varepsilon_{ijk} = \sum_{i,j} \varepsilon_{kij} A_i B_j.$$

If $k = 1$, we obtain the usual geometrical result:

$$(\mathbf{A} \times \mathbf{B})_1 = \sum_{i,j} \varepsilon_{1ij} A_i B_j = \varepsilon_{123} A_2 B_3 + \varepsilon_{132} A_3 B_2 = A_2 B_3 - A_3 B_2.$$

The triple scalar product $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$

We now briefly discuss the scalar $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$. This scalar represents the volume of the parallelepiped formed by the coterminous sides **A**, **B**, **C**, since

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = ABC \sin\theta \cos\alpha = hS = \text{volume},$$

S being the area of the parallelogram with sides **B** and **C**, and h the height of the parallelogram (Fig. 1.10).

Now

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = (A_1\hat{e}_1 + A_2\hat{e}_2 + A_3\hat{e}_3) \cdot \begin{vmatrix} \hat{e}_1 & \hat{e}_2 & \hat{e}_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix}$$

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

so that

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix}. \tag{1.20}$$

The exchange of two rows (or two columns) changes the sign of the determinant but does not change its absolute value. Using this property, we find

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix} = -\begin{vmatrix} C_1 & C_2 & C_3 \\ B_1 & B_2 & B_3 \\ A_1 & A_2 & A_3 \end{vmatrix} = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}),$$

that is, the dot and the cross may be interchanged in the triple scalar product.

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C} \tag{1.21}$$

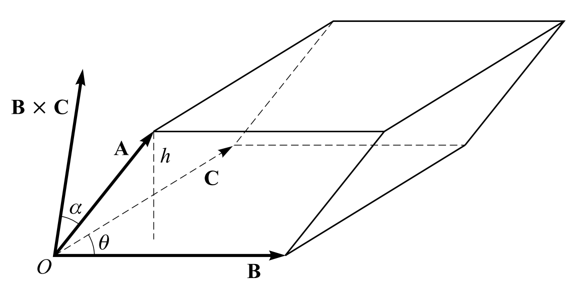


Figure 1.10. The triple scalar product of three vectors **A**, **B**, **C**.

In fact, as long as the three vectors appear in cyclic order, $\mathbf{A} \to \mathbf{B} \to \mathbf{C} \to \mathbf{A}$, then the dot and cross may be inserted between any pairs:

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

It should be noted that the scalar resulting from the triple scalar product changes sign on an inversion of coordinates. For this reason, the triple scalar product is sometimes called a pseudoscalar.

The triple vector product

The triple product $\mathbf{A} \times (\mathbf{B} \times \mathbf{C})$ is a vector, since it is the vector product of two vectors: $\mathbf{A}$ and $\mathbf{B} \times \mathbf{C}$. This vector is perpendicular to $\mathbf{B} \times \mathbf{C}$ and so it lies in the plane of $\mathbf{B}$ and $\mathbf{C}$. If $\mathbf{B}$ is not parallel to $\mathbf{C}$, $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = x\mathbf{B} + y\mathbf{C}$. Now dot both sides with $\mathbf{A}$ and we obtain $x(\mathbf{A} \cdot \mathbf{B}) + y(\mathbf{A} \cdot \mathbf{C}) = 0$, since $\mathbf{A} \cdot [\mathbf{A} \times (\mathbf{B} \times \mathbf{C})] = 0$. Thus

$$x/(\mathbf{A} \cdot \mathbf{C}) = -y/(\mathbf{A} \cdot \mathbf{B}) \equiv \lambda \quad (\lambda \text{ is a scalar})$$

and so

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = x\mathbf{B} + y\mathbf{C} = \lambda[\mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})].$$

We now show that $\lambda = 1$. To do this, let us consider the special case when $\mathbf{B} = \mathbf{A}$. Dot the last equation with $\mathbf{C}$:

$$\mathbf{C} \times [\mathbf{A} \times (\mathbf{A} \times \mathbf{C})] = \lambda[(\mathbf{A} \cdot \mathbf{C})^2 - \mathbf{A}^2\mathbf{C}^2],$$

or, by an interchange of dot and cross

$$-(\mathbf{A} \cdot \mathbf{C})^2 = \lambda[(\mathbf{A} \cdot \mathbf{C})^2 - \mathbf{A}^2\mathbf{C}^2].$$

In terms of the angles between the vectors and their magnitudes the last equation becomes

$$-A^2C^2 \sin^2\theta = \lambda(A^2C^2\cos^2\theta - A^2C^2) = -\lambda A^2C^2\sin^2\theta;$$

hence $\lambda = 1$. And so

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}). \qquad (1.22)$$

Change of coordinate system

Vector equations are independent of the coordinate system we happen to use. But the components of a vector quantity are different in different coordinate systems. We now make a brief study of how to represent a vector in different coordinate systems. As the rectangular Cartesian coordinate system is the basic type of coordinate system, we shall limit our discussion to it. Other coordinate systems will be introduced later. Consider the vector **A** expressed in terms of the unit coordinate vectors $(\hat{e}_1, \hat{e}_2, \hat{e}_3)$:

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

Relative to a new system $(\hat{e}_1', \hat{e}_2', \hat{e}_3')$ that has a different orientation from that of the old system $(\hat{e}_1, \hat{e}_2, \hat{e}_3)$, vector **A** is expressed as

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

Note that the dot product $\mathbf{A} \cdot \hat{e}_1'$ is equal to A_1', the projection of **A** on the direction of $\hat{e}_1'$; $\mathbf{A} \cdot \hat{e}_2'$ is equal to A_2', and $\mathbf{A} \cdot \hat{e}_3'$ is equal to A_3'. Thus we may write

$$\left.\begin{aligned} A_1' &= (\hat{e}_1 \cdot \hat{e}_1')A_1 + (\hat{e}_2 \cdot \hat{e}_1')A_2 + (\hat{e}_3 \cdot \hat{e}_1')A_3, \\ A_2' &= (\hat{e}_1 \cdot \hat{e}_2')A_1 + (\hat{e}_2 \cdot \hat{e}_2')A_2 + (\hat{e}_3 \cdot \hat{e}_2')A_3, \\ A_3' &= (\hat{e}_1 \cdot \hat{e}_3')A_1 + (\hat{e}_2 \cdot \hat{e}_3')A_2 + (\hat{e}_3 \cdot \hat{e}_3')A_3. \end{aligned}\right\} \qquad (1.23)$$

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

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

The 3×3 matrix in the above equation is called the rotation (or transformation) matrix, and is an orthogonal matrix. One advantage of using a matrix is that successive transformations can be handled easily by means of matrix multiplication. Let us digress for a quick review of some basic matrix algebra. A full account of matrix method is given in Chapter 3.

A matrix is an ordered array of scalars that obeys prescribed rules of addition and multiplication. A particular matrix element is specified by its row number followed by its column number. Thus a_{ij} is the matrix element in the ith row and jth column. Alternative ways of representing matrix $\tilde{A}$ are $[a_{ij}]$ or the entire array

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

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

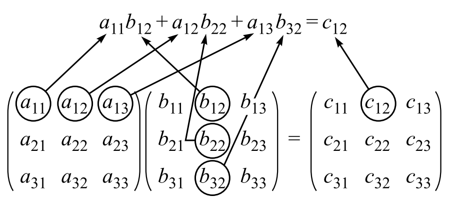
$$\tilde{B} = (b_{11}\ b_{12}\ b_{13}) \quad \text{or} \quad \tilde{C} = \begin{pmatrix} c_{11} \\ c_{21} \\ c_{31} \end{pmatrix},$$

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

The multiplication of a matrix $\tilde{A}$ and a matrix $\tilde{B}$ is defined only when the number of columns of $\tilde{A}$ is equal to the number of rows of $\tilde{B}$, and is performed in the same way as the multiplication of two determinants: if $\tilde{C} = \tilde{A}\tilde{B}$, then

$$c_{ij} = \sum_k a_{ik} b_{kl}.$$

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



If we denote the direction cosines $\hat{e}'_i \cdot \hat{e}_j$ by λ_{ij}, then Eq. (1.23) can be written as

$$A'_i = \sum_{j=1}^{3} \hat{e}'_i \cdot \hat{e}_j A_j = \sum_{j=1}^{3} \lambda_{ij} A_j. \tag{1.23a}$$

It can be shown (Problem 1.9) that the quantities λ_{ij} satisfy the following relations

$$\sum_{i=1}^{3} \lambda_{ij} \lambda_{ik} = \delta_{jk} \quad (j, k = 1, 2, 3). \tag{1.24}$$

Any linear transformation, such as Eq. (1.23a), that has the properties required by Eq. (1.24) is called an orthogonal transformation, and Eq. (1.24) is known as the orthogonal condition.

The linear vector space V_n

We have found that it is very convenient to use vector components, in particular, the unit coordinate vectors $\hat{e}_i$ ($i = 1, 2, 3$). The three unit vectors $\hat{e}_i$ are orthogonal and normal, or, as we shall say, orthonormal. This orthonormal property is conveniently written as Eq. (1.8). But there is nothing special about these orthonormal unit vectors $\hat{e}_i$. If we refer the components of the vectors to a different system of rectangular coordinates, we need to introduce another set of three orthonormal unit vectors $\hat{f}_1, \hat{f}_2$, and $\hat{f}_3$:

$$\hat{f}_i \hat{f}_j = \delta_{ij} \quad (i, j = 1, 2, 3). \tag{1.8a}$$

For any vector **A** we now write

$$\mathbf{A} = \sum_{i=1}^{3} c_i \hat{f}_i, \quad \text{and} \quad c_i = \hat{f}_i \cdot \mathbf{A}.$$

We see that we can define a large number of different coordinate systems. But the physically significant quantities are the vectors themselves and certain functions of these, which are independent of the coordinate system used. The orthonormal condition (1.8) or (1.8a) is convenient in practice. If we also admit oblique Cartesian coordinates then the $\hat{f}_i$ need neither be normal nor orthogonal; they could be any three non-coplanar vectors, and any vector **A** can still be written as a linear superposition of the $\hat{f}_i$

$$\mathbf{A} = c_1 \hat{f}_1 + c_2 \hat{f}_2 + c_3 \hat{f}_3. \tag{1.25}$$

Starting with the vectors $\hat{f}_i$, we can find linear combinations of them by the algebraic operations of vector addition and multiplication of vectors by scalars, and then the collection of all such vectors makes up the three-dimensional linear space often called V_3 (V for vector) or R_3 (R for real) or E_3 (E for Euclidean). The vectors $\hat{f}_1, \hat{f}_2, \hat{f}_3$ are called the base vectors or bases of the vector space V_3. Any set of vectors, such as the $\hat{f}_i$, which can serve as the bases or base vectors of V_3 is called complete, and we say it spans the linear vector space. The base vectors are also linearly independent because no relation of the form

$$c_1 \hat{f}_1 + c_2 \hat{f}_2 + c_3 \hat{f}_3 = 0 \tag{1.26}$$

exists between them, unless $c_1 = c_2 = c_3 = 0$.

The notion of a vector space is much more general than the real vector space V_3. Extending the concept of V_3, it is convenient to call an ordered set of n matrices, or functions, or operators, a 'vector' (or an n-vector) in the n-dimensional space V_n. Chapter 5 will provide justification for doing this. Taking a cue from V_3, vector addition in V_n is defined to be

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n) \tag{1.27}$$

and multiplication by scalars is defined by

$$\alpha(x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n), \tag{1.28}$$

where α is real. With these two algebraic operations of vector addition and multiplication by scalars, we call V_n a vector space. In addition to this algebraic structure, V_n has geometric structure derived from the length defined to be

$$\left(\sum_{j=1}^{n} x_j^2\right)^{1/2} = \sqrt{x_1^2 + \cdots + x_n^2} \tag{1.29}$$

The dot product of two n-vectors can be defined by

$$(x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) = \sum_{j=1}^{n} x_j y_j. \tag{1.30}$$

In V_n, vectors are not directed line segments as in V_3; they may be an ordered set of n operators, matrices, or functions. We do not want to become sidetracked from our main goal of this chapter, so we end our discussion of vector space here.

Vector differentiation

Up to this point we have been concerned mainly with vector algebra. A vector may be a function of one or more scalars and vectors. We have encountered, for example, many important vectors in mechanics that are functions of time and position variables. We now turn to the study of the calculus of vectors.

Physicists like the concept of field and use it to represent a physical quantity that is a function of position in a given region. Temperature is a scalar field, because its value depends upon location: to each point (x, y, z) is associated a temperature $T(x, y, z)$. The function $T(x, y, z)$ is a scalar field, whose value is a real number depending only on the point in space but not on the particular choice of the coordinate system. A vector field, on the other hand, associates with each point a vector (that is, we associate three numbers at each point), such as the wind velocity or the strength of the electric or magnetic field. When described in a rotated system, for example, the three components of the vector associated with one and the same point will change in numerical value. Physically and geometrically important concepts in connection with scalar and vector fields are the gradient, divergence, curl, and the corresponding integral theorems.

The basic concepts of calculus, such as continuity and differentiability, can be naturally extended to vector calculus. Consider a vector **A**, whose components are functions of a single variable u. If the vector **A** represents position or velocity, for example, then the parameter u is usually time t, but it can be any quantity that determines the components of **A**. If we introduce a Cartesian coordinate system, the vector function $\mathbf{A}(u)$ may be written as

$$\mathbf{A}(u) = A_1(u)\hat{e}_1 + A_2(u)\hat{e}_2 + A_3(u)\hat{e}_3. \tag{1.31}$$

$\mathbf{A}(u)$ is said to be continuous at $u = u_0$ if it is defined in some neighborhood of u_0 and

$$\lim_{u \to u_0} A(u) = A(u_0). \tag{1.32}$$

Note that $\mathbf{A}(u)$ is continuous at u_0 if and only if its three components are continuous at u_0.

$\mathbf{A}(u)$ is said to be differentiable at a point u if the limit

$$\frac{d\mathbf{A}(u)}{du} = \lim_{\Delta u \to 0} \frac{\mathbf{A}(u + \Delta u) - \mathbf{A}(u)}{\Delta u} \tag{1.33}$$

exists. The vector $\mathbf{A}'(u) = d\mathbf{A}(u)/du$ is called the derivative of $\mathbf{A}(u)$; and to differentiate a vector function we differentiate each component separately:

$$\mathbf{A}'(u) = A_1'(u)\hat{e}_1 + A_2'(u)\hat{e}_2 + A_3'(u)\hat{e}_3. \tag{1.33a}$$

Note that the unit coordinate vectors are fixed in space. Higher derivatives of $\mathbf{A}(u)$ can be similarly defined.

If $\mathbf{A}$ is a vector depending on more than one scalar variable, say u, v for example, we write $\mathbf{A} = \mathbf{A}(u, v)$. Then

$$d\mathbf{A} = (\partial \mathbf{A}/\partial u)du + (\partial \mathbf{A}/\partial v)dv \tag{1.34}$$

is the differential of $\mathbf{A}$, and

$$\frac{\partial \mathbf{A}}{\partial u} = \lim_{\Delta u \to 0} \frac{\mathbf{A}(u + \Delta u, v) - \mathbf{A}(u, v)}{\partial u} \tag{1.34a}$$

and similarly for $\partial \mathbf{A}/\partial v$.

Derivatives of products obey rules similar to those for scalar functions. However, when cross products are involved the order may be important.

Space curves

As an application of vector differentiation, let us consider some basic facts about curves in space. If $\mathbf{A}(u)$ is the position vector $\mathbf{r}(u)$ joining the origin of a coordinate system and any point $P(x_1, x_2, x_3)$ in space as shown in Fig. 1.11, then Eq. (1.31) becomes

$$\mathbf{r}(u) = x_1(u)\hat{e}_1 + x_2(u)\hat{e}_2 + x_3(u)\hat{e}_3. \tag{1.35}$$

As u changes, the terminal point P of $\mathbf{r}$ describes a curve C in space. Eq. (1.35) is called a parametric representation of the curve C, and u is the parameter of this representation. Then

$$\frac{\Delta \mathbf{r}}{\Delta u} \left(= \frac{\mathbf{r}(u + \Delta u) - \mathbf{r}(u)}{\Delta u} \right)$$

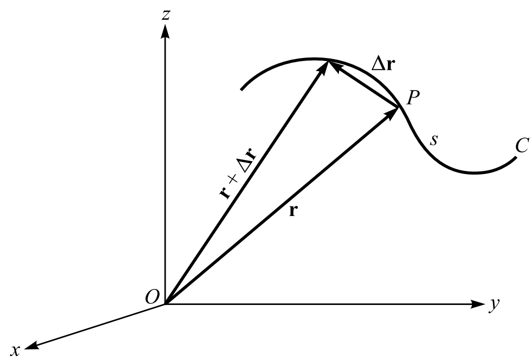


Figure 1.11. Parametric representation of a curve.

is a vector in the direction of $\Delta\mathbf{r}$, and its limit (if it exists) $d\mathbf{r}/du$ is a vector in the direction of the tangent to the curve at (x_1, x_2, x_3). If u is the arc length s measured from some fixed point on the curve C, then $d\mathbf{r}/ds = \hat{T}$ is a unit tangent vector to the curve C. The rate at which $\hat{T}$ changes with respect to s is a measure of the curvature of C and is given by $d\hat{T}/ds$. The direction of $d\hat{T}/ds$ at any given point on C is normal to the curve at that point: $\hat{T} \cdot \hat{T} = 1$, $d(\hat{T} \cdot \hat{T})/ds = 0$, from this we get $\hat{T} \cdot d\hat{T}/ds = 0$, so they are normal to each other. If $\hat{N}$ is a unit vector in this normal direction (called the principal normal to the curve), then $d\hat{T}/ds = \kappa\hat{N}$, and κ is called the curvature of C at the specified point. The quantity $\rho = 1/\kappa$ is called the radius of curvature. In physics, we often study the motion of particles along curves, so the above results may be of value.

In mechanics, the parameter u is time t, then $d\mathbf{r}/dt = \mathbf{v}$ is the velocity of the particle which is tangent to the curve at the specific point. Now we can write

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}}{ds}\frac{ds}{dt} = v\hat{T}$$

where v is the magnitude of $\mathbf{v}$, called the speed. Similarly, $\mathbf{a} = d\mathbf{v}/dt$ is the acceleration of the particle.

Motion in a plane

Consider a particle P moving in a plane along a curve C (Fig. 1.12). Now $\mathbf{r} = r\hat{e}_r$, where $\hat{e}_r$ is a unit vector in the direction of $\mathbf{r}$. Hence

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{dr}{dt}\hat{e}_r + r\frac{d\hat{e}_r}{dt}.$$

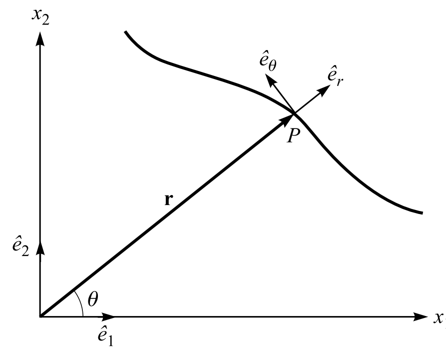


Figure 1.12. Motion in a plane.

Now $d\hat{e}_r/dt$ is perpendicular to $\hat{e}_r$. Also $|d\hat{e}_r/dt| = d\theta/dt$; we can easily verify this by differentiating $\hat{e}_r = \cos\theta\hat{e}_1 + \sin\theta\hat{e}_2$. Hence

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{dr}{dt}\hat{e}_r + r\frac{d\theta}{dt}\hat{e}_\theta;$$

$\hat{e}_\theta$ is a unit vector perpendicular to $\hat{e}_r$.

Differentiating again we obtain

$$\begin{aligned}\mathbf{a} = \frac{d\mathbf{v}}{dt} &= \frac{d^2r}{dt^2}\hat{e}_r + \frac{dr}{dt}\frac{d\hat{e}_r}{dt} + \frac{dr}{dt}\frac{d\theta}{dt}\hat{e}_\theta + r\frac{d^2\theta}{dt^2}\hat{e}_\theta + r\frac{d\theta}{dt}\hat{e}_\theta \\ &= \frac{d^2r}{dt^2}\hat{e}_r + 2\frac{dr}{dt}\frac{d\theta}{dt}\hat{e}_\theta + r\frac{d^2\theta}{dt^2}\hat{e}_\theta - r\left(\frac{d\theta}{dt}\right)^2\hat{e}_r \quad \left(\because \frac{d\hat{e}_\theta}{dt} = -\frac{d\theta}{dt}\hat{e}_r\right).\end{aligned}$$

Thus

$$\mathbf{a} = \left[\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2\right]\hat{e}_r + \frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right)\hat{e}_\theta.$$

A vector treatment of classical orbit theory

To illustrate the power and use of vector methods, we now employ them to work out the Keplerian orbits. We first prove Kepler's second law which can be stated as: angular momentum is constant in a central force field. A central force is a force whose line of action passes through a single point or center and whose magnitude depends only on the distance from the center. Gravity and electrostatic forces are central forces. A general discussion on central force can be found in, for example, Chapter 6 of *Classical Mechanics*, Tai L. Chow, John Wiley, New York, 1995.

Differentiating the angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ with respect to time, we obtain

$$d\mathbf{L}/dt = d\mathbf{r}/dt \times \mathbf{p} + \mathbf{r} \times d\mathbf{p}/dt.$$

The first vector product vanishes because $\mathbf{p} = md\mathbf{r}/dt$ so $d\mathbf{r}/dt$ and $\mathbf{p}$ are parallel. The second vector product is simply $\mathbf{r} \times \mathbf{F}$ by Newton's second law, and hence vanishes for all forces directed along the position vector $\mathbf{r}$, that is, for all central forces. Thus the angular momentum $\mathbf{L}$ is a constant vector in central force motion. This implies that the position vector $\mathbf{r}$, and therefore the entire orbit, lies in a fixed plane in three-dimensional space. This result is essentially Kepler's second law, which is often stated in terms of the conservation of area velocity, $|\mathbf{L}|/2m$.

We now consider the inverse-square central force of gravitational and electrostatics. Newton's second law then gives

$$md\mathbf{v}/dt = -(k/r^2)\hat{n}, \tag{1.36}$$

where $\hat{n} = \mathbf{r}/r$ is a unit vector in the $\mathbf{r}$-direction, and $k = Gm_1m_2$ for the gravitational force, and $k = q_1q_2$ for the electrostatic force in cgs units. First we note that

$$\mathbf{v} = d\mathbf{r}/dt = dr/dt\hat{n} + rd\hat{n}/dt.$$

Then $\mathbf{L}$ becomes

$$\mathbf{L} = \mathbf{r} \times (m\mathbf{v}) = mr^2[\hat{n} \times (d\hat{n}/dt)]. \tag{1.37}$$

Now consider

$$\frac{d}{dt}(\mathbf{v} \times \mathbf{L}) = \frac{d\mathbf{v}}{dt} \times \mathbf{L} = -\frac{k}{mr^2}(\hat{n} \times \mathbf{L}) = -\frac{k}{mr^2}[\hat{n} \times mr^2(\hat{n} \times d\hat{n}/dt)]$$
$$= -k[\hat{n}(d\hat{n}/dt \cdot \hat{n}) - (d\hat{n}/dt)(\hat{n} \cdot \hat{n})].$$

Since $\hat{n} \cdot \hat{n} = 1$, it follows by differentiation that $\hat{n} \cdot d\hat{n}/dt = 0$. Thus we obtain

$$\frac{d}{dt}(\mathbf{v} \times \mathbf{L}) = kd\hat{n}/dt;$$

integration gives

$$\mathbf{v} \times \mathbf{L} = k\hat{n} + \mathbf{C}, \tag{1.38}$$

where $\mathbf{C}$ is a constant vector. It lies along, and fixes the position of, the major axis of the orbit as we shall see after we complete the derivation of the orbit. To find the orbit, we form the scalar quantity

$$L^2 = \mathbf{L} \cdot (\mathbf{r} \times m\mathbf{v}) = m\mathbf{r} \cdot (\mathbf{v} \times \mathbf{L}) = mr(k + C\cos\theta), \tag{1.39}$$

where θ is the angle measured from $\mathbf{C}$ (which we may take to be the x-axis) to $\mathbf{r}$. Solving for r, we obtain

$$r = \frac{L^2/km}{1 + C/(k\cos\theta)} = \frac{A}{1 + \varepsilon\cos\theta}. \tag{1.40}$$

Eq. (1.40) is a conic section with one focus at the origin, where ε represents the eccentricity of the conic section; depending on its values, the conic section may be a circle, an ellipse, a parabola, or a hyperbola. The eccentricity can be easily determined in terms of the constants of motion:

$$\varepsilon = \frac{C}{k} = \frac{1}{k}|(\mathbf{v} \times \mathbf{L}) - k\hat{n}|$$
$$= \frac{1}{k}[|\mathbf{v} \times \mathbf{L}|^2 + k^2 - 2k\hat{n} \cdot (\mathbf{v} \times \mathbf{L})]^{1/2}$$

Now $|\mathbf{v} \times \mathbf{L}|^2 = v^2 L^2$ because $\mathbf{v}$ is perpendicular to $\mathbf{L}$. Using Eq. (1.39), we obtain

$$\varepsilon = \frac{1}{k}\left[v^2L^2 + k^2 - \frac{2kL^2}{mr}\right]^{1/2} = \left[1 + \frac{2L^2}{mk^2}\left(\frac{1}{2}mv^2 - \frac{k}{r}\right)\right]^{1/2} = \left[1 + \frac{2L^2E}{mk^2}\right]^{1/2},$$

where E is the constant energy of the system.

Vector differentiation of a scalar field and the gradient

Given a scalar field in a certain region of space given by a scalar function $\phi(x_1, x_2, x_3)$ that is defined and differentiable at each point with respect to the position coordinates (x_1, x_2, x_3), the total differential corresponding to an infinitesimal change $d\mathbf{r} = (dx_1, dx_2, dx_3)$ is

$$d\phi = \frac{\partial \phi}{\partial x_1}dx_1 + \frac{\partial \phi}{\partial x_2}dx_2 + \frac{\partial \phi}{\partial x_3}dx_3. \tag{1.41}$$

We can express $d\phi$ as a scalar product of two vectors:

$$d\phi = \frac{\partial \phi}{\partial x_1}dx_1 + \frac{\partial \phi}{\partial x_2}dx_2 + \frac{\partial \phi}{\partial x_3}dx_3 = (\nabla\phi) \cdot d\mathbf{r}, \tag{1.42}$$

where

$$\nabla\phi \equiv \frac{\partial \phi}{\partial x_1}\hat{e}_1 + \frac{\partial \phi}{\partial x_2}\hat{e}_2 + \frac{\partial \phi}{\partial x_3}\hat{e}_3 \tag{1.43}$$

is a vector field (or a vector point function). By this we mean to each point $\mathbf{r} = (x_1, x_2, x_3)$ in space we associate a vector $\nabla\phi$ as specified by its three components $(\partial\phi/\partial x_1, \partial\phi/\partial x_2, \partial\phi/\partial x_3)$: $\nabla\phi$ is called the *gradient* of ϕ and is often written as grad ϕ.

There is a simple geometric interpretation of $\nabla\phi$. Note that $\phi(x_1, x_2, x_3) = c$, where c is a constant, represents a surface. Let $\mathbf{r} = x_1\hat{e}_1 + x_2\hat{e}_2 + x_3\hat{e}_3$ be the position vector to a point $P(x_1, x_2, x_3)$ on the surface. If we move along the surface to a nearby point $Q(\mathbf{r} + d\mathbf{r})$, then $d\mathbf{r} = dx_1\hat{e}_1 + dx_2\hat{e}_2 + dx_3\hat{e}_3$ lies in the tangent plane to the surface at P. But as long as we move along the surface ϕ has a constant value and $d\phi = 0$. Consequently from (1.41),

$$d\mathbf{r} \cdot \nabla\phi = 0. \tag{1.44}$$