1

Introduction and Mathematics Review

1.1 The Nature of Celestial Mechanics

Celestial mechanics has a long and venerable history as a discipline. It would be fair to say that it was the first area of physical science to emerge from Newton's theory of mechanics and gravitation put forth in the Principia. It was Newton's ability to describe accurately the motion of the planets under the concept of a single universal set of laws that led to his fame in the seventeenth century. The application of Newtonian mechanics to planetary motion was honed to so fine an edge during the next two centuries that by the advent of the twentieth century the description of planetary motion was refined enough that the departure of prediction from observation by 43 arcsec in the precession of the perihelion of Mercury's orbit was a major factor in the replacement of Newton's theory of gravity by the General Theory of Relativity.

At the turn of the century no professional astronomer would have been considered properly educated if he could not determine the location of a planet in the local sky given the orbital elements of that planet. The reverse would also have been expected. That is, given three or more positions of the planet in the sky for three different dates, he should be able to determine the orbital elements of that planet preferably in several ways. It is reasonably safe to say that few contemporary astronomers could accomplish this without considerable study. The emphasis of astronomy has shifted dramatically during the past fifty years. The techniques of classical celestial mechanics developed by Gauss, Lagrange, Euler and many others have more or less been consigned to the history books. Even in the situation where the orbits of spacecraft are required, the accuracy demanded is such that much more complicated mechanics is necessary than for planetary motion, and these problems tend to be dealt with by techniques suited to modern computers.

However, the foundations of classical celestial mechanics contain elements of modern physics that should be understood by every physical scientist. It is the understanding of these elements that will form the primary aim of the book while their application to celestial mechanics will be incidental. A mastery of these fundamentals will enable the student to perform those tasks required of an astronomer at the turn of the century and also equip him to deal with more complicated problems in many other fields.

The traditional approach to celestial mechanics well into the twentieth century was incredibly narrow and encumbered with an unwieldy notation that tended to confound rather than elucidate. It wasn't until the 1950s that vector notation was even introduced into the subject at the textbook level. Since throughout this book we shall use the now familiar vector notation along with the broader view of classical mechanics and linear algebra, it is appropriate that we begin with a review of some of these concepts.

1.2 Scalars, Vectors, Tensors, Matrices and Their Products

While most students of the physical sciences have encountered scalars and vectors throughout their college career, few have had much to do with tensors and fewer still have considered the relations between these concepts. Instead they are regarded as separate entities to be used under separate and specific conditions. Other students regard tensors as the unfathomable language of General Relativity and therefore comprehensible only to the intellectually elite. This latter situation is unfortunate since tensors are incredibly useful in the wide range of modern theoretical physics and the sooner one vanquishes his fear of them the better. Thus, while we won't make great use of them in this book, we will introduce them and describe their relationship to vectors and scalars.

a. Scalars

The notion of a scalar is familiar to anyone who has completed a freshman course in physics. A single number or symbol is used to describe some physical quantity. In truth, as any mathematician will tell you, it is not necessary for the scalar to represent anything physical. But since this is a book about physical science we shall narrow our view to the physical world. There is, however, an area of mathematics that does provide a basis for defining scalars, vectors, etc. That area is set theory and its more specialized counterpart, group theory. For a collection or set of objects to form a group there must be certain relations between the elements of the set. Specifically, there must be a "Law" which describes the result of "combining" two members of the set. Such a "Law" could be *addition*. Now if the action of the law upon any two members of the set produces a third member of the set, the set is said to be "closed" with respect to that law. If the set contains an element which, when combined under the law with any other member of the set, yields that member unchanged, that element is said to be the identity element. Finally, if the set contains elements which are inverses, so that the combination of a member of the set with its inverse under the "Law" yields the identity element, then the set is said to form a group under the "Law".

The integers (positive and negative, including zero) form a group under addition. In this instance, the identity element is zero and the operation that generates inverses is subtraction so that the negative integers represent the inverse elements of the positive integers. However, they do not form a group under multiplication as each inverse is a fraction. On the other hand the rational numbers do form a group under both addition and multiplication. Here the identity element for addition is again zero, but under multiplication it is one. The same is true for the real and complex numbers. Groups have certain nice properties; thus it is useful to know if the set of objects forms a group or not. Since scalars are generally used to represent real or complex numbers in the physical world, it is nice to know that they will form a group under multiplication and addition so that the inverse operations of subtraction and division are defined. With that notion alone one can develop all of algebra and calculus which are so useful in describing the physical world. However, the notion of a vector is also useful for describing the physical world and we shall now look at their relation to scalars.

b. Vectors

A vector has been defined as "an ordered n-tuple of numbers". Most find that this technically correct definition needs some explanation. There are some physical quantities that require more than a single number to fully describe them. Perhaps the most obvious is an object's location in space. Here we require three numbers to define its location (four if we include time). If we insist that the order of those three numbers be the same, then we can represent them by a single symbol called a vector. In general, vectors need not be limited to three numbers; one may use as many as is necessary to characterize the quantity. However, it would be useful if the vectors also formed a group and for this we need some "Laws" for which the group is closed. Again addition and multiplication seem to be the logical laws to impose. Certainly vector addition satisfies the group condition, namely that the application of the "law" produces an element of the set. The identity element is a 'zero-vector' whose components are all zero. However, the commonly defined "laws" of multiplication do not satisfy this condition.

Consider the vector scalar product, also known as the inner product, which is defined as

$$\vec{A} \bullet \vec{B} = c = \sum_i A_i B_i \tag{1.2.1}$$

Here the result is a scalar which is clearly a different type of quantity than a vector. Now consider the other well known 'vector product', sometimes called the cross product, which in ordinary Cartesian coordinates is defined as

$$\vec{A} \times \vec{B} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ A_i & A_j & A_k \\ B_i & B_j & B_k \end{vmatrix} = \hat{i}(A_j B_k - A_k B_j) - \hat{j}(A_i B_k - A_k B_i) + \hat{k}(A_i B_j - A_j B_i) \,. \tag{1.2.2}$$

This appears to satisfy the condition that the result is a vector. However as we shall see, the vector produced by this operation does not behave in the way in which we would like all vectors to behave.

Finally, there is a product law known as the tensor, or outer product that is useful to define as

$$\left.\begin{array}{l} \vec{A}\vec{B} = \mathbf{C} \quad , \\ C_{ij} = A_i B_j \end{array}\right\} \tag{1.2.3}$$

Here the result of applying the "law" is an ordered array of (n x m) numbers where n and m are the dimensionalities of the vectors $\vec{A}$ and $\vec{B}$ respectively. Such a result is clearly not a vector and so vectors under this law do not form a group. In order to provide a broader concept wherein we can understand scalars and vectors as well as the results of the outer product, let us briefly consider the quantities knows as tensors.

c. Tensors and Matrices

In general a tensor has N^n components or elements. N is known as the dimensionality of the tensor by analogy with the notion of a vector while n is called the rank. Thus vectors are simply tensors of rank unity while scalars are tensors of rank zero. If we consider the set of all tensors, then they form a group under addition and all of the vector products. Indeed the inner product can be generalized for tensors of rank m and n. The result will be a tensor of rank $|m-n|$. Similarly the outer product can be so defined that the outer product of tensors with rank m and n is a tensor of rank $|m+n|$.

One obvious way of representing tensors of rank two is by denoting them as matrices. Thus the arranging of the N^2 components in an (N x N) array will produce the familiar square matrix. The scalar product of a matrix and vector should then yield a vector by

$$\left.\begin{array}{l}\mathbf{A}\bullet\vec{B}=\vec{C} \quad , \\ C_i = \sum_j A_{ij} B_j \end{array}\right\} \quad , \tag{1.2.4}$$

while the outer product would result in a tensor of rank three from

$$\left.\begin{array}{l}\mathbf{A}\vec{B}=\overline{\mathbf{C}} \quad , \\ C_{ijk} = A_{ij} B_k \end{array}\right\} \quad . \tag{1.2.5}$$

An important tensor of rank two is called the unit tensor whose elements are the Kronecker delta and for two dimensions is written as

$$\mathbf{1} = \begin{pmatrix} 1\,0 \\ 0\,1 \end{pmatrix} = \delta_{ij} \quad . \tag{1.2.6}$$

Clearly the scalar product of this tensor with a vector yields the vector itself. There is a parallel tensor of rank three known as the Levi-Civita tensor (or more correctly tensor density) which is a three index tensor whose elements are zero when any two indices are equal. When the indices are all different the value is +l or -1 depending on whether the index sequence can be obtained by an even or odd permutation of 1,2,3 respectively. Thus the elements of the Levi-Civita tensor can be written in terms of three matrices as

$$\left.\varepsilon_{1jk} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & +1 \\ 0 & -1 & 0 \end{pmatrix} \quad , \quad \varepsilon_{2jk} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ +1 & 0 & 0 \end{pmatrix} \quad , \quad \varepsilon_{3jk} = \begin{pmatrix} 0 & +1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right\} \quad . \tag{2.1.7}$$

One of the utilities of this tensor is that it can be used to express the vector cross product as follows

$$\vec{A}\times\vec{B} = \overline{\varepsilon}\bullet(\vec{A}\vec{B}) = \sum_j \sum_k \varepsilon_{ijk} A_j B_k = C_i \quad . \tag{1.2.8}$$

As we shall see later, while the rule for calculating the rank correctly implies that the vector cross product as expressed by equation (1.2.8) will yield a vector, there are reasons for distinguishing between this type of vector and the normal vectors $\vec{A}$ and $\vec{B}$. These same reasons extend to the correct naming of the Levi-Civita tensor as the Levi-Civita tensor density. However, before this distinction can be made clear, we shall have to understand more about coordinate transformations and the behavior of both vectors and tensors that are subject to them.

The normal matrix product is certainly different from the scalar or outer product and serves as an additional multiplication "law" for second rank tensors. The standard definition of the matrix product is

$$\left.\begin{array}{l} \mathbf{AB} = \mathbf{C} \quad , \\ C_{ij} = \sum_k A_{ik} B_{kj} \end{array}\right\} \quad . \tag{1.2.9}$$

Only if the matrices can be resolved into the outer product of two vectors so that

$$\left.\begin{array}{l} \mathbf{A} = \vec{a}\vec{\alpha} \\ \mathbf{B} = \vec{\beta}\vec{b} \end{array}\right\} , \tag{1.2.10}$$

can the matrix product be written in terms of the products that we have already defined -namely

$$\mathbf{AB} = \vec{a}\vec{b}(\vec{\alpha} \bullet \vec{\beta}) \quad . \tag{1.2.11}$$

There is much more that can, and perhaps should, be said about matrices. Indeed, entire books have been written about their properties. However, we shall consider only some of those properties within the notion of a group. Clearly the unit tensor (or unit matrix) given by equation (1.2.6) represents the unit element of the matrix group under matrix multiplication. The unit under addition is simply a matrix whose elements are all zero, since matrix addition is defined by

$$\left.\begin{array}{l} \mathbf{A} + \mathbf{B} = \mathbf{C} \\ \\ A_{ij} + B_{ij} = C_{ij} \end{array}\right\} \quad . \tag{1.2.12}$$

Remember that the unit element of any group forms the definition of the inverse element. Clearly the inverse of a matrix *under addition* will simply be that matrix whose elements are the negative of the original matrix, so that their sum is zero. However, the inverse of a matrix under matrix multiplication is quite another matter. We can certainly define the process by

$$\mathbf{AA^{-1}} = \mathbf{1} \quad , \tag{1.2.13}$$

but the process by which $\mathbf{A^{-1}}$ is actually computed is lengthy and beyond the scope of this book. We can further define other properties of a matrix such as the *transpose* and the *determinant*. The transpose of a matrix A with elements Aij is just

$$\mathbf{A^T} = \mathrm{A}_{ij} \quad , \tag{1.2.14}$$

while the determinant is obtained by expanding the matrix by minors as is done in Kramer's rule for the solution of linear algebraic equations. For a (3 x 3) matrix, this would become

$$\begin{aligned} \det\mathbf{A} = \det\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} &= +a_{11}(a_{22}a_{33} - a_{23}a_{32}) \\ &\quad - a_{12}(a_{21}a_{33} - a_{23}a_{31}) \quad . \\ &\quad + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \end{aligned} \tag{1.2.15}$$

The matrix is said to be symmetric if $\mathrm{A}_{ij} = \mathrm{A}_{ji}$. Finally, if the matrix elements are complex so that the transpose element is the complex conjugate of its counterpart, the matrix is said to be Hermitian. Thus for a Hermitian matrix **H** the elements obey

$$\mathrm{H}_{ij} = \tilde{\mathrm{H}}_{ji} \quad , \tag{1.2.16}$$

where $\tilde{\mathrm{H}}_{ji}$ is the complex conjugate of H_{ij} .

1.3 Commutativity, Associativity, and Distributivity

Any "law" that is defined on the elements of a set may have certain properties that are important for the implementation of that "law" and the resultant elements. For the sake of generality, let us denote the "law" by ^, which can stand for any of the products that we have defined. Now any such law is said to be commutative if

$$A \wedge B = B \wedge A \quad . \tag{1.3.1}$$

Of all the laws we have discussed only addition and the scalar product are commutative. This means that considerable care must be observed when using the outer, vector-cross, or matrix products, as the order in which terms appear in a product will make a difference in the result.

Associativity is a somewhat weaker condition and is said to hold for any law when

$$(A \wedge B) \wedge C = A \wedge (B \wedge C) \quad . \tag{1.3.2}$$

In other words the order in which the law is applied to a string of elements doesn't matter if the law is associative. Here addition, the scalar, and matrix products are associative while the vector cross product and outer product are, in general, not. Finally, the notion of distributivity involves the relation between two different laws. These are usually addition and one of the products. Our general purpose law ^ is said to be distributive with respect to addition if

$$A \wedge (B + C) = (A \wedge B) + (A \wedge C) \,. \tag{1.3.3}$$

This is usually the weakest of all conditions on a law and here all of the products we have defined pass the test. They are all distributive with respect to addition. The main function of remembering the properties of these various products is to insure that mathematical manipulations on expressions involving them are done correctly.

1.4 Operators

The notion of operators is extremely important in mathematical physics and there are entire books written on the subject. Most students usually first encounter operators in calculus when the notation [d/dx] is introduced to denote the derivative of a function. In this instance the operator stands for taking the limit of the difference between adjacent values of some function of x divided by the difference between the adjacent values of x as that difference tends toward zero.

This is a fairly complicated set of instructions represented by a relatively simple set of symbols. The designation of some symbol to represent a collection of operations is said to represent the definition of an operator. Depending on the details of the definition, the operators can often be treated as if they were quantities and subjected to algebraic manipulations. The extent to which this is possible is determined by how well the operators satisfy the conditions for the group on which the algebra or mathematical system in question is defined.

We shall make use of a number of operators in this book, the most common of which is the "del" operator or "nabla". It is usually denoted by the symbol ∇ and is a vector operator defined in Cartesian coordinates as

$$\nabla \equiv \hat{i}\frac{\partial}{\partial x} + \hat{j}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z} \quad . \tag{1.4.1}$$

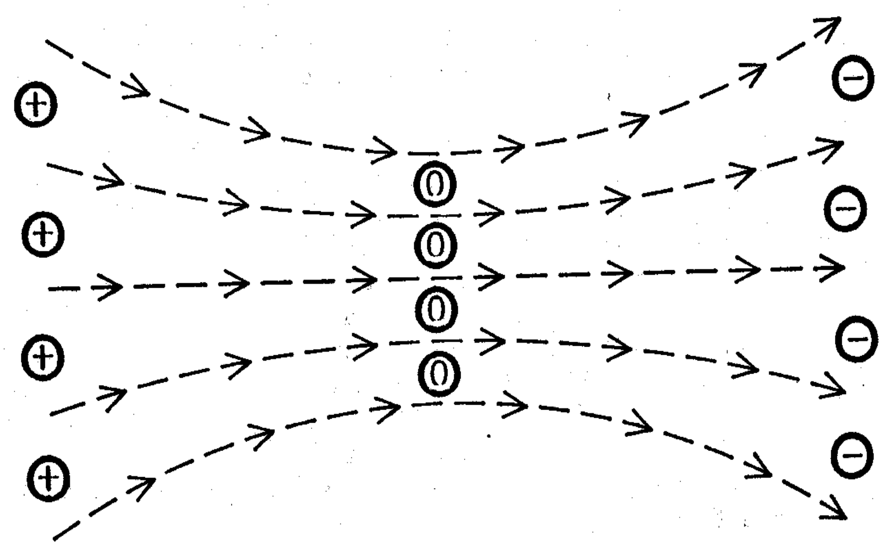

Figure 1.1 schematically shows the divergence of a vector field. In the region where the arrows of the vector field converge, the divergence is positive, implying an increase in the source of the vector field. The opposite is true for the region where the field vectors diverge.

This single operator, when combined with the some of the products defined above, constitutes the foundation of vector calculus. Thus the divergence, gradient, and curl are defined as

$$
\left.\begin{aligned}
\nabla \bullet \vec{A} &= \beta \\
\nabla\alpha &= \vec{B} \\
\nabla \times \vec{A} &= \vec{C}
\end{aligned}\right\} \qquad (1.4.2)
$$

respectively. If we consider $\vec{A}$ to be a continuous vector function of the independent variables that make up the space in which it is defined, then we may give a physical interpretation for both the divergence and curl. The divergence of a vector field is a measure of the amount that the field spreads or contracts at some given point in the space (see Figure 1.1).

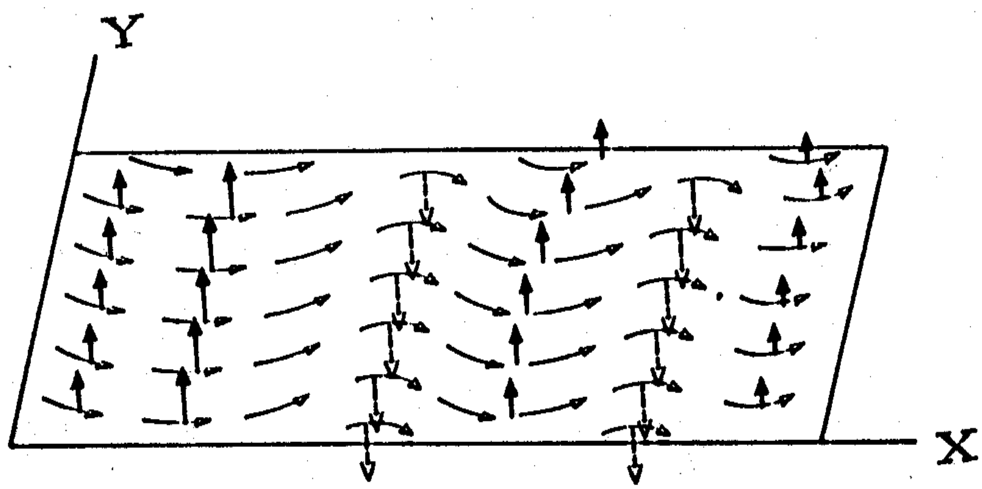

Figure 1.2 schematically shows the curl of a vector field. The direction of the curl is determined by the "right hand rule" while the magnitude depends on the rate of change of the x- and y-components of the vector field with respect to y and x.

The curl is somewhat harder to visualize. In some sense it represents the amount that the field rotates about a given point. Some have called it a measure of the "swirliness" of the field. If in the vicinity of some point in the field, the vectors tend to veer to the left rather than to the right, then the curl will be a vector pointing up normal to the net rotation with a magnitude that measures the degree of rotation (see Figure 1.2). Finally, the gradient of a scalar field is simply a measure of the direction and magnitude of the maximum rate of change of that scalar field (see Figure 1.3).

With these simple pictures in mind it is possible to generalize the notion of the Del-operator to other quantities. Consider the gradient of a vector field. This represents the outer product of the Del-operator with a vector. While one doesn't see such a thing often in freshman physics, it does occur in more advanced descriptions of fluid mechanics (and many other places). We now know enough to understand that the result of this operation will be a tensor of rank two which we can represent as a matrix.

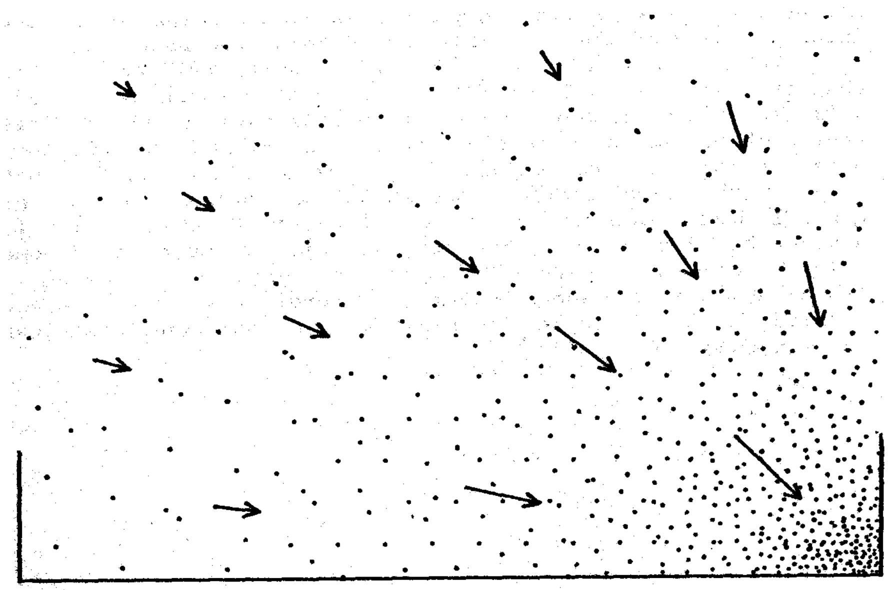

Figure 1.3 schematically shows the gradient of the scalar dot-density in the form of a number of vectors at randomly chosen points in the scalar field. The direction of the gradient points in the direction of maximum increase of the dot-density, while the magnitude of the vector indicates the rate of change of that density.

What do the components mean? Generalize from the scalar case. The nine elements of the vector gradient can be viewed as three vectors denoting the direction of the maximum rate of change of *each* of the components of the original vector. The nine elements represent a perfectly well defined quantity and it has a useful purpose in describing many physical situations. One can also consider the divergence of a second rank tensor, which is clearly a vector. In hydrodynamics, the divergence of the pressure tensor may reduce to the gradient of the scalar gas pressure if the macroscopic flow of the material is small compared to the internal speed of the particles that make up the material.

Thus by combining the various products defined in this chapter with the familiar notions of vector calculus, we can formulate a much richer description of the physical world. This review of scalar and vector mathematics along with the all-too-brief introduction to tensors and matrices will be useful, not only in the development of celestial mechanics, but in the general description of the physical world. However, there is another broad area of mathematics on which we must spend some time. To describe events in the physical world, it is common to frame them within some system of coordinates. We will now consider some of these coordinates and the transformations between them.

Common Del-Operators

Cylindrical Coordinates

Orthogonal Line Elements

$dr, \quad rd\theta, dz$

Divergence

$$\nabla \bullet \vec{A} = \frac{1}{r}\frac{\partial(rA_r)}{\partial r} + \frac{1}{r}\frac{\partial A_\vartheta}{\partial \vartheta} + \frac{\partial A_z}{\partial z}$$

Components of the Gradient

$$(\nabla a)_r = \frac{\partial a}{\partial r}$$

$$(\nabla a)_\vartheta = \frac{1}{r}\frac{\partial a}{\partial \vartheta}$$

$$(\nabla a)_z = \frac{\partial a}{\partial z}$$

Components of the Curl

$$(\nabla \times \vec{A})_r = \frac{1}{r}\frac{\partial A_z}{\partial \vartheta} - \frac{\partial A_\vartheta}{\partial z}$$

$$(\nabla \times \vec{A})_\vartheta = \frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r}$$

$$(\nabla \times \vec{A})_z = \frac{1}{r}\left[\frac{\partial(rA_\vartheta)}{\partial r} - \frac{\partial A_r}{\partial \vartheta}\right]$$

Spherical Coordinates

Orthogonal Line Elements

$dr, \quad rd\theta \quad r\sin\theta\, d\phi$

Divergence

$$\nabla \bullet \vec{A} = \frac{1}{r^2}\frac{\partial(r^2 A_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(A_\theta \sin\theta)}{\partial \theta} + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial \phi}$$

Components of the Gradient

$$(\nabla a)_r = \frac{\partial a}{\partial r}$$

$$(\nabla a)_\theta = \frac{1}{r}\frac{\partial a}{\partial \theta}$$

$$(\nabla a)_\phi = \frac{1}{r\sin\theta}\frac{\partial a}{\partial \phi}$$

Components of the Curl

$$(\nabla \times \vec{A})_r = \frac{1}{r\sin\theta}\left[\frac{\partial(A_\phi \sin\theta)}{\partial \theta} - \frac{\partial A_\theta}{\partial \phi}\right]$$

$$(\nabla \times \vec{A})_\theta = \frac{1}{r\sin\theta}\frac{\partial A_r}{\partial \phi} - \frac{1}{r}\frac{\partial(rA_\phi)}{\partial r}$$

$$(\nabla \times \vec{A})_\phi = \frac{1}{r}\left[\frac{\partial(rA_\theta)}{\partial r} - \frac{\partial A_r}{\partial \theta}\right]$$

Chapter 1: Exercises

1. Find the components of the vector $\vec{A} = \nabla \times (\nabla \times \vec{B})$ in spherical coordinates.

2. Show that:

   a. $\nabla \times (\vec{A} \times \vec{B}) = (\vec{B} \bullet \nabla)\vec{A} - \vec{B}(\nabla \bullet \vec{A}) - (\vec{A} \bullet \nabla)\vec{B} + \vec{A}(\nabla \bullet \vec{B})$.

   b. $\nabla \bullet (\vec{A} \times \vec{B}) = \vec{B} \bullet (\nabla \times \vec{A}) - \vec{A} \bullet (\nabla \times \vec{B})$.

3. Show that:

   $\nabla(\vec{A} \bullet \vec{B}) = (\vec{B} \bullet \nabla)\vec{A} + (\vec{A} \bullet \nabla)\vec{B} + \vec{B} \times (\nabla \times \vec{A}) + \vec{A} \times (\nabla \times \vec{B})$.

4. If **T** is a tensor of rank 2 with components T_{ij}, show that $\nabla \bullet \mathbf{T}$ is a vector and find the components of that vector.

Useful Vector Identities

$$\nabla \bullet (a\vec{A}) = a\nabla \bullet \vec{A} + \vec{A} \bullet \nabla a \quad . \tag{a1}$$

$$\nabla \times (a\vec{A}) = a(\nabla \times \vec{A}) + \vec{A} \times \nabla a \, . \tag{a2}$$

$$\nabla \times (\nabla \times \vec{A}) = \nabla(\nabla \bullet \vec{A}) - \nabla \bullet (\nabla \vec{A}) \tag{a3}$$

$$\nabla \times (\vec{A} \times \vec{B}) = (\vec{B} \bullet \nabla)\vec{A} - \vec{B}(\nabla \bullet \vec{A}) - (\vec{A} \bullet \nabla)\vec{B} + \vec{A}(\nabla \bullet \vec{B}) \quad . \tag{a4}$$

$$\nabla \bullet (\vec{A} \times \vec{B}) = \vec{B} \bullet (\nabla \times \vec{A}) - \vec{A} \bullet (\nabla \times \vec{B}) \quad . \tag{a5}$$

$$\nabla(\vec{A} \bullet \vec{B}) = (\vec{B} \bullet \nabla)\vec{A} + (\vec{A} \bullet \nabla)\vec{B} + \vec{B} \times (\nabla \times \vec{A}) + \vec{A} \times (\nabla \times \vec{B}) \tag{a6}$$

$$\nabla \bullet (\nabla a) \equiv \nabla^2 a = \text{Laplacian of } a \quad . \tag{a7}$$

In Cartesian coordinates:

$$(\vec{A} \bullet \nabla)\vec{B} = \hat{i}\left(A_x \frac{\partial B_x}{\partial x} + A_y \frac{\partial B_x}{\partial y} + A_z \frac{\partial B_x}{\partial z} \right) + \hat{j}\left(A_x \frac{\partial B_y}{\partial x} + A_y \frac{\partial B_y}{\partial y} + A_z \frac{\partial B_y}{\partial z} \right)$$
$$+ \hat{k}\left(A_x \frac{\partial B_z}{\partial x} + A_y \frac{\partial B_z}{\partial y} + A_z \frac{\partial B_z}{\partial z} \right) \tag{a8}$$