1

Introduction and Fundamental Concepts

• • •

The numerical expression of a scientific statement has traditionally been the manner by which scientists have verified a theoretical description of the physical world. During this century there has been a revolution in both the nature and extent to which this numerical comparison can be made. Indeed, it seems likely that when the history of this century is definitively written, it will be the development of the computer, which will be regarded as its greatest technological achievement - not nuclear power. While it is true that the origins of the digital computer can be traced through the work of Isaac Babbitt, Hermann Hollerith, and others in the nineteenth century, the real advance came after the Second World War when machines were developed that were able to carry out an extended sequence of instructions at a rate that was very much greater than a human could manage. We call such machines programmable.

The electronic digital computer of the sort developed by John von Neumann and others in the 1950s really ushered in the present computer revolution. While it is still to soon to delineate the form and consequences of this revolution, it is already clear that it has forever changed the way in which science and engineering will be done. The entire approach to numerical analysis has changed in the past two decades and that change will most certainly continue rapidly into the future. Prior to the advent of the electronic digital computer, the emphasis in computing was on short cuts and methods of verification which insured that computational errors could be caught before they propagated through the solution. Little attention was paid to "round off error" since the "human computer" could easily control such problems when they were encountered. Now the reliability of electronic machines has nearly eliminated concerns of random error, but round off error can be a persistent problem.

The extreme speed of contemporary machines has tremendously expanded the scope of numerical problems that may be considered as well as the manner in which such computational problems may even be approached. However, this expansion of the degree and type of problem that may be numerically solved has removed the scientist from the details of the computation. For this, most would shout "Hooray"! But this removal of the investigator from the details of computation may permit the propagation of errors of various types to intrude and remain undetected. Modern computers will almost always produce numbers, but whether they represent the solution to the problem or the result of error propagation may not be obvious. This situation is made worse by the presence of programs designed for the solution of broad classes of problems. Almost every class of problems has its pathological example for which the standard techniques will fail. Generally little attention is paid to the recognition of these pathological cases which have an uncomfortable habit of turning up when they are least expected.

Thus the contemporary scientist or engineer should be skeptical of the answers presented by the modern computer unless he or she is completely familiar with the numerical methods employed in obtaining that solution. In addition, the solution should always be subjected to various tests for "reasonableness". There is often a tendency to regard the computer and the programs which they run as "black boxes" from which come infallible answers. Such an attitude can lead to catastrophic results and belies the attitude of "healthy skepticism" that should pervade all science. It is necessary to understand, at least at some level, what the "Black Boxes" do. That understanding is one of the primary aims of this book.

It is not my intention to teach the techniques of programming a computer. There are many excellent texts on the multitudinous languages that exist for communicating with a computer. I will assume that the reader has sufficient capability in this area to at least conceptualize the manner by which certain processes could be communicated to the computer or at least recognize a computer program that does so. However, the programming of a computer does represent a concept that is not found in most scientific or mathematical presentations. We will call that concept an algorithm. An algorithm is simply a sequence of mathematical operations which, when preformed in sequence, lead to the numerical answer to some specified problem. Much time and effort is devoted to ascertaining the conditions under which a particular algorithm will work. In general, we will omit the proof and give only the results when they are known. The use of algorithms and the ability of computers to carry out vastly more operations in a short interval of time than the human programmer could do in several lifetimes leads to some unsettling differences between numerical analysis and other branches of mathematics and science.

Much as the scientist may be unwilling to admit it, some aspects of art creep into numerical analysis. Knowing when a particular algorithm will produce correct answers to a given problem often involves a non-trivial amount of experience as well as a broad based knowledge of machines and computational procedures. The student will achieve some feeling for this aspect of numerical analysis by considering problems for which a given algorithm should work, but doesn't. In addition, we shall give some "rules of thumb" which indicate when a particular numerical method is failing. Such "rules of thumb" are not guarantees of either success or failure of a specific procedure, but represent instances when a greater height of skepticism on the part of the investigator may be warranted.

As already indicated, a broad base of experience is useful when trying to ascertain the validity of the results of any computer program. In addition, when trying to understand the utility of any algorithm for

calculation, it is useful to have as broad a range of mathematical knowledge as possible. Mathematics is indeed the language of science and the more proficient one is in the language the better. So a student should realize as soon as possible that there is essentially one subject called mathematics, which for reasons of convenience we break down into specific areas such as arithmetic, algebra, calculus, tensors, group theory, etc. The more areas that the scientist is familiar with, the more he/she may see the relations between them. The more the relations are apparent, the more useful mathematics will be. Indeed, it is all too common for the modern scientist to flee to a computer for an answer. I cannot emphasize too strongly the need to analyze a problem thoroughly before any numerical solution is attempted. Very often a better numerical approach will suggest itself during the analyses and occasionally one may find that the answer has a closed form analytic solution and a numerical solution is unnecessary.

However, it is too easy to say "I don't have the background for this subject" and thereby never attempt to learn it. The complete study of mathematics is too vast for anyone to acquire in his or her lifetime. Scientists simply develop a base and then continue to add to it for the rest of their professional lives. To be a successful scientist one cannot know too much mathematics. In that spirit, we shall "review" some mathematical concepts that are useful to understanding numerical methods and analysis. The word review should be taken to mean a superficial summary of the area mainly done to indicate the relation to other areas. Virtually every area mentioned has itself been a subject for many books and has occupied the study of some investigators for a lifetime. This short treatment should not be construed in any sense as being complete. Some of this material will indeed be viewed as elementary and if thoroughly understood may be skimmed. However many will find some of these concepts as being far from elementary. Nevertheless they will sooner or later be useful in understanding numerical methods and providing a basis for the knowledge that mathematics is "all of a piece".

1.1 Basic Properties of Sets and Groups

Most students are introduced to the notion of a set very early in their educational experience. However, the concept is often presented in a vacuum without showing its relation to any other area of mathematics and thus it is promptly forgotten. Basically *a set is a collection of elements*. The notion of an element is left deliberately vague so that it may represent anything from cows to the real numbers. The number of elements in the set is also left unspecified and may or may not be finite. Just over a century ago Georg Cantor basically founded set theory and in doing so clarified our notion of infinity by showing that there are different types of infinite sets. He did this by generalizing what we mean when we say that two sets have the same number of elements. Certainly if we can identify each element in one set with a unique element in the second set and there are none left over when the identification is completed, then we would be entitled in saying that the two sets had the same number of elements. Cantor did this formally with the infinite set composed of the positive integers and the infinite set of the real numbers. He showed that it is not possible to identify each real number with a integer so that there are more real numbers than integers and thus different degrees of infinity which he called cardinality. He used the first letter of the Hebrew alphabet to denote the cardinality of an infinite set so that the integers had cardinality $\aleph_0$ and the set of real numbers had cardinality of $\aleph_1$. Some of the brightest minds of the twentieth century have been concerned with the properties of infinite sets.

Our main interest will center on those sets which have constraints placed on their elements for it will be possible to make some very general statements about these restricted sets. For example, consider a set wherein the elements are related by some "law". Let us denote the "law" by the symbol ‡. If two elements are combined under the "law" so as to yield another element in the set, the set is said to be closed with respect to that law. Thus if a, b, and c are elements of the set and

$$a‡b = c \ , \tag{1.1.1}$$

then the set is said to be closed with respect to ‡. We generally consider ‡ to be some operation like + or $\times$, but we shouldn't feel that the concept is limited to such arithmetic operations alone. Indeed, one might consider operations such as b 'follows' a to be an example of a law operating on a and b.

If we place some additional conditions of the elements of the set, we can create a somewhat more restricted collection of elements called a group. Let us suppose that one of the elements of the set is what we call a unit element. Such an element is one which, when combined with any other element of the set under the law, produces that same element. Thus

$$a‡i = a \ . \tag{1.1.2}$$

This suggests another useful constraint, namely that there are elements in the set that can be designated "*inverses*". An inverse of an element is one that when combined with its element under the law produces the unit element i or

$$a^{-1}‡a = i \ . \tag{1.1.3}$$

Now with one further restriction on the law itself, we will have all the conditions required to produce a group. The restriction is known as *associativity*. A law is said to be associative if the order in which it is applied to three elements does not determine the outcome of the application. Thus

$$(a‡b)‡c = a‡(b‡c) \ . \tag{1.1.4}$$

If a set possess a unit element and inverse elements and is closed under an associative law, that set is called a group under the law. Therefore the normal integers form a group under addition. The unit is zero and the inverse operation is clearly subtraction and certainly the addition of any two integers produces another integer. The law of addition is also associative. However, it is worth noting that the integers do not form a group under multiplication as the inverse operation (reciprocal) does not produce a member of the group (an integer). One might think that these very simple constraints would not be sufficient to tell us much that is new about the set, but the notion of a group is so powerful that an entire area of mathematics known as group theory has developed. It is said that Eugene Wigner once described all of the essential aspects of the thermodynamics of heat transfer on one sheet of paper using the results of group theory.

While the restrictions that enable the elements of a set to form a group are useful, they are not the only restrictions that frequently apply. The notion of commutivity is certainly present for the laws of addition and scalar multiplication and, if present, may enable us to say even more about the properties of our set. A law is said to be *communitative* if

$$a‡b = b‡a \ . \tag{1.1.5}$$

A further restriction that may be applied involves two laws say ‡ and $\wedge$. These laws are said to be distributive with respect to one another if

$$a‡(b\wedge c) = (a‡b)\wedge(a‡c) \ . \tag{1.1.6}$$

Although the laws of addition and scalar multiplication satisfy all three restrictions, we will encounter common laws in the next section that do not. Subsets that form a group under addition and scalar multiplication are called *fields*. The notion of a field is very useful in science as most theoretical descriptions of the physical world are made in terms of fields. One talks of gravitational, electric, and magnetic fields in physics. Here one is describing scalars and vectors whose elements are real numbers and for which there are laws of addition and multiplication which cause these quantities to form not just groups, but fields. Thus all the abstract mathematical knowledge of groups and fields is available to the scientist to aid in understanding physical fields.

1.2 Scalars, Vectors, and Matrices

In the last section we mentioned specific sets of elements called scalars and vectors without being too specific about what they are. In this section we will define the elements of these sets and the various laws that operate on them. In the sciences it is common to describe phenomena in terms of specific quantities which may take on numerical values from time to time. For example, we may describe the atmosphere of the planet at any point in terms of the temperature, pressure, humidity, ozone content or perhaps a pollution index. Each of these items has a single value at any instant and location and we would call them scalars. The common laws of arithmetic that operate on scalars are addition and multiplication. As long as one is a little careful not to allow division by zero (often known as the cancellation law) such scalars form not only groups, but also fields.

Although one can generally describe the condition of the atmosphere locally in terms of scalar fields, the location itself requires more than a single scalar for its specification. Now we need two (three if we include altitude) numbers, say the latitude and longitude, which locate that part of the atmosphere for further description by scalar fields. A quantity that requires more than one number for its specification may be called a vector. Indeed, some have defined a vector as an "*ordered n-tuple of numbers*". While many may not find this too helpful, it is essentially a correct statement, which emphasizes the multi-component side of the notion of a vector. The number of components that are required for the vector's specification is usually called the *dimensionality of the vector*. We most commonly think of vectors in terms of spatial vectors, that is, vectors that locate things in some coordinate system. However, as suggested in the previous section, vectors may represent such things as an electric or magnetic field where the quantity not only has a magnitude or scalar length associated with it at every point in space, but also has a direction. As long as such quantities obey laws of addition and some sort of multiplication, they may indeed be said to form vector fields. Indeed, there are various types of products that are associated with vectors. The most common of these and the one used to establish the field nature of most physical *vector fields* is called the "*scalar product*" or inner product, or sometimes simply the dot product from the manner in which it is usually written. Here the result is a scalar and we can operationally define what we mean by such a product by

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

One might say that as the result of the operation is a scalar not a vector, but that would be to put to restrictive an interpretation on what we mean by a vector. Specifically, any scalar can be viewed as vector having only one component (i.e. a 1-dimensional vector). Thus scalars become a subgroup of vectors and since the vector scalar product degenerates to the ordinary scalar product for 1-dimensional vectors, they are actually a sub-field of the more general notion of a vector field.

It is possible to place additional constraints (laws) on a field without destroying the field nature of the elements. We most certainly do this with vectors. Thus we can define an additional type of product known as the "vector product" or simply cross product again from the way it is commonly written. Thus in Cartesian coordinates the cross product can be written as

$$\vec{A}\times\vec{B}=\begin{vmatrix}\hat{i} & \hat{j} & \hat{k}\\ A_i & A_j & A_k\\ B_i & B_j & B_k\end{vmatrix}=\hat{i}(A_jB_k-A_kB_j)-\hat{j}(A_iB_k-A_kB_i)+\hat{k}(A_iB_j-A_jB_i)\,. \qquad (1.2.2)$$

The result of this operation is a vector, but we shall see later that it will be useful to sharpen our definition of vectors so that this result is a special kind of vector.

Finally, there is the "tensor product" or vector outer product that is defined as

$$\left.\begin{aligned}&\vec{A}\vec{B}=\mathbf{C}\\ &C_{ij}=A_iB_j\end{aligned}\right\}\,. \qquad (1.2.3)$$

Here the result of applying the "law" is an ordered array of ($n\times m$) numbers where n and m are the dimensions of the vectors $\vec{A}$ and $\vec{B}$ respectively. Again, here the result of applying the law is not a vector in any sense of the normal definition, but is a member of a larger class of objects we will call tensors. But before discussing tensors in general, let us consider a special class of them known as matrices.

The result of equation (1.2.3) while needing more than one component for its specification is clearly not simply a vector with dimension ($n\times m$). The values of n and m are separately specified and to specify only the product would be to throw away information that was initially specified. Thus, in order to keep this information, we can represent the result as an array of numbers having n columns and m rows. Such an array can be called a matrix. For matrices, the products already defined have no simple interpretation. However, there is an additional product known as a *matrix product*, which will allow us to at least define a matrix group. Consider the product defined by

$$\left.\begin{aligned}&\mathbf{AB}=\mathbf{C}\\ &C_{ij}=\sum_k A_{ik}B_{kj}\end{aligned}\right\}\,. \qquad (1.2.4)$$

With this definition of a product, the unit matrix denoted by **1** will have elements δ_{ij} specified for n = m = 2 by

$$\delta_{ij}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\,. \qquad (1.2.5)$$

The quantity δ_{ij} is called the *Kronecker delta* and may be generalized to n-dimensions.

Thus the inverse elements of the group will have to satisfy the relation

$$\mathbf{A}\mathbf{A}^{-1}=\mathbf{1}\,, \qquad (1.2.6)$$

and we shall spend some time in the next chapter discussing how these members of the group may be calculated. Since matrix addition can simply be defined as the scalar addition of the elements of the matrix, and the 'unit' matrix under addition is simply a matrix with zero elements, it is tempting to think that the group of matrices also form a field. However, the matrix product as defined by equation (1.2.4), while being distributive with respect to addition, is not communitative. Thus we shall have to be content with matrices forming a group under both addition and matrix multiplication but not a field.

There is much more that can be said about matrices as was the case with other subjects of this chapter, but we will limit ourselves to a few properties of matrices which will be particularly useful later. For example, the transpose of a matrix with elements $\mathbf{A}_{ij}$ is defined as

$$\mathbf{A}^{T} = A_{ji} \quad . \tag{1.2.7}$$

We shall see that there is an important class of matrices (i.e. the orthonormal matrices) whose inverse is their transpose. This makes the calculation of the inverse trivial.

Another important scalar quantity is the *trace of a matrix* defined as

$$\mathrm{Tr}\mathbf{A} = \sum_{i} A_{ii} \quad . \tag{1.2.8}$$

A matrix is said to be *symmetric* if $A_{i\,j} = A_{ji}$. If, in addition, the elements are themselves complex numbers, then should the elements of the transpose be the complex conjugates of the original matrix, the matrix is said to be *Hermitian* or *self-adjoint*. The conjugate transpose of a matrix $\mathbf{A}$ is usually denoted by $\mathbf{A}^{\dagger}$. If the Hermitian conjugate of $\mathbf{A}$ is also $\mathbf{A}^{-1}$, then the matrix is said to be *unitary*. Should the matrix $\mathbf{A}$ commute with it Hermitian conjugate so that

$$\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{A}^{\dagger}\mathbf{A} \quad , \tag{1.2.9}$$

then the matrix is said to be *normal*. For matrices with only real elements, Hermitian is the same as symmetric, unitary means the same as orthonormal and both classes would be considered to be normal.

Finally, a most important characteristic of a matrix is its *determinant*. It may be calculated by expansion of the matrix by "minors" so that

$$\det A = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{13}) . \tag{1.2.10}$$

Fortunately there are more straightforward ways of calculating the determinant which we will consider in the next chapter. There are several theorems concerning determinants that are useful for the manipulation of determinants and which we will give without proof.

1. *If each element in a row or column of a matrix is zero, the determinant of the matrix is zero.*

2. *If each element in a row or column of a matrix is multiplied by a scalar q, the determinant is multiplied by q.*

3. *If each element of a row or column is a sum of two terms, the determinant equals the sum of the two corresponding determinants.*

4. *If two rows or two columns are proportional, the determinant is zero. This clearly follows from theorems 1, 2 and 3.*

5. *If two rows or two columns are interchanged, the determinant changes sign.*

6. *If rows and columns of a matrix are interchanged, the determinant of the matrix is unchanged.*

7. *The value of a determinant of a matrix is unchanged if a multiple of one row or column is added to another.*

8. *The determinant of the product of two matrices is the product of the determinants of the two matrices.*

One of the important aspects of the determinant is that it is a single parameter that can be used to characterize the matrix. Any such single parameter (i.e. the sum of the absolute value of the elements) can be so used and is often called a matrix norm. We shall see that various matrix norms are useful in determining which numerical procedures will be useful in operating on the matrix. Let us now consider a broader class of objects that include scalars, vectors, and to some extent matrices.

1.3 Coordinate Systems and Coordinate Transformations

There is an area of mathematics known as *topology,* which deals with the description of spaces. To most students the notion of a space is intuitively obvious and is restricted to the three dimensional Euclidian space of every day experience. A little reflection might persuade that student to include the flat plane as an allowed space. However, a little further generalization would suggest that any time one has several independent variables that they could be used to form a space for the description of some phenomena. In the area of topology the notion of a space is far more general than that and many of the more exotic spaces have no known counterpart in the physical world.

We shall restrict ourselves to spaces of independent variables, which generally have some physical interpretation. These variables can be said to constitute a coordinate frame, which describes the space and are fairly high up in the hierarchy of spaces catalogued by topology. To understand what is meant by a coordinate frame, imagine a set of rigid rods or vectors all connected at a point. We shall call such a collection of rods a reference frame. If every point in space can be projected onto the rods so that a unique set of rod-points represent the space point, the vectors are said to span the space.

If the vectors that define the space are locally perpendicular, they are said to form an orthogonal coordinate frame. If the vectors defining the reference frame are also unit vectors say $\hat{e}_i$ then the condition for orthogonality can be written as

$$\hat{e}_i \bullet \hat{e}_j = \delta_{ij} \quad , \tag{1.3.1}$$

where δ_{ij} is the Kronecker delta. Such a set of vectors will span a space of dimensionality equal to the number of vectors $\hat{e}_j$. Such a space need not be Euclidian, but if it is then the coordinate frame is said to be a *Cartesian coordinate frame*. The conventional xyz-coordinate frame is Cartesian, but one could imagine such a coordinate system drawn on a rubber sheet, and then distorted so that locally the orthogonality conditions are still met, but the space would no longer be Euclidian or Cartesian.

Of the orthogonal coordinate systems, there are several that are particularly useful for the description of the physical world. Certainly the most common is the rectangular or Cartesian coordinate frame where coordinates are often denoted by x, y, z or x_1, x_2, x_3. Other common three dimensional frames include spherical polar coordinates (r,θ,φ) and cylindrical coordinates (ρ,ϑ,z). Often the most important part of solving a numerical problem is choosing the proper coordinate system to describe the problem. For example, there are a total of thirteen orthogonal coordinate frames in which Laplace's equation is separable (see Morse and Feshbach[1]).

In order for coordinate frames to be really useful it is necessary to know how to get from one to another. That is, if we have a problem described in one coordinate frame, how do we express that same problem in another coordinate frame? For quantities that describe the physical world, we wish their meaning to be independent of the coordinate frame that we happen to choose. Therefore we should expect the process to have little to do with the problem, but rather involve relationships between the coordinate frames themselves. These relationships are called *coordinate transformations*. While there are many such transformations in mathematics, for the purposes of this summary we shall concern ourselves with *linear transformations*. Such coordinate transformations relate the coordinates in one frame to those in a second frame by means of a system of linear algebraic equations. Thus if a vector $\vec{x}$ in one coordinate system has components x_j, in a primed-coordinate system a vector $\vec{x}'$ to the same point will have components x'_j

$$x_i = \sum_j A_{ij} x_j + B_i \,. \tag{1.3.2}$$

In vector notation we could write this as

$$\vec{x}' = \mathbf{A}\vec{x} + \vec{B} \quad . \tag{1.3.3}$$

This defines the general class of linear transformation where **A** is some matrix and $\vec{B}$ is a vector. This general linear form may be divided into two constituents, the matrix **A** and the vector. It is clear that the vector $\vec{B}$ may be interpreted as a shift in the origin of the coordinate system, while the elements A_{ij} are the cosines of the angles between the axes X_i and X'_i, and are called the directions cosines (see Figure 1.1). Indeed, the vector $\vec{B}$ is nothing more than a vector from the origin of the un-primed coordinate frame to the origin of the primed coordinate frame. Now if we consider two points that are fixed in space and a vector connecting them, then the length and orientation of that vector will be independent of the origin of the coordinate frame in which the measurements are made. That places an additional constraint on the types of linear transformations that we may consider. For instance, transformations that scaled each coordinate by a constant amount, while linear, would change the length of the vector as measured in the two coordinate systems. Since we are only using the coordinate system as a convenient way to describe the vector, the coordinate system can play no role in controlling the length of the vector. Thus we shall restrict our investigations of linear transformations to those that transform orthogonal coordinate systems while preserving the length of the vector.

Thus the matrix **A** must satisfy the following condition

$$\vec{x}' \bullet \vec{x}' = (\mathbf{A}\vec{x}) \bullet (\mathbf{A}\vec{x}) = \vec{x} \bullet \vec{x} \quad , \tag{1.3.4}$$

which in component form becomes

$$\sum_i \left(\sum_j A_{ij} x_j \right) \left(\sum_k A_{ik} x_k \right) = \sum_j \sum \left(\sum_i A_{ij} A_{ik} x_j x_k \right) = \sum_i x_i^2 \quad . \tag{1.3.5}$$

This must be true for all vectors in the coordinate system so that

$$\sum_i A_{ij} A_{ik} = \delta_{jk} = \sum_i A_{ji}^{-1} A_{ik} \quad . \tag{1.3.6}$$

Now remember that the Kronecker delta δ_{ij} is the unit matrix and any element of a group that multiplies another and produces that group's unit element is defined as the inverse of that element. Therefore

$$A_{ji} = [A_{ij}]^{-1} \quad . \tag{1.3.7}$$

Interchanging the rows with the columns of a matrix produces a new matrix which we have called the transpose of the matrix. Thus orthogonal transformations that preserve the length of vectors have inverses that are simply the transpose of the original matrix so that

$$\mathbf{A}^{-1} = \mathbf{A}^{T} \quad . \tag{1.3.8}$$

This means that given the transformation **A** in the linear system of equations (1.3.3), we may invert the transformation, or solve the linear equations, by multiplying those equations by the transpose of the original matrix or

$$\vec{x} = \mathbf{A}^{T} \vec{x}' - \mathbf{A}^{T} \vec{B} \quad . \tag{1.3.9}$$

Such transformations are called orthogonal unitary transformations, or orthonormal transformations, and the result given in equation (1.3.9) greatly simplifies the process of carrying out a transformation from one coordinate system to another and back again.

We can further divide orthonormal transformations into two categories. These are most easily described by visualizing the relative orientation between the two coordinate systems. Consider a transformation that carries one coordinate into the negative of its counterpart in the new coordinate system while leaving the others unchanged. If the changed coordinate is, say, the x-coordinate, the transformation matrix would be

$$\mathbf{A} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad , \tag{1.3.10}$$

which is equivalent to viewing the first coordinate system in a mirror. Such transformations are known as reflection transformations and will take a right handed coordinate system into a left handed coordinate system.

The length of any vectors will remain unchanged. The x-component of these vectors will simply be replaced by its negative in the new coordinate system. However, this will not be true of "vectors" that result from the vector cross product. The values of the components of such a vector will remain unchanged implying that a reflection transformation of such a vector will result in the orientation of that vector being changed. If you will, this is the origin of the "right hand rule" for vector cross products. A left hand rule

results in a vector pointing in the opposite direction. Thus such vectors are not invariant to *reflection transformations* because their orientation changes and this is the reason for putting them in a separate class, namely the axial (pseudo) vectors. It is worth noting that an orthonormal reflection transformation will have a determinant of -1. The unitary magnitude of the determinant is a result of the magnitude of the vector being unchanged by the transformation, while the sign shows that some combination of coordinates has undergone a reflection.

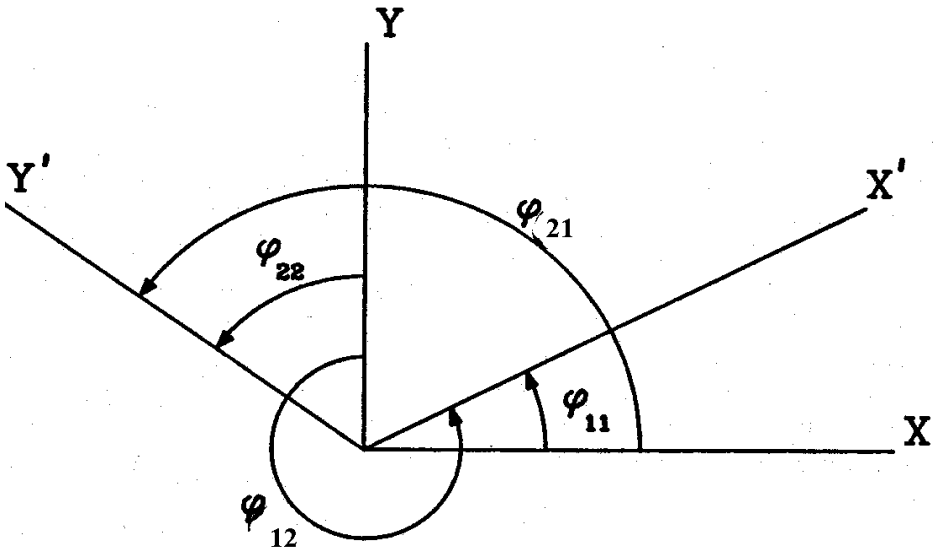


Figure 1.1 shows two coordinate frames related by the transformation angles φ_{ij}*. Four coordinates are necessary if the frames are not orthogonal*

As one might expect, the elements of the second class of orthonormal transformations have determinants of +1. These represent transformations that can be viewed as a rotation of the coordinate system about some axis. Consider a transformation between the two coordinate systems displayed in Figure 1.1. The components of any vector $\vec{C}$ in the primed coordinate system will be given by

$$\begin{pmatrix} C_{x'} \\ C_{y'} \\ C_{z'} \end{pmatrix} = \begin{pmatrix} \cos\varphi_{11} & \cos\varphi_{12} & 0 \\ \cos\varphi_{21} & \cos\varphi_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C_x \\ C_y \\ C_z \end{pmatrix} . \tag{1.3.11}$$

If we require the transformation to be orthonormal, then the direction cosines of the transformation will not be linearly independent since the angles between the axes must be $\pi/2$ in both coordinate systems. Thus the angles must be related by

$$\left.\begin{aligned} &\varphi_{11} = \varphi_{22} = \varphi \\ &\varphi_{12} = \varphi_{11} + \pi/2 = \varphi + \pi/2 \\ &(2\pi - \varphi_{21}) = \pi/2 - \varphi_{11} = \pi/2 - \varphi \end{aligned}\right\} . \tag{1.3.12}$$

Using the addition identities for trigonometric functions, equation (1.3.11) can be given in terms of the single

angle φ by

$$\begin{pmatrix} C_{x'} \\ C_{y'} \\ C_{z'} \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C_x \\ C_y \\ C_z \end{pmatrix} . \tag{1.3.13}$$

This transformation can be viewed as a simple rotation of the coordinate system about the Z-axis through an angle φ. Thus,

$$\mathrm{Det} \begin{vmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{vmatrix} = \cos^2\varphi + \sin^2\varphi = +1 . \tag{1.3.14}$$

In general, the rotation of any Cartesian coordinate system about one of its principal axes can be written in terms of a matrix whose elements can be expressed in terms of the rotation angle. Since these transformations are about one of the coordinate axes, the components along that axis remain unchanged. The rotation matrices for each of the three axes are

$$\left.\begin{aligned} P_x(\phi) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{pmatrix} \\ P_y(\phi) &= \begin{pmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{pmatrix} \\ P_z(\phi) &= \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}\right\} . \tag{1.3.15}$$

It is relatively easy to remember the form of these matrices for the row and column of the matrix corresponding to the rotation axis always contains the elements of the unit matrix since that component is not affected by the transformation. The diagonal elements always contain the cosine of the rotation angle while the remaining off diagonal elements always contain the sine of the angle modulo a sign. For rotations about the x- or z-axes, the sign of the upper right off diagonal element is positive and the other negative. The situation is just reversed for rotations about the y-axis. So important are these rotation matrices that it is worth remembering their form so that they need not be re-derived every time they are needed.

One can show that it is possible to get from any given orthogonal coordinate system to another through a series of three successive coordinate rotations. Thus a general orthonormal transformation can always be written as the product of three coordinate rotations about the orthogonal axes of the coordinate systems. It is important to remember that the matrix product is not commutative so that the order of the

rotations is important.

1.4 Tensors and Transformations

Many students find the notion of tensors to be intimidating and therefore avoid them as much as possible. After all Einstein was once quoted as saying that there were not more than ten people in the world that would understand what he had done when he published General Theory of Relativity. Since tensors are the foundation of general relativity that must mean that they are so esoteric that only ten people could manage them. Wrong! This is a beautiful example of misinterpretation of a quote taken out of context. What Einstein meant was that the notation he used to express the General Theory of Relativity was sufficiently obscure that there were unlikely to be more than ten people who were familiar with it and could therefore understand what he had done. So unfortunately, tensors have generally been represented as being far more complex than they really are. Thus, while readers of this book may not have encountered them before, it is high time they did. Perhaps they will be somewhat less intimidated the next time, for if they have any ambition of really understanding science, they will have to come to an understanding of them sooner or later.

In general a tensor has N^n components or elements. N is known as the dimensionality of the tensor by analogy with vectors, while n is called the rank of the tensor. Thus scalars are tensors of rank zero and vectors of any dimension are rank one. So scalars and vectors are subsets of tensors. We can define the law of addition in the usual way by the addition of the tensor elements. Thus the null tensor (i.e. one whose elements are all zero) forms the unit under addition and arithmetic subtraction is the inverse operation. Clearly tensors form a communitative group under addition. Furthermore, the scalar or dot product can be generalized for tensors so that the result is a tensor of rank $|m-n|$. In a similar manner the outer product can be defined so that the result is a tensor of rank $|m+n|$. It is clear that all of these operations are closed; that is, the results remain tensors. However, while these products are in general distributive, they are not communitative and thus tensors will not form a field unless some additional restrictions are made.

One obvious way of representing tensors of rank 2 is as $N_{\times}N$ square matrices Thus, the scalar product of a tensor of rank 2 with a vector would be written as

$$\left.\begin{aligned} \mathbf{A}\bullet\vec{B} &= \vec{C} \\ C_i &= \sum_j A_{ij}B_j \end{aligned}\right\}, \tag{1.4.1}$$

while the tensor outer product of the same tensor and vector could be written as

$$\left.\begin{aligned} \mathbf{A}B &= \vec{\vec{C}} \\ C_{ijk} &= A_{ij}B_k \end{aligned}\right\}. \tag{1.4.2}$$

It is clear from the definition and specifically from equation (1.4.2) that tensors may frequently have a rank of more than two. However, it becomes more difficult to display all the elements in a simple geometrical fashion so they are generally just listed or described. A particularly important tensor of rank three is known as the *Levi-Civita Tensor* (or correctly the Levi-Civita Tensor Density). It plays a role that is somewhat complimentary to that of the Kronecker delta in that when any two indices are equal the tensor element is zero. When the indices are all different the tensor element is +1 or -1 depending on whether the index sequence can be obtained as an even or odd permutation from the sequence 1, 2, 3 respectively. If we try to represent the tensor ε_{ijk} as a succession of 3×3 matrices we would get

$$\left.\begin{aligned}
\varepsilon_{1jk} &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & +1 \\ 0 & -1 & 0 \end{pmatrix} \\
\varepsilon_{2jk} &= \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ +1 & 0 & 0 \end{pmatrix} \\
\varepsilon_{3jk} &= \begin{pmatrix} 0 & -1 & 0 \\ +1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}
\end{aligned}\right\} . \qquad (1.4.3)$$

This somewhat awkward looking third rank tensor allows us to write the equally awkward vector cross product in summation notation as

$$\vec{A}\times\vec{B} = \vec{\vec{\varepsilon}} : (\vec{A}\vec{B}) = \sum_j \sum_k \varepsilon_{ijk} A_j B_k = C_i \quad . \qquad (1.4.4)$$

Here the symbol **:** denotes the *double dot product* which is explicitly specified by the double sum of the right hand term. The quantity ε_{ijk} is sometimes called the permutation symbol as it changes sign with every permutation of its indices. This, and the identity

$$\sum_i \varepsilon_{ijk}\varepsilon_{ipq} = \delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp} \quad , \qquad (1.4.5)$$

makes the evaluation of some complicated vector identities much simpler (see exercise 13).

In section 1.3 we added a condition to what we meant by a vector, namely we required that the length of a vector be invariant to a coordinate transformation. Here we see the way in which additional constraints of what we mean by vectors can be specified by the way in which they transform. We further limited what we meant by a vector by noting that some vectors behave strangely under a reflection transformation and calling these pseudo-vectors. Since the Levi-Civita tensor generates the vector cross product from the elements of ordinary (polar) vectors, it must share this strange transformation property. Tensors that share this transformation property are, in general, known as tensor densities or pseudo-tensors. Therefore we should call ε_{ijk} defined in equation (1.4.3) the Levi-Civita tensor density. Indeed, it is the invariance of tensors, vectors, and scalars to orthonormal transformations that is most correctly used to define the elements of the group called tensors.

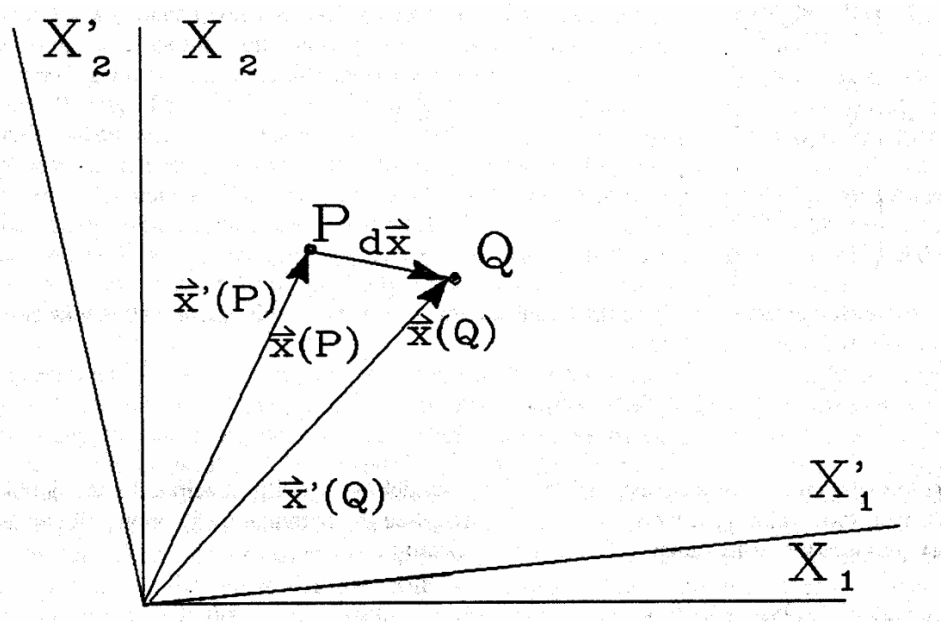


Figure 1.2 shows two neighboring points P *and* Q *in two adjacent coordinate systems* X *and* X'. *The differential distance between the two is* $d\vec{x}$. *The vectorial distance to the two points is* $\vec{X}$ *(P) or* $\vec{X}'$ *(P) and* $\vec{X}$(Q) *or* $\vec{X}'$(Q) *respectively.*

Since vectors are just a special case of the broader class of objects called tensors, we should expect these transformation constraints to extend to the general class. Indeed the only fully appropriate way to define tensors is to define the way in which they transform from one coordinate system to another. To further refine the notion of tensor transformation, we will look more closely at the way vectors transform. We have written a general linear transformation for vectors in equation (1.3.2). However, except for rotational and reflection transformations, we have said little about the nature of the transformation matrix **A**. So let us consider how we would express a coordinate transformation from some point P in a space to a nearby neighboring point Q. Each point can be represented in any coordinate system we choose. Therefore, let us consider two coordinate systems having a common origin where the coordinates are denoted by x_i and x'_i respectively.

Since P and Q are near each other, we can represent the coordinates of Q to those of P in either coordinate system by

$$\left.\begin{aligned} x_i(Q) &= x_i(P) + dx_i \\ x'_i(Q) &= x'_i(P) + dx'_i \end{aligned}\right\} . \qquad (1.4.6)$$

Now the coordinates of the vector from P to Q will be dx_i and dx'_i, in the un-primed and primed coordinate systems respectively. By the chain rule the two coordinates will be related by

$$dx'_i = \sum_j \frac{\partial x'_i}{\partial x_j} dx_j \quad . \tag{1.4.7}$$

Note that equation (1.4.7) does not involve the specific location of point Q but rather is a general expression of the local relationship between the two coordinate frames. Since equation (1.4.7) basically describes how the coordinates of P or Q will change from one coordinate system to another, we can identify the elements A_{ij} from equation (1.3.2) with the partial derivatives in equation (1.4.6). Thus we could expect any vector $\mathbf{x}$ to transform according to

$$x'_i = \sum_j \frac{\partial x'_i}{\partial x_j} x_j \quad . \tag{1.4.8}$$

Vectors that transform in this fashion are called *contravariant vectors*. In order to distinguish them from covariant vectors, which we shall shortly discuss, we shall denote the components of the vector with superscripts instead of subscripts. Thus the correct form for the transformation of a contravariant vector is

$$x'^i = \sum_j \frac{\partial x'^i}{\partial x^j} x^j \quad . \tag{1.4.9}$$

We can generalize this transformation law to contravariant tensors of rank two by

$$T'^{ij} = \sum_k \sum_l T^{kl} \frac{\partial x'^i}{\partial x^l} \frac{\partial x'^j}{\partial x^k} \quad . \tag{1.4.10}$$

Higher rank contravariant tensors transform as one would expect with additional coordinate changes. One might think that the use of superscripts to represent contravariant indices would be confused with exponents, but such is generally not the case and the distinction between this sort of vector transformation and covariance is sufficiently important in physical science to make the accommodation. The sorts of objects that transform in a contravariant manner are those associated with, but not limited to, *geometrical objects*. For example, the infinitesimal displacements of coordinates that makes up the tangent vector to a curve show that it is a contravariant vector. While we have used vectors to develop the notion of contravariance, it is clear that the concept can be extended to tensors of any rank including rank zero. The transformation rule for such a tensor would simply be

$$T' = T \quad . \tag{1.4.11}$$

In other words scalars will be invariant to contravariant coordinate transformations.

Now instead of considering vector representations of geometrical objects imbedded in the space and their transformations, let us consider a scalar function of the coordinates themselves. Let such a function be $\Phi(x^i)$. Now consider components of the gradient of Φ in the x'^i-coordinate frame. Again by the chain rule

$$\frac{\partial \Phi}{\partial x'_i} = \sum_j \frac{\partial x^j}{\partial x_i} \frac{\partial \Phi}{\partial x'^j} \quad . \tag{1.4.12}$$

If we call $\partial\Phi / \partial x'_i$ a vector with components V_i , then the transformation law given by equation (1.4.12) appears very like equation (1.4.8), but with the partial derivatives inverted. Thus we would identify the elements A_i^j of the linear vector transformation represented by equation (1.3.2) as

$$A_i^{\ j} = \partial x^j / \partial x'_i \quad , \tag{1.4.13}$$

and the vector transformation would have the form

$$V_i = \sum_j A_i{}^j V_j \quad . \tag{1.4.14}$$

Vectors that transform in this manner are called *covariant vectors*. In order to distinguish them from contravariant vectors, the component indices are written as subscripts. Again, it is not difficult to see how the concept of covariance would be extended to tensors of higher rank and specifically for a second rank covariant tensor we would have

$$T'_{ij} = \sum_k \sum_l T_{lk} \frac{\partial x^l}{\partial x'_i} \frac{\partial x^k}{\partial x'_j} \quad . \tag{1.4.15}$$

The use of the scalar invariant Φ to define what is meant by a covariant vector is a clue as to the types of vectors that behave as covariant vectors. Specifically the gradient of physical scalar quantities such as temperature and pressure would behave as a covariant vector while coordinate vectors themselves are contravariant. Basically equations (1.4.15) and (1.4.10) define what is meant by a covariant or contravariant tensor of second rank. It is possible to have a mixed tensor where one index represents covariant transformation while the other is contravariant so that

$$T'_i{}^j = \sum_k \sum_l T_l{}^k \frac{\partial x^l}{\partial x'_i} \frac{\partial x^j}{\partial x_k} \quad . \tag{1.4.16}$$

Indeed the Kronecker delta can be regarded as a tensor as it is a two index symbol and in particular it is a mixed tensor of rank two and when covariance and contravariance are important should be written as δ^i_j.

Remember that both contravariant and covariant transformations are locally linear transformations of the form given by equation (1.3.2). That is, they both preserve the length of vectors and leave scalars unchanged. The introduction of the terms contravariance and covariance simply generate two subgroups of what we earlier called tensors and defined the members of those groups by means of their detailed transformation properties. One can generally tell the difference between the two types of transformations by noting how the components depend on the coordinates. If the components denote 'distances' or depend directly on the coordinates, then they will transform as contravariant tensor components. However, should the components represent quantities that change with the coordinates such as gradients, divergences, and curls, then dimensionally the components will depend inversely on the coordinates and the will transform covariantly. The use of subscripts and superscripts to keep these transformation properties straight is particularly useful in the development of tensor calculus as it allows for the development of rules for the manipulation of tensors in accord with their specific transformation characteristics. While coordinate systems have been used to define the tensor characteristics, those characteristics are properties of the tensors themselves and do not depend on any specific coordinate frame. This is of considerable importance when developing theories of the physical world as anything that is fundamental about the universe should be independent of man made coordinate frames. This is not to say that the choice of coordinate frames is unimportant when actually solving a problem. Quite the reverse is true. Indeed, as the properties of the physical world represented by tensors are independent of coordinates and their explicit representation and transformation properties from one coordinate system to another are well defined, they may be quite useful in reformulating numerical problems in different coordinate systems.

1.5 Operators

The notion of a mathematical operator is extremely important in mathematical physics and there are entire books written on the subject. Most students first encounter operators in calculus when the notation [d/dx] is introduced to denote the operations involved in finding 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 subject 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. The operator [d/dx] is a scalar operator. That is, it provides a single result after operating on some function defined in an appropriate coordinate space. It and the operator $\int$ represent the fundamental operators of the infinitesimal calculus. Since [d/dx] and $\int$ carry out inverse operations on functions, one can define an identity operator by [d/dx]$\int$ so that continuous differentionable functions will form a group under the action of these operators.

In numerical analysis there are analogous operators Δ and Σ that perform similar functions but without taking the limit to vanishingly small values of the independent variable. Thus we could define the forward finite difference operator Δ by its operation on some function f(x) so that

$$\Delta f(x) = f(x+h) - f(x) , \qquad (1.5.1)$$

where the problem is usually scaled so that h = 1. In a similar manner Σ can be defined as

$$\sum_{i=0}^{n} f(x_i) = f(x) + f(x+h) + f(x+2h) + f(x+ih) \cdots + f(x+nh) \quad . \qquad (1.5.2)$$

Such operators are most useful in expressing formulae in numerical analysis. Indeed, it is possible to build up an entire calculus of finite differences. Here the base for such a calculus is 2 instead of e=2.7182818... as in the infinitesimal calculus . Other operators that are useful in the finite difference calculus are the shift operator E[f(x)] and the Identity operator I[f(x)] which are defined as

$$\left.\begin{aligned} E[f(x)] &\equiv f(x+h) \\ I[f(x)] &\equiv f(x) \end{aligned}\right\} \quad . \qquad (1.5.3)$$

These operators are not linearly independent as we can write the forward difference operator as

$$\Delta = E - I \quad . \qquad (1.5.4)$$

The finite difference and summation calculus are extremely powerful when summing series or evaluating convergence tests for series. Before attempting to evaluate an infinite series, it is useful to know if the series converges. If possible, the student should spend some time studying the calculus of finite differences.

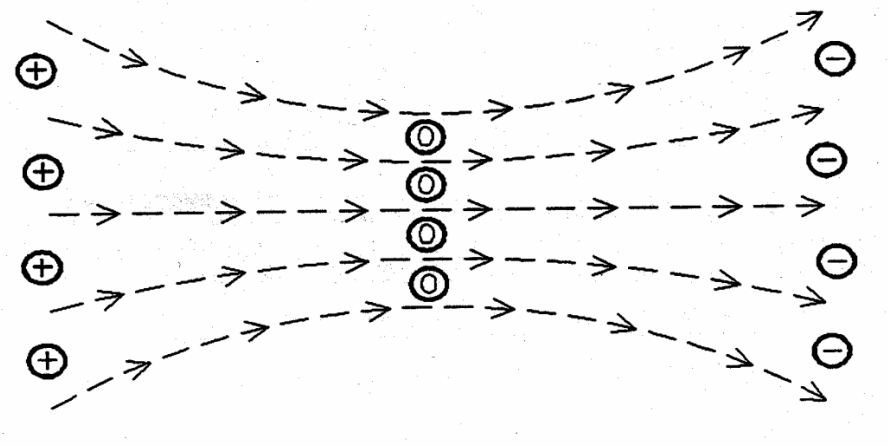
In addition to scalar operators, it is possible to define vector and tensor operators. One of the most common vector operators is the "del" operator or "nabla". It is usually denoted by the symbol ∇ and is defined in Cartesian coordinates as

$$\nabla = \hat{i}\frac{\partial}{\partial x} + \hat{j}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z} \quad . \tag{1.5.5}$$

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

$$\left.\begin{aligned} \nabla \bullet \vec{A} &= b \\ \nabla a &= \vec{B} \\ \nabla \times \vec{A} &= \vec{C} \end{aligned}\right\} , \tag{1.5.6}$$

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 1.3).



.Figure 1.3 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.

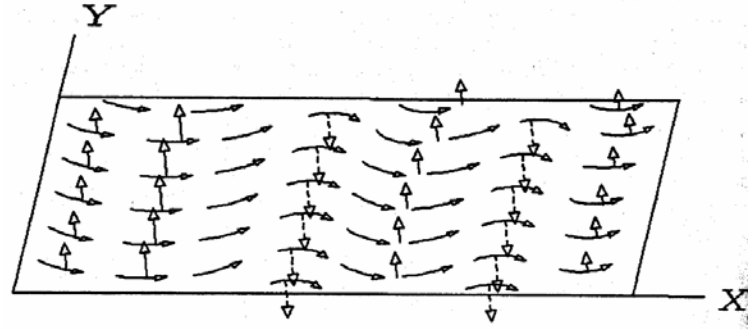


Figure 1.4 schematically shows the curl of a vector field. The direction of the curl is determined by the "right hand rule" while the magnitude depends on the rate of change of the x- *and* y- *components of the vector field with respect to* y *and* x..

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 1.4). 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 1.5).

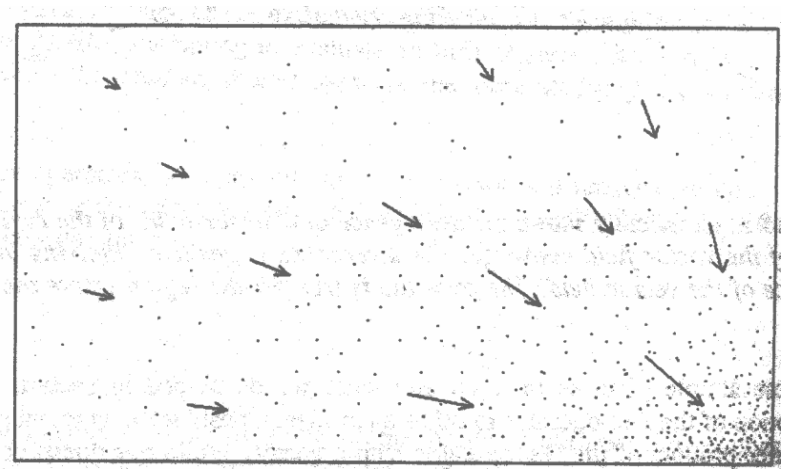


Figure 1.5 schematically shows the gradient of the scalar dot-density in the form of a number of vectors at randomly chosen points in the scalar field. The direction of the gradient points in the direction of maximum increase of the dot-density while the magnitude of the vector indicates the rate of change of that density.

With these simple pictures in mind and what we developed in section 1.4 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. 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. With some care in the definition of a collection of operators, their action on the elements of a field or group will preserve the field or group nature of the original elements. These are the operators that are of the greatest use in mathematical physics.

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 in setting up problems for their eventual numerical solution. Indeed, it is clear from the transformations described in the last sections that a prime aspect in numerically solving problems will be dealing with linear equations and matrices and that will be the subject of the next chapter

Chapter 1 Exercises

1. Show that the rational numbers (not including zero) form a group under addition and multiplication. Do they also form a scalar field?

2. Show that it is not possible to put the rational numbers into a one to one correspondence with the real numbers.

3. Show that the vector cross product is not communitative.

4. Show that the matrix product is not communitative.

5. Is the scalar product of two second rank tensors communitative? If so show how you know. If not, give a counter example.

6. Give the necessary and sufficient conditions for a tensor field.

7. Show that the Kronecker delta δ^i_j is indeed a mixed tensor.

8. Determine the nature (i.e. contravariant, covariant, or mixed) of the Levi-Civita tensor density.

9. Show that the vector cross product does indeed give rise to a pseudo-vector.

10. Use the forward finite difference operator to define a second order finite difference operator and use it to evaluate $\Delta^2[f(x)]$, where $f(x) = x^2 + 5x + 12$.

11. If $g_n(x) = x^{(n)} \equiv x(x-1)(x-2)(x-3) \cdot \cdot \cdot (x-n+1)$, show that $\Delta[g_n(x)] = n g_{n-1}(x)$. $g_n(x)$ is known as the factorial function.

12. Show that if f(x) is a polynomial of degree n, then it can be expressed as a sum of factorial functions (see problem 11).

13. Show that

$$\sum_i \varepsilon_{i\,j\,k}\varepsilon_{\,i\,p\,q} = \delta_{j\,p}\delta_{k\,q} - \delta_{j\,q}\delta_{k\,p} \quad ,$$

and use the result to prove

$$\nabla \times (\nabla \times \vec{F}) = \nabla(\nabla \bullet \vec{F}) - \nabla^2 \vec{F}$$

Chapter 1 References and Additional Reading

One of the great books in theoretical physics, and the only one I know that gives a complete list of the coordinate frames for which Laplace's equation is separable is

1. Morse, P.M., and Feshbach, H., "Methods of Theoretical Physics" (1953) McGraw-Hill Book Co., Inc. New York, Toronto, London, pp. 665-666.

It is a rather formidable book on theoretical physics, but any who aspire to a career in the area should be familiar with its contents.

While many books give excellent introductions to modern set and group theory, I have found

2. Andree, R.V., "Selections from Modern Abstract Algebra" (1958) Henry Holt & Co. New York,to be clear and concise. A fairly complete and concise discussion of determinants can be found in

3. Sokolnikoff, I.S., and Redheffer, R.M., "Mathematics of Physics and Modern Engineering" (1958) McGraw-Hill Book Co., Inc. New York, Toronto, London, pp. 741-753.

A particularly clear and approachable book on Tensor Calculus which has been reprinted by Dover is

4. Synge, J.L., and Schild, A., "Tensor Calculus" (1949) University of Toronto Press, Toronto.

I would strongly advise any student of mathematical physics to become familiar with this book before attempting books on relativity theory that rely heavily on tensor notation. While there are many books on operator calculus, a venerable book on the calculus of finite differences is

5. Milne-Thomson, L.M., "The Calculus of Finite Differences" (1933) Macmillan and Co., LTD, London.

A more elementary book on the use of finite difference equations in the social sciences is

6. Goldberg, S., "Introduction to Difference Equations", (1958) John Wiley & Sons, Inc., London.

There are many fine books on numerical analysis and I will refer to many of them in later chapters. However, there are certain books that are virtually unique in the area and foremost is

7. Abramowitz, M. and Stegun, I.A., "Handbook of Mathematical Functions" National Bureau of Standards Applied Mathematics Series 55 (1964) U.S. Government Printing Office, Washington D.C.

While this book has also been reprinted, it is still available from the Government Printing Office and represents an exceptional buy. Approximation schemes and many numerical results have been collected and are clearly presented in this book. One of the more obscure series of books are collectively known as the Bateman manuscripts, or

8. Bateman, H., "The Bateman Manuscript Project" (1954) Ed. A. Erdélyi, 5 Volumns, McGraw-Hill Book Co., Inc. New York, Toronto, London.

Harry Bateman was a mathematician of considerable skill who enjoyed collecting obscure functional relationships. When he died, this collection was organized, catalogued, and published as the Bateman Manuscripts. It is a truly amazing collection of relations. When all else fails in an analysis of a problem, before fleeing to the computer for a solution, one should consult the Bateman Manuscripts to see if the problem could not be transformed to a different more tractable problem by means of one of the remarkable relations collected by Harry Bateman. A book of similar utility but easier to obtain and use is

9. Lebedev, N.N., "Special Functions and Their Applications" (1972), Trans. R.A. Silverman. Dover Publications, Inc. New York.