LINEAR EQUATIONS
AND MATRIX ALGEBRA

Many physical problems are described by sets of simultaneous algebraic equations. Further, more difficult physical problems lead to approximations involving sets of these equations. For instance, the numerical approximation for the multigroup diffusion method results in rather simple algebraic equations.

The frequency with which sets of simultaneous algebraic equations arise motivates the introduction of a matrix notation. This notation provides a compact and convenient statement of physical and mathematical relationships, and lends itself readily to theoretical investigations and deductions. Furthermore, matrix notation leads to useful interpretation of simultaneous equations and greater understanding, which in turn induces improved methods of solution.

In this chapter we shall introduce this simplified formulation of linear algebra. We first define matrices and operations with matrices and then discuss properties of special matrices. Following the introduction of a geometric interpretation of matrix equations, we shall derive many matrix relations applied later in the text. Special attention is directed to relations of use in nuclear engineering.

1.1 Linear Equations and Matrix Notation

A simple set of linear equations in three variables might be given as

\[
\begin{align*}
3x + 2y + z &= 1, \\
x - 2y + 4z &= 2, \\
-x - y + 2z &= -1.
\end{align*}
\]
I. LINEAR EQUATIONS AND MATRIX ALGEBRA

The solution of Eqs. (1.1.1) may be found by substitution, determinants, or other means. For the moment, we postpone a discussion of solving the equations. In problems with more than three variables, the notation of Eqs. (1.1.1) is inconvenient, and we adopt a more general subscript notation. Equations (1.1.1) are written in the form

\[ a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = y_1, \]
\[ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = y_2, \]  \hspace{1cm} (1.1.2)
\[ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = y_3. \]

The quantities \( x_1, x_2, x_3 \) are the variables or unknowns. The elements on the right-hand side, \( y_1, y_2, y_3 \), are assumed known, as are the coefficients \( a_{ij} \). The notation of Eqs. (1.1.2) is conveniently extended to problems of many unknowns. Each equation of (1.1.2) is represented by one line or row of the set of equations. The first equation can be written in the compact form

\[ \sum_{j=1}^{3} a_{1j}x_j = y_1. \]  \hspace{1cm} (1.1.3)

Note the summation is over the index identifying the column of the set of equations. In a similar manner, the entire set of equations may be written

\[ \sum_{j=1}^{3} a_{ij}x_j = y_i \quad (i = 1, 2, 3). \]  \hspace{1cm} (1.1.4)

For \( n \) equations in \( n \) unknowns, the set of equations may be written

\[ \sum_{j=1}^{n} a_{ij}x_j = y_i \quad (i = 1, 2, ..., n). \]  \hspace{1cm} (1.1.5)

The notation may be simplified even further by defining several arrays of elements. We define the one-column arrays

\[ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \]  \hspace{1cm} (1.1.6)

as column matrices. Similarly, we define the one row array

\[ [a_{i1}, a_{i2}, ..., a_{in}] \]
as a row matrix. The \( i \)th equation of the set (1.1.5) may then be written

\[
\begin{bmatrix}
  a_{11}, a_{12}, \ldots, a_{1n} \\
  a_{21}, a_{22}, \ldots, a_{2n} \\
  \vdots \\
  a_{n1}, a_{n2}, \ldots, a_{nn}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{bmatrix}
= \sum_{j=1}^{n} a_{ij} x_j = y_i .
\] (1.1.7)

The definition (1.1.7) implies that the element in the \( j \)th column of the row matrix multiplies the element in the \( j \)th row of the column matrix. We define the entire array of coefficients as the square matrix

\[
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
\]

The entire set of equations may then be written

\[
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{bmatrix}
= \begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix} .
\] (1.1.8)

The \( i \)th equation of the set is found by multiplying the elements of the \( i \)th row of the square matrix into the column matrix of \( x_j \)'s.

The notation may be further simplified by denoting the one column matrices as single quantities, such as

\[
\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} .
\] (1.1.9)

Similarly, we denote the square array as

\[
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
= \mathbf{A} = [a_{ij}] .
\] (1.1.10)
The set of equations (1.1.8) become

$$\sum_{j=1}^{n} a_{ij}x_j = [y_i],$$

or, equivalently

$$Ax = y. \quad (1.1.12)$$

The form of equation (1.1.12) suggests that the quantity $A$ multiplies the quantity $x$. We shall call this multiplication of a column matrix by a square matrix. Obviously the multiplication is defined only when the number of columns of $A$ equals the number of rows of $x$. It is easily seen that the definition of multiplication may be extended to the case where the matrix $A$ is rectangular rather than square, provided only that the number of columns of $A$ equals the number of rows of $x$. A matrix of $m$ rows and $n$ columns is referred to as an $m$ by $n$ matrix.

### 1.2 Matrix Operations

Matrices may be manipulated in a manner similar to numbers. The rules for manipulation are derivable from previous results. We define two matrices as equal if corresponding elements are equal. The rule for the addition of matrices can be derived by noting that

$$\sum_j a_{ij}x_j + \sum_j b_{ij}x_j = \sum_j (a_{ij} + b_{ij})x_j,$$

and hence

$$[a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}]. \quad (1.2.2)$$

Thus addition of matrices is performed by adding corresponding elements. The definition applies only when $A$ and $B$ have the same number of rows and columns. Addition of matrices is commutative and associative.

$$A + B = B + A, \quad (1.2.3)$$

$$A + (B + C) = (A + B) + C. \quad (1.2.4)$$

The rule for multiplication of two matrices may be derived by considering two sets of simultaneous equations. Consider the sets of equations

$$Ax = y, \quad (1.2.5)$$

and

$$By = z, \quad (1.2.6)$$
where the products are assumed to exist. The \( i \)th equation of (1.2.5) is

\[
y_i = \sum_j a_{ij}x_j,
\]  

(1.2.7)

whereas the \( k \)th equation of (1.2.6) is

\[
z_k = \sum_i b_{ki}y_i.
\]  

(1.2.8)

Thus,

\[
z_k = \sum_i b_{ki} \sum_j a_{ij}x_j = \sum_j \left( \sum_i b_{ki}a_{ij} \right) x_j.
\]  

(1.2.9)

In matrix notation we have

\[
z = By,
\]  

(1.2.10)

\[
y = Ax,
\]  

(1.2.11)

\[
z = BAx.
\]  

(1.2.12)

Consequently,

\[
[B(A)]_{kj} = \left[ \sum_i b_{ki}a_{ij} \right].
\]  

(1.2.13)

The summation in Eq. (1.2.13) is to extend over the columns of \( B \) and the rows of \( A \). Therefore, matrix multiplication is defined only when the number of columns of the first matrix equals the number of rows of the second matrix. The product matrix will have as many rows as \( B \) and as many columns as \( A \).

It is easily seen that matrix multiplication is associative and distributive

\[
A(BC) = (AB)C,
\]  

(1.2.14)

\[
A(B + C) = AB + AC.
\]  

(1.2.15)

It is easily shown that matrix multiplication is not commutative; that is,

\[
AB \neq BA
\]  

(1.2.16)

in general. Note that if \( A \) and \( B \) are not square, the products cannot be equal. Even for square matrices, the matrices do not commute in general. For the special case when \( AB = BA \), we say the matrices are commutative.
Occasionally it is convenient to partition a matrix into smaller matrices or submatrices. Thus, if

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$  \hspace{1cm} (1.2.17)

then a partition of $\mathbf{A}$ might be

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$  \hspace{1cm} (1.2.18)

where the submatrices $\mathbf{A}_{11}, \mathbf{A}_{12}, \mathbf{A}_{21}, \mathbf{A}_{22}$ are

$$\mathbf{A}_{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

$$\mathbf{A}_{12} = \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix},$$

$$\mathbf{A}_{21} = [a_{31}, a_{32}],$$

$$\mathbf{A}_{22} = [a_{33}].$$

The matrix $\mathbf{A}$ is called a supermatrix. Although we shall not use more than the two levels of matrices illustrated here, it is apparent that any number of levels could be used. The usual rules of matrix algebra apply at all levels.

1.3 Determinants

A determinant may be associated with any square matrix. Whereas a matrix is an ordered collection of numbers, a determinant represents a quantity having just one value. The determinant of a square matrix is the sum of the $n!$ terms, all formed differently, each of which is constructed of $n$ factors, one and only one factor being chosen from each row and column. No two factors may come from either the same row or the same column. The sign of each term is determined by drawing straight lines connecting each factor with every other factor in the given term. If the number of these lines from all factors sloping upward to the
right is odd, the sign of the term is negative, and if the number of these lines is even, the sign of the term is positive. The determinant is represented as follows:

\[
\begin{vmatrix}
 a_{11} & a_{12} & \cdots & a_{1n} \\
 a_{21} & a_{22} & \cdots & a_{2n} \\
 \vdots & \vdots & \ddots & \vdots \\
 a_{n1} & a_{n2} & \cdots & a_{nn}
\end{vmatrix}
\]  

(1.3.1)

As an example, note that

\[
\begin{vmatrix}
 a_{11} & a_{12} & a_{13} \\
 a_{21} & a_{22} & a_{23} \\
 a_{31} & a_{32} & a_{33}
\end{vmatrix} = (a_{11}a_{22}a_{33}) - (a_{11}a_{23}a_{32}) + (a_{12}a_{23}a_{31}) - (a_{12}a_{21}a_{33}) + (a_{13}a_{21}a_{32}) - (a_{13}a_{22}a_{31}).
\]

The determinant formed from an \( n \) by \( n \) array of numbers is said to be of order \( n \).

There are a number of useful theorems that facilitate the evaluation of a determinant:

1. If all elements of a row or column are zero, the determinant is zero.
2. If all elements of a row or column are multiplied by the same factor, the determinant is multiplied by that factor.
3. The interchange of two rows or two columns changes the sign of the determinant but otherwise leaves its value unaltered.
4. Interchanging the rows and columns of a determinant does not change the value of a determinant.
5. If each element of a row of a determinant is the sum of two terms, then the value of the determinant equals the sum of the values of two determinants, one formed by omission of the first term of each binomial and the other formed by omission of the second term in each binomial.
6. The value of a determinant is not altered by adding to the elements of any row a multiple times the corresponding elements of any other row. Likewise, for columns.

The proof of these theorems is left to the problems.

The Laplace development of a determinant will enable us to find the rule for solving an array of linear equations. The minor \( M_{ij} \) of an element \( a_{ij} \) is the determinant of the matrix formed by deleting the \( i \)th row and the \( j \)th column of the original determinant. The cofactor \( C_{ij} \) of \( a_{ij} \) is then defined by

\[
C_{ij} = (-)^{i+j}M_{ij}.
\]  

(1.3.2)
The Laplace development of a determinant is then given by

\[ |A| = \sum_{j=1}^{n} a_{ij}C_{ij}, \]  

or

\[ |A| = \sum_{i=1}^{n} a_{ij}C_{ij}. \]

In words, the determinant is equal to the sum of the products of the elements in any row or column by their corresponding cofactors. The validity of this theorem follows immediately from the definition of the determinant since \( a_{ij}C_{ij} \) is just the sum over all terms containing the element \( a_{ij} \).

The sum of the products of the elements in any row by the cofactors of corresponding elements in another row is zero:

\[ \sum_{j=1}^{n} a_{ij}C_{kj} = 0 \quad (i \neq k). \]  

Similarly for columns

\[ \sum_{i=1}^{n} a_{ij}C_{ik} = 0 \quad (k \neq j). \]

The proof follows from the observation that the sum (1.3.4) is merely the determinant itself with one of its original rows replaced by another of its original rows. Such a determinant is zero since by Theorem 6 above relating to the evaluation of determinants, we could reduce one of the identical rows to zero by subtracting the other from it. Then by Theorem 1 above, the determinant would be zero. A similar development for columns applies.

The unknown \( x_k \) in a set of \( n \) linear equations in \( n \) unknowns is easily found by multiplying the equations (1.1.5) by \( C_{ik} \), by summing over \( i \) from 1 to \( n \), and by use of the relation (1.3.4b) above.

\[ |A| x_k = \sum_{i=1}^{n} C_{ik}y_i. \]

This result is known as Cramer’s rule. A solution exists only if

\[ |A| \neq 0. \]
Matrices satisfying this last condition are called nonsingular; matrices whose determinants are zero are called singular. We note that the solution exists and is unique if the number of unknowns equals the number of equations and if the determinant of the matrix formed from the coefficients is nonsingular.

The product of two determinants \(|A|\) and \(|B|\) is equal to the determinant \(|AB|\) of the product. This fact is proved in a straightforward manner. By Theorem 5 for the evaluation of determinants, the determinant of the product can be expanded in \(n^n\) determinants of the form

\[
\begin{vmatrix}
a_{1k_1}b_{k_11} & a_{1k_2}b_{k_21} & \ldots & a_{1k_n}b_{k_n1} \\
a_{2k_1}b_{k_11} & a_{2k_2}b_{k_21} & \ldots & a_{2k_n}b_{k_n1} \\
\vdots & \vdots & \ddots & \vdots \\
a_{nk_1}b_{k_11} & a_{nk_2}b_{k_21} & \ldots & a_{nk_n}b_{k_n1}
\end{vmatrix}
\]

(1.3.7)

where the \(k_1, k_2, \ldots, k_n\) stands for any \(n\) values of the subscript \(j\). Only the determinants in which the values of all \(j\) are different contribute to the sum in the expansion of the determinant of the product by Theorems 1, 2, and 6 above. (If any two columns are multiples of each other, the determinant is zero.) Therefore, the sum of \(n^n\) terms in the expansion of the determinant of the product consists of only \(n!\) terms each of the form

\[
(b_{k_11}b_{k_22} \ldots b_{k_n1})
\]

in which all \(k_j\) are different. An interchange of the columns of the determinant shown reduces it to the exact form of \(|A|\). We have thus \(n!\) terms, all different, in the expansion of \(|AB|\) made above, each comprising \(|A|\) times one term of \(|B|\) together with the correct sign. Thus

\[
|AB| = |A| \cdot |B|,
\]

(1.3.9)

as was to be proved.

1.4 Solution of Simultaneous Equations

We now consider a systematic procedure for solving sets of equations and determine conditions under which solutions do exist. The procedure
to be outlined is called the Gauss reduction. Consider a set of \( m \) equations in \( n \) unknowns

\[
\sum_{j=1}^{n} a_{ij} x_j = y_i \quad (i = 1, 2, \ldots, m),
\]

(1.4.1)

or

\[
a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n = y_1,
\]

\[
a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n = y_2,
\]

\[
\vdots
\]

\[
a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n = y_m.
\]

(1.4.2)

We assume the coefficient \( a_{11} \neq 0 \), otherwise renumber the equations so that we have an \( a_{11} \neq 0 \). We may eliminate the variable \( x_1 \) from the other \( m-1 \) equations. To this end divide the first equation by \( a_{11} \) to obtain

\[
x_1 + \frac{a_{12}}{a_{11}} x_2 + \ldots + \frac{a_{1n}}{a_{11}} x_n = \frac{y_1}{a_{11}},
\]

(1.4.3a)

or

\[
x_1 + a'_{12} x_2 + \ldots + a'_{1n} x_n = y'_1.
\]

(1.4.3b)

We multiply Eq. (1.4.3b) successively by \( a_{21}, a_{31}, \ldots, a_{m1} \) and subtract the resultant equations from the second, third, etc. equations of (1.4.2). The result is a set of equations of the form

\[
x_1 + a'_{12} x_2 + \ldots + a'_{1n} x_n = y'_1,
\]

\[
a_{22} x_2 + a'_{23} x_3 + \ldots + a'_{2n} x_n = y'_2,
\]

\[
\vdots
\]

\[
a_{m2} x_2 + a'_{m3} x_3 + \ldots + a'_{mn} x_n = y'_m.
\]

(1.4.4)

We now divide the second equation of (1.4.4) by \( a_{22} \) and eliminate \( x_2 \) from the remaining \( m-2 \) equations as before. We continue in this manner to eliminate the unknowns \( x_i \). If \( m = n \), the set of equations takes the form

\[
x_1 + a'_{12} x_2 + \ldots + a'_{1n} x_n = y'_1,
\]

\[
x_2 + a'_{23} x_3 + \ldots + a'_{2n} x_n = y'_2,
\]

\[
\vdots
\]

\[
x_{n-1} + a'_{n-1,n} x_n = y'_{n-1},
\]

\[
a'_{nn} x_n = y'_n.
\]

(1.4.5)
If \( a_{nn}'' \neq 0 \), then by back substitution we may evaluate the \( x_i \). If \( a_{nn}'' = 0 \) and \( y_n'' = 0 \), then \( x_n \) is indeterminate, and we do not obtain a unique solution. It is easily shown \( a_{nn}'' = 0 \) only if \( |A| = 0 \). If \( a_{nn}'' = 0 \) and \( y_n'' \neq 0 \), then no solution to the equations exists.

The results may be generalized for \( m \neq n \). If \( m > n \) the reduction process will lead to a set of equations of the form

\[
\begin{align*}
  x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1', \\
  x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= y_2', \\
  \vdots \\
  x_n &= y_n', \\
  0 &= y_{n+1}' \\
  \vdots \\
  0 &= y_m'.
\end{align*}
\]  

(1.4.6)

If the \( y_{n+1}', y_{n+2}', \ldots, y_m' \) are all zero, then we again have a unique solution; the last \( m-n \) equations are merely linear combinations of the first \( n \) equations. On the other hand, if any \( y_p'' (n < p \leq m) \) are not zero, then the equations are inconsistent and no solution exists. In like manner, for \( m < n \) the reduction leads to

\[
\begin{align*}
  x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1', \\
  x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= y_2', \\
  \vdots \\
  x_m + a_{m+1}x_{m+1} + \cdots + a_{mn}x_n &= y_n'.
\end{align*}
\]  

(1.4.7)

In this case the variables \( x_{m+1}, x_{m+2}, \ldots, x_n \) may be assigned arbitrarily and the remaining \( x_i \) determined in terms of the arbitrary variables. Obviously there is not a unique solution in this case.

The above results may be expressed in a compact theorem. To this end we introduce the concept of the rank of a matrix and define the coefficient and augmented matrices associated with a set of linear equations. Consider the set of equations (1.4.1). The coefficient matrix associated with this set of equations is

\[
\begin{bmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}
\]  

(1.4.8)
The augmented matrix is defined as the \( m \) by \( n + 1 \) matrix formed by appending the column matrix \([y_i]\) to the coefficient matrix. Thus, the augmented matrix associated with Eq. (1.4.1) is
\[
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} & y_1 \\
a_{21} & a_{22} & \cdots & a_{2n} & y_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn} & y_m
\end{bmatrix}
\] (1.4.9)

The rank of a matrix is defined to be the order of the largest nonvanishing determinant contained in the matrix. Obviously, the rank of the coefficient matrix can never exceed the rank of the augmented matrix.

The rank of a matrix is unaltered by multiplying all the elements of a row by a constant or by adding a row times a constant to another row of the matrix. The result follows from Theorems 2 and 6 relating to the evaluation of determinants.

The simple theorem relating to the solution of a system of linear equations may now be stated as follows: a solution to a system of linear equations exists if and only if the ranks of the coefficient and augmented matrices are equal. The proof of the theorem follows from the Gauss reduction. The ranks of the coefficient matrices of the array of Eqs. (1.4.2) and (1.4.6) are equal by Theorems 2 and 6. Likewise, the ranks of the augmented matrices are equal. Accordingly, consider Eq. (1.4.6); the coefficient matrix is of rank \( n \). If any \( y_i \neq 0, n < p \leq m \), then the augmented matrix is of rank greater than \( n \). But under these circumstances, no solution exists and hence the theorem follows.

From the previous work, it is also clear that if the common rank, \( r \), of the augmented and coefficient matrices is less than the number, \( n \), of unknowns, then \( n-r \) of the unknowns may have their values assigned arbitrarily. In this case, the remaining variables are uniquely determined as linear functions of the \( n-r \) unknowns whose values have been arbitrarily chosen.

A special case occurs if all the inhomogeneous terms \( y_i \) are zero. In this case the coefficient and augmented matrices are always of the same rank, and hence a solution always exists. However, this result is evident, since in this case we have the trivial solution \( x_i = 0 \). A nontrivial solution will exist only if the rank of the coefficient matrix is less than the number of unknowns, otherwise only the trivial solution exists.

1.5 Special Matrices and Their Properties

There are many special matrices that are of frequent interest. We shall assume throughout the chapter that all of the matrices have real ele-
1.5 SPECIAL MATRICES AND THEIR PROPERTIES

There are generalizations of the results to matrices with complex elements, but the generalizations are not of interest for this work. The zero matrix \( \mathbf{0} \) is a square matrix, each of whose elements is zero. The product of the zero matrix and any other matrix is a zero matrix. Note that if the product of two matrices is zero, we cannot conclude one of the matrices is zero, however. Consider the simple example

\[
\begin{bmatrix}
1 & -1 \\
-1 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 1 \\
1 & 1
\end{bmatrix}
= 
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}.
\]  

(1.5.1)

The unit matrix \( \mathbf{I} \) is a square matrix, each of whose nondiagonal elements is zero, and each of whose diagonal elements is unity:

\[
\mathbf{I} = [\delta_{ij}],
\]  

(1.5.2)

where \( \delta_{ij} \) is the Kronecker delta function

\[
\delta_{ij} = \begin{cases} 
0, & i \neq j, \\
1, & i = j.
\end{cases}
\]  

(1.5.3)

The product of the unit matrix with any other matrix \( \mathbf{A} \) of the same order is merely \( \mathbf{A} \). Further, the unit matrix commutes with any other matrix.

A diagonal matrix is a matrix each of whose nondiagonal elements is zero.

\[
\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\
0 & d_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & d_n
\end{bmatrix}.
\]  

(1.5.4)

Two diagonal matrices commute, but a diagonal matrix does not commute with other matrices in general. A scalar matrix is a diagonal matrix all of whose diagonal elements are equal. A scalar matrix commutes with any other matrix.

The transpose of a matrix \( \mathbf{A} \), denoted \( \mathbf{A}^T \), is formed from \( \mathbf{A} \) by interchanging rows and columns of \( \mathbf{A} \). Therefore

\[
\mathbf{A}^T = [a_{ij}] = [a_{ji}].
\]  

(1.5.5)

The transpose of a product of matrices satisfies the relation

\[
(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T,
\]  

(1.5.6)

a result readily proved. Note that

\[
(\mathbf{AB})^T = \left[ \sum_k a_{ik} b_{kj} \right]^T = \left[ \sum_k a_{kj} b_{ki} \right] = \mathbf{B}^T \mathbf{A}^T.
\]  

(1.5.7)

If \( \mathbf{A}^T = \mathbf{A} \), then the matrix is said to be symmetric. If \( \mathbf{A}^T = -\mathbf{A} \), then the matrix is said to be antisymmetric.
The adjoint\(^1\) of a matrix is defined only for square matrices. We define the adjoint of \(A\), written \(\text{adj } A\), as the matrix formed by replacing each element of \(A\) by the cofactor of its transpose:

\[
\text{adj } A = [\text{adj } a_{ij}] = [C_{ij}], \tag{1.5.8}
\]

where \(C_{ij}\) is the cofactor of the \(ij\)th element of \(A\). It is easily proved that

\[
\text{adj}(AB) = (\text{adj } B)(\text{adj } A). \tag{1.5.9}
\]

The inverse of a matrix \(A\), written \(A^{-1}\), is a matrix such that

\[
AA^{-1} = I. \tag{1.5.10}
\]

Note that

\[
A^{-1}A = I. \tag{1.5.11}
\]

Since

\[
|A| |A^{-1}| = |AA^{-1}| = |I| = 1, \tag{1.5.12}
\]

the inverse of a matrix exists only for nonsingular square matrices. Let \(a_{jk}^{-1}\) be the \(jk\)th element of the inverse matrix \(A^{-1}\). Then

\[
\left[ \sum_j a_{ij} a_{jk}^{-1} \right] = I. \tag{1.5.13}
\]

To find the elements of \(A^{-1}\), we recall the Laplace expansion theorem (1.3.3), which can be written

\[
\sum_j a_{ij} \frac{C_{kj}}{|A|} = I. \tag{1.5.14}
\]

Hence, if

\[
a_{jk}^{-1} = \frac{C_{kj}}{|A|} = \frac{\text{adj } a_{jk}}{|A|}, \tag{1.5.15}
\]

\[
A^{-1} = \frac{\text{adj}[a_{jk}]}{|A|}, \tag{1.5.16}
\]

Eq. (1.5.13) will be satisfied. The uniqueness of \(A^{-1}\) is proved by supposing that there were a second inverse, say \(B\). In this case,

\[
A(A^{-1} - B) = I - I = 0.
\]

\(^1\)In this book we have no need to define the Hermitian adjoint, often called merely the adjoint, of a matrix. The Hermitian adjoint and the adjoint are not related.
1.6 VECTOR INTERPRETATION

Now, multiply on the left by either inverse to learn that

$$B = A^{-1}$$

and the two inverses are identical. It is easily shown that

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{1.5.17}$$

The inverse matrix is essentially that which has been calculated in Cramer’s rule (1.3.5). If

$$Ax = y, \tag{1.5.18}$$

then

$$x = A^{-1}y. \tag{1.5.19}$$

If, for a real matrix $A$

$$A^T = A^{-1}, \tag{1.5.20}$$

then the matrix $A$ is called orthogonal. Note that

$$|A^T| = |A|, \tag{1.5.21}$$

and

$$|A^{-1}| = |A|^{-1}, \tag{1.5.22}$$

and consequently the determinant of an orthogonal matrix is $\pm 1$.

1.6 Vector Interpretation

Matrix equations may be given a very convenient and useful interpretation in terms of vectors and operations among these vectors. Vectors may be interpreted as special cases of matrices of such importance that a special abbreviated notation is used. As we shall see, operations on these vectors may then be given a geometric interpretation. We recall that a vector$^2$ in three dimensions may be written

$$t = t_1i + t_2j + t_3k, \tag{1.6.1}$$

where $i$, $j$, and $k$ are unit vectors along three mutually perpendicular coordinate axes, and $t_1$, $t_2$ and $t_3$ are the components of $t$ along the various axes. If we define the row matrix $E$ as

$$E = (i, j, k), \tag{1.6.2}$$

$^2$ We define a vector here as an ordered collection of $n$ entities, called components, in an $n$-dimensional space, without implying any particular transformation properties. [A vector is also often defined to be a quantity whose components transform as the coordinates. We do not use this definition in this book].
then Eq. (1.6.1) can be written
\[
\mathbf{t} = \mathbf{E} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}.
\] (1.6.3)

It is usually convenient to assume the underlying coordinate system \( \mathbf{E} \) is fixed throughout the discussion and to denote the vector \( \mathbf{t} \) as a column matrix
\[
\mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}.
\] (1.6.4)

We shall adopt this shorthand notation and shall further assume the coordinate system \( \mathbf{E} \) is constructed of mutually orthogonal axes.\(^3\)

The scalar product of two vectors, \( \mathbf{t} \) and \( \mathbf{u} \), in vector analysis is
\[
(\mathbf{t}, \mathbf{u}) = t_1u_1 + t_2u_2 + t_3u_3.
\] (1.6.5)

In matrix notation the scalar product is
\[
(\mathbf{t}, \mathbf{u}) = \mathbf{u}^T \mathbf{t} = \mathbf{t}^T \mathbf{u},
\] (1.6.6)

where the transpose of a column matrix is a row matrix. Frequently we shall refer to a column matrix as a column vector.

Matrix equations may also be given a useful vector interpretation. The equation
\[
\mathbf{y} = \mathbf{A} \mathbf{x}
\] (1.6.7)

is interpreted as a relation between two vectors \( \mathbf{y} \) and \( \mathbf{x} \). In particular, the matrix \( \mathbf{A} \) acts as a transformation which transforms the vector \( \mathbf{x} \) into another vector \( \mathbf{y} \). An alternative viewpoint is to consider \( \mathbf{x} \) and \( \mathbf{y} \) as the same vector expressed in two different coordinate systems. The matrix \( \mathbf{A} \) then specifies the relation between the components of the vector in the two different coordinate systems. A geometric portrayal of the two different interpretations is given in Figs. 1.6.1 and 1.6.2.

Either interpretation of the equation is found to be useful. For our later purposes, the first viewpoint will be more frequently employed. The concepts of the vector interpretation of matrices may be extended

\(^3\) If the coordinate system is not an orthogonal system, the results to be obtained subsequently must be generalized. See Section 1.12.
Fig. 1.6.1. Geometric view of the matrix equation $Ax = y$ considered as a transformation of a vector.

Fig. 1.6.2. Geometric view of the matrix equation $Ax = y$ considered as a transformation of the coordinate system.
I. LINEAR EQUATIONS AND MATRIX ALGEBRA

to \( n \)-dimensional spaces in a straightforward manner. The column matrix

\[
x = \begin{bmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_n
\end{bmatrix}
\]

(1.6.8)
is interpreted as a vector in \( n \) space where \( x_i \) are components of \( x \) along the \( i \)th coordinate axis.

1.7 Matrix Functions and Similarity Transformations

We can now define under certain conditions a function \( f(A) \) of a nonsingular matrix \( A \), since both positive and negative integral powers of this matrix are available. If \( f(x) \) may be expanded in a Laurent series so that

\[
f(x) = \sum_{i=-\infty}^{\infty} b_i x^i,
\]

then

\[
f(A) = \sum_{i=-\infty}^{\infty} b_i A^i,
\]

(1.7.2)

where \( b_i \) is the coefficient of \( A^i \), \( b_i \) not being a matrix. If \( A \) is symmetric, then \( f(A) \) will be symmetric. We observe that two functions \( f \) and \( g \) of the same matrix \( A \) commute:

\[
f(A)g(A) = g(A)f(A).
\]

(1.7.3)

Two matrices \( A \) and \( B \) are called equivalent if and only if they are related by two nonsingular matrices \( R \) and \( Q \) as follows:

\[
RAQ = B.
\]

(1.7.4)
The factor \( R \) merely causes each new row of \( B \) to be a linear combination of the original rows of \( A \), and the factor \( Q \) merely linearly combines the old columns of \( A \) into new columns of \( B \), as follows from the definition of a product. The matrix operators \( R \) may also exchange rows; the matrix \( Q \) may exchange columns. Since these operations leave the rank of a matrix unchanged, \( A \) and \( B \) have the same rank.

The matrices \( R \) and \( Q \) that linearly combine the rows or columns of \( A \) in a particular way are easily constructed by linearly combining the
1.7 MATRIX FUNCTIONS AND SIMILARITY TRANSFORMATIONS

rows or columns, respectively, of the unit matrix in the same way. The first and second rows, for example, are interchanged by the nonsingular operator

\[
\begin{bmatrix}
0 & 1 & 0 & 0 & \ldots \\
1 & 0 & 0 & 0 & \ldots \\
0 & 0 & 1 & 0 & \ldots \\
0 & 0 & 0 & 1 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\]

Again, a multiple \( C \) of the second row of \( A \) is added to the first row of \( A \) by the operator

\[
\begin{bmatrix}
1 & C & 0 & 0 & \ldots \\
0 & 1 & 0 & 0 & \ldots \\
0 & 0 & 1 & 0 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\]

The matrix is nonsingular. Since exchanging the rows or columns of a matrix and since linearly combining the rows or columns of a matrix do not alter the value of any minor, the matrices \( R \) and \( Q \) are clearly nonsingular since the unit matrix is.

If \( R = Q^{-1} \), the transformation is called a similarity transformation:

\[
B = Q^{-1}AQ. \tag{1.7.5}
\]

If, on the other hand, \( R = Q^T \), the transformation is called a congruence transformation:

\[
B = Q^TAQ. \tag{1.7.6}
\]

If \( R = Q^{-1} = Q^T \), so that \( Q \) is orthogonal, the transformation is called an orthogonal transformation.

All matrix relations are equally valid if all matrices occurring in these relations are subjected to the same similarity transformation. If \( AB = C \)

\[
Q^{-1}CQ = (Q^{-1}AQ)(Q^{-1}BQ), \tag{1.7.7}
\]

and if \( A + B = C \),

\[
Q^{-1}CQ = Q^{-1}AQ + Q^{-1}BQ. \tag{1.7.8}
\]

Again, suppose we had two vectors \( x_0 \) and \( y_0 \) related by

\[
y_0 = Ax_0. \tag{1.7.9}
\]
If we introduce new vectors, \( \chi \) and \( y \), defined by
\[
\begin{align*}
x_0 &= Qx, \\
y_0 &= Qy,
\end{align*}
\] (1.7.10)
where \( Q \) is nonsingular, then
\[
\begin{align*}
y &= Q^{-1}AQx = Bx,
\end{align*}
\] (1.7.11)
whence we see that the two new vectors, \( \chi \) and \( y \), are related to each other exactly like the old ones, \( x_0 \) and \( y_0 \), providing the new and old operators are related by
\[
\begin{align*}
\mathbf{B} &= Q^{-1}AQ.
\end{align*}
\] (1.7.12)

If \( Q \) be a real orthogonal matrix, then \( Q \) satisfies the definition (1.5.20), and the scalar product of two vectors \( x_0 \) and \( y_0 \) is given by
\[
\begin{align*}
y_0^T x_0 &= y^T Q^T Q x = y^T x
\end{align*}
\] (1.7.13)
from which we see that the length of a vector is unaltered (i.e., if we let \( y_0 = x_0 \), then the present result shows that the length of \( x \) equals the length of \( x_0 \)), and the angle between two original vectors is also unchanged by an orthogonal transformation. Thus, unit vectors which are originally orthogonal will remain orthogonal unit vectors, hence the name orthogonal transformation.

A particularly useful orthogonal transformation is the permutation transformation. A permutation matrix is any matrix for which there is one and only one nonzero element in each row and column of the matrix, and the nonzero element is unity. Thus, the unit matrix is a permutation matrix. If we denote a permutation matrix as \( P \), then a permutation transformation is a similarity (orthogonal) transformation of the form
\[
PAP^T = PAP^{-1}.
\]

A permutation matrix merely interchanges certain rows and columns of a matrix.

The trace of a matrix is the sum of the diagonal elements:
\[
\begin{align*}
\text{Tr} \ A &= \sum_i a_{ii}.
\end{align*}
\] (1.7.14)

The trace of the product of two matrices is independent of the order of the factors
\[
\begin{align*}
\text{Tr}(AB) &= \sum_{i=1}^n \sum_{j=1}^n a_{ij}b_{ji} = \text{Tr}(BA).
\end{align*}
\] (1.7.15)
The trace of a matrix is unaltered by a similarity transformation:

$$\text{Tr}(Q^{-1}AQ) = \sum_{i,j,k=1}^{n} (Q^{-1})_{ij}a_{jk}(Q)_{kl} = \sum_{j=1}^{n} a_{jj} = \text{Tr} A.$$  \hspace{1cm} (1.7.16)

### 1.8 Linear Independence of Vectors and Orthogonalization of Vectors

An array $x_i, 1 \leq i \leq n,$ of vectors is said to be linearly dependent when

$$\sum_{i=1}^{n} b_i x_i = 0,$$  \hspace{1cm} (1.8.1)

where not all $b_i$ are zero. When no set of $b_i$ exists in which at least one $b_i$ differs from zero for which equation (1.8.1) is true, the array of vectors is said to be linearly independent. We can easily generate a linearly independent array of vectors from a linearly dependent array by discarding all zero vectors (which are not very interesting anyway), by examining each of the remaining vectors one by one, and by keeping only those which are linearly independent of all the vectors already selected. The remaining vectors are then linearly related to those selected, since otherwise they would have been selected.

A test for linear independence is readily constructed by observing that the equation (1.8.1) may be regarded as an array of $n$ linear homogeneous equations in which the components $x_{ij}$ of the vectors $x_i$ are the coefficients of the unknowns $b_i$:

$$\sum_{i=1}^{n} b_i x_{ji} = 0 \quad (j = 1, \ldots, n).$$  \hspace{1cm} (1.8.2)

Indeed, one can associate the $j$th element $x_{ji}$ of a matrix with the $j$th component of the $i$th vector, in which case each vector forms one column of a matrix, or with the $i$th component of the $j$th vector, in which case each vector forms one row of a matrix. The vectors will then be linearly dependent if and only if nontrivial solutions $b_i$ of the array of linear homogeneous equations (1.8.2) exist. By Section 1.4, we have seen that the necessary and sufficient condition for the existence of nontrivial solutions of such equations is that the determinant $|x_{ji}|$ of $x_{ji}$ vanish. Consequently, an array of vectors is linearly dependent if and only if the determinant formed from their components vanishes. The square of this determinant is called the Gram determinant of the vector array.

If and only if the Gram determinant vanishes, the array of vectors is
I. LINEAR EQUATIONS AND MATRIX ALGEBRA

linearly dependent. The present test requires only that the components of each vector along the others be known.

\[
\begin{vmatrix}
    x_1^2 & (x_1, x_2) & (x_1, x_3) & \cdots & (x_1, x_n) \\
    (x_2, x_1) & x_2^2 & (x_2, x_3) & \cdots & (x_2, x_n) \\
    (x_3, x_1) & (x_3, x_2) & x_3^2 & \cdots & (x_3, x_n) \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    (x_n, x_1) & (x_n, x_2) & (x_n, x_3) & \cdots & x_n^2
\end{vmatrix} = 0.
\]

The condition

\[
\begin{vmatrix}
    x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\
    x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\
    x_{31} & x_{32} & x_{33} & \cdots & x_{3n} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    x_{n1} & x_{n2} & x_{n3} & \cdots & x_{nn}
\end{vmatrix} = 0
\]

requires that the components along some arbitrary coordinate system be known.

There cannot be more than \(n\) linearly independent vectors each of which is of dimension \(n\). If then a space has \(n\) dimensions, any vector \(\mathbf{u}\) can be expanded in terms of any set of \(n\) linearly independent vectors.

\[
\mathbf{u} = \sum_{i=1}^{n} b_i \mathbf{x}_i
\]  

(1.8.3)

where the \(b_i\) can be found by Cramer's rule if the equation (1.8.3) be written out in component form. A set of \(n\) linearly independent vectors in a space of \(n\) dimensions and in terms of which other vectors are expanded is called a basis. An incomplete set of \(r\) vectors is said to be of rank \(r\) for evident reasons and to be of defect \(n - r\). A basis is usually chosen to be orthogonal and normal.

Should the basis not be orthogonal, it may be made orthogonal quite easily by the Schmidt procedure, which essentially consists in subtracting the projection of any particular vector on any previously orthogonalized vectors from that particular vector in forming a new vector. Consider the set of vectors \(\mathbf{x}_i\) which are not orthogonal. The first vector of the orthogonal set, say the \(\mathbf{t}_1\) set, is defined by

\[
\mathbf{t}_1 = \mathbf{x}_1,
\]  

(1.8.4)
and the second by
\[ t_2 = x_2 - \frac{(x_2, t_1)t_1}{(t_1, t_1)}. \] (1.8.5)

The vector \( t_2 \) is orthogonal to \( t_1 \) because any component of \( x_2 \) that lies along \( t_1 \) has been subtracted from \( x_2 \). The third orthogonal vector \( t_3 \) is then given by
\[ t_3 = x_3 - \frac{(x_3, t_1)}{(t_1, t_1)} t_1 - \frac{(x_3, t_2)}{(t_2, t_2)} t_2. \] (1.8.6)

The remaining vectors of the set \( t_i \) are computed in like manner. If there are as many vectors \( t_i \) as dimensions of the space, then these vectors form an orthogonal, linearly independent set which span the space, i.e., are such that any arbitrary vector can be expressed in terms of them.

The method of orthogonalization cannot fail. Suppose it were to fail. Then some vector \( t_r \) would be zero. Thus, \( x_r \) would be some linear combination of \( x_1, x_2, \ldots, x_{r-1} \) contrary to the hypothesis that the original basis was linearly independent. Therefore, all \( t_r \) must differ from zero. The new vectors may now be normalized by dividing them by their own length.

1.9 Eigenvalues and Eigenvectors

The transformation applied to a vector by a matrix may conceivably merely lead to a multiple of the original vector.
\[ A x = \lambda x. \] (1.9.1)

Such a vector \( x \) is called an eigenvector and the multiple \( \lambda \) is called an eigenvalue. These two concepts are of transcendent importance in theoretical work. There may be a number of eigenvectors and eigenvalues associated with a particular operator.

We see that the eigenvector-eigenvalue equation (1.9.1) actually represents a series of linear, homogeneous equations. In order that there be a nontrivial solution in Section 1.3, we have seen it is necessary and sufficient that
\[ |A - \lambda I| = 0. \] (1.9.2)

This equation determines the possible eigenvalues and is called the characteristic equation. In a space of \( n \) dimensions, it is a polynomial equation of order \( n \), which will therefore have \( n \) roots. The roots will occur in complex conjugate pairs; some of the roots may have the same
value. The number that do is called the multiplicity of the root. If the \( n \) roots are distinct, then there are \( n \) associated eigenvectors. For repeated roots, there may be less than \( n \) eigenvectors.

A similarity transformation does not change the eigenvalues, since the characteristic equation is unaltered.

\[
| Q^{-1} A Q - \gamma I | = | Q^{-1} (A - \gamma I) Q | = | Q^{-1} (A - \gamma I) | | Q | = | A - \gamma I | = 0. \tag{1.9.3}
\]

Therefore, \( \gamma_i = \lambda_i \) if roots of the two polynomials be properly ordered.

Since Eq. (1.9.1) is homogeneous, only the directions of the eigenvectors are determined. The eigenvectors may be multiplied by any arbitrary constant and still be eigenvectors. It is usually convenient to scale the eigenvectors so that they have unit length.

The eigenvalues of a real symmetric matrix are real. To prove the result, let \( x_i \) be such that

\[
A x_i = \lambda_i x_i. \tag{1.9.4}
\]

Since the characteristic equation is a polynomial with real coefficients, there is also a root \( \bar{\lambda}_i \), which is the complex conjugate of \( \lambda_i \). The corresponding eigenvector \( \bar{x}_i \) will have components which are complex conjugate to those of \( x_i \). Therefore, we also have

\[
A \bar{x}_i = \bar{\lambda}_i \bar{x}_i. \tag{1.9.5}
\]

We multiply Eq. (1.9.4) by \( \bar{x}_i^T \), Eq. (1.9.5) by \( x_i^T \), subtract and obtain

\[
\bar{x}_i^T A x_i - x_i^T A \bar{x}_i = (\lambda_i - \bar{\lambda}_i) \bar{x}_i^T x_i. \tag{1.9.6}
\]

But

\[
x_i^T A \bar{x}_i = \bar{x}_i^T A^T x_i = \bar{x}_i^T A x_i, \tag{1.9.7}
\]

the last result since \( A \) is symmetric. Equation (1.9.6) becomes

\[
(\lambda_i - \bar{\lambda}_i) \bar{x}_i^T x_i = 0. \tag{1.9.8}
\]

The quantity

\[
\bar{x}_i^T x_i = (\bar{x}_i, x_i) \tag{1.9.9}
\]

is the generalization of the length of a vector for complex components. Since the elements are complex conjugate, the length is a positive real number. Equation (1.9.8) can be true only if

\[
\lambda_i = \bar{\lambda}_i, \tag{1.9.10}
\]

which proves the theorem.
The eigenvectors associated with eigenvalues of different value of a real symmetric matrix are orthogonal. To prove this, let $\lambda_1$, $x_1$ and $\lambda_2$, $x_2$ be such that

$$A x_1 = \lambda_1 x_1,$$  
(1.9.11)  
$$A x_2 = \lambda_2 x_2.$$  
(1.9.12)

with $\lambda_1 \neq \lambda_2$. We again multiply by $x_1^T$ and $x_2^T$ respectively and subtract. We have

$$(x_1^T A x_1 - x_2^T A x_2) = (\lambda_1 - \lambda_2) x_2^T x_1 = 0.$$  
(1.9.13)

Since $\lambda_1 \neq \lambda_2$, we must have

$$x_1^T x_2 = x_2^T x_1 = 0.$$  
(1.9.14)

If the eigenvalues of a real symmetric matrix are all distinct, then for each eigenvalue there is an eigenvector which is orthogonal to all of the other eigenvectors. If there are $n$ vectors in all, then these $n$ vectors are complete: that is, the vectors span the $n$-dimensional space and may therefore be used as a basis. The orthogonal basis of eigenvectors is a particularly useful coordinate system for a given problem. As an example, suppose we desire to study the effect of a transformation $A$ on an arbitrary vector $x$. If the eigenvectors of $A$ are the complete orthonormal set $e_i$, then we may expand $x$ in the form

$$x = \sum_i a_i e_i$$  
(1.9.15)

where the $a_i$ are expansion coefficients given by

$$a_i = x^T e_i.$$  
(1.9.16)

We then have

$$A x = A \left( \sum_i a_i e_i \right) = \sum_i a_i A e_i = \sum_i a_i \lambda_i e_i.$$  
(1.9.17)

Hence the operation of multiplying by $A$ merely multiplies the various components of $x$ by the corresponding eigenvalues. In our later work we shall make frequent use of this result.

In the event that not all the eigenvalues of a real symmetric matrix are distinct, it is still possible to construct a set of complete orthogonal eigenvectors. For any repeated root of multiplicity $k$, there are $k$ associated eigenvectors, which may be made orthogonal.\(^4\)

\(^4\) The proof of these remarks is simple but detailed. See Reference 1, pp. 59–61.
A real symmetric matrix $A$ may be transformed into a particularly simple form by a similarity transformation. Let the components of the eigenvectors be written as column matrices, thus

$$
e_i = \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{in} \end{bmatrix}.$$  

Let the matrix $M$ be defined as

$$M = [e_1, e_2, \ldots, e_n] = \begin{bmatrix} e_{11} & e_{21} & \cdots & e_{n1} \\ e_{12} & e_{22} & \cdots & e_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1n} & e_{2n} & \cdots & e_{nn} \end{bmatrix}.$$  

The eigenvectors are orthogonal, and we assume they are normalized. The matrix $M$ is called the normalized modal matrix. The product $AM$ is then

$$AM = \begin{bmatrix} \lambda_1 e_{11} & \lambda_2 e_{21} & \cdots & \lambda_n e_{n1} \\ \lambda_1 e_{12} & \lambda_2 e_{22} & \cdots & \lambda_n e_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 e_{1n} & \lambda_2 e_{2n} & \cdots & \lambda_n e_{nn} \end{bmatrix} = MD,$$  

where

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$  

Thus, we have

$$M^{-1}AM = D.$$  

The inversion of $M$ is always possible since $M$ cannot vanish by the orthogonality and consequent independence of the $e_i$.

The result shows that a real symmetric matrix is similar to a diagonal matrix. The process of so transforming a matrix is called diagonalization.
1.10 NONSYMMETRIC MATRICES

It is interesting to note that the similarity transformation used above is also an orthogonal transformation. To see this, we form the product

\[ M^T M = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nn} \end{bmatrix} \begin{bmatrix} e_{11} & e_{21} & \cdots & e_{n1} \\ e_{12} & e_{22} & \cdots & e_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1n} & e_{2n} & \cdots & e_{nn} \end{bmatrix} \]  

(1.9.23)

Since the vectors \( e_i \) are orthogonal and normalized, we have

\[ M^T M = I, \]  

(1.9.24)

and hence

\[ M^T = M^{-1}. \]  

(1.9.25)

The normalized modal matrix is an orthogonal matrix.

1.10 Nonsymmetric Matrices

The results of the previous section do not apply in full generality to a nonsymmetric matrix. We again consider only matrices with real elements.

We show first that, if the characteristic roots of a square \( n \times n \) nonsymmetric matrix \( A \) are distinct, then there are \( n \) linearly independent eigenvectors associated with the matrix. Let the eigenvalues be denoted \( \lambda_i \) and the eigenvectors as \( e_i \). We have

\[ A e_i = \lambda_i e_i \]  

(1.10.1)

for \( i = 1, 2, \ldots, n \). If the eigenvectors are linearly dependent, then at least one of the eigenvectors, say \( e_n \), is a linear combination of the remaining \( n - 1 \). Then

\[ e_n = \sum_{i=1}^{n-1} a_i e_i. \]  

(1.10.2)

Not all of the \( a_i \) are zero. Applying the operator \( A \) to both sides of Eq. (1.10.2), we have

\[ A e_n = \lambda_n e_n = \sum_{i=1}^{n-1} a_i A e_i = \sum_{i=1}^{n-1} a_i \lambda_i e_i. \]  

(1.10.3)
We use the expansion (1.10.2) in (1.10.3) to find
\[
\sum_{i=1}^{n-1} a_i e_i \lambda_i = \sum_{i=1}^{n-1} a_i \lambda_i e_i , \tag{1.10.4}
\]
or
\[
\sum_{i=1}^{n-1} a_i e_i (\lambda_n - \lambda_i) = 0. \tag{1.10.5}
\]
Thus if the \( \lambda_i \) are all distinct, and since not all the \( a_i \) are zero, Eq. (1.10.5) cannot be true. Consequently, the assumption (1.10.2) is invalid.

Even though the eigenvectors are linearly independent, we cannot assume they are orthogonal. In fact they cannot be. Of course, the eigenvectors may be normalized. A nonsymmetric matrix with distinct eigenvalues may be diagonalized by using the modal matrix constructed from the eigenvectors of the matrix. However, since the eigenvectors are not generally orthogonal, the diagonalization is accomplished by a similarity transformation which is not an orthogonal transformation in general.

If the eigenvalues of a nonsymmetric matrix are not all distinct, it may not be possible to find a complete set of eigenvectors, but, nevertheless, it is always possible to find a complete set of some other vectors, called principal vectors, which permit some simplification of the original matrix (see Reference 6, pp. 32–36). Let the matrix \( A \) have \( n \) roots, \( \lambda_i \), where some of the roots are repeated. Let \( \lambda_1 \) be repeated \( k \) times. We can always find one eigenvector \( e_1 \) such that
\[
A e_1 = \lambda_1 e_1 . \tag{1.10.6}
\]
We assume this is the only eigenvector associated with \( \lambda_1 \). We now seek a vector \( t_1 \) satisfying
\[
(A - \lambda_1 I) t_1 = e_1 . \tag{1.10.7}
\]
Any solution of Eq. (1.10.7) may be chosen orthogonal to \( e_1 \) (see Problem 17). Let us assume we have found the vector \( t_1 \). We then seek another vector \( t_2 \) from the relation
\[
(A - \lambda_1 I) t_2 = t_1 . \tag{1.10.8}
\]
This implies that \( t_2 \) must be orthogonal to \( t_1 \) and furthermore, from Eq. (1.10.7) we have
\[
(A - \lambda_1 I) t_2 = e_1 . \tag{1.10.9}
\]
Consequently, \( t_2 \) may also be chosen orthogonal to \( e_1 \). We continue generating vectors in sequence of the form

\[
(A - \lambda_1 I)t_p = t_{p-1}.
\]  

(1.10.10)

Each new vector \( t_p \) will be orthogonal to \( t_{p-1}, t_{p-2}, \ldots, e_1 \). It can be shown that we can only find \( k - 1 \) vectors \( t_p \) in this manner (see below).

Let us assume for the moment that \( \lambda_1 \) is the only repeated root of \( A \). Thus, for the \( n - k \) remaining distinct roots \( \lambda_i \), we have \( n - k \) linearly independent eigenvectors \( e_i \). We assert that the set of vectors \( e_1, t_1, t_2, \ldots, t_{k-1}, e_{k+1}, e_{k+2}, \ldots, e_n \) are linearly independent. This follows from the fact that if any of the \( t_j \) were linear combinations of the \( e_i, i \neq 1 \), then we would have

\[
t_j = \sum_{i=k+1}^n a_i e_i.
\]

But from the definition of \( t_j \), we also have

\[
e_1 = (A - \lambda_1 I)t_j = \sum_{i=k+1}^n a_i(A - \lambda_1 I)e_i,
\]

(1.10.12)

which implies \( e_1 \) is a linear combination of the \( e_i \). Therefore, the \( t_j \) are linearly independent of the \( e_i, i \neq 1 \). Since the set \( e_1, t_1, t_2, \ldots, t_{k-1} \) are orthogonal, they are linearly independent of each other. Thus, the sets \( e_i, t_j \) are linearly independent of each other and constitute a basis. Since we have \( n - k + 1 \) eigenvectors, we see that we cannot find more than \( k - 1 \) independent principal vectors, hence the statement in the preceding paragraph.

The particular advantage of the set of vectors so chosen can be seen by constructing the modal matrix. We again define \( M \) as

\[
M = \begin{bmatrix}
e_{11} & t_{11} & t_{21} & \cdots & e_{n1} \\
e_{12} & t_{12} & t_{22} & \cdots & e_{n2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
e_{1n} & t_{1n} & t_{2n} & \cdots & e_{nn}
\end{bmatrix}.
\]

(1.10.13)

We operate on \( M \) by the matrix \( A \). The first column of the product is merely \( \lambda_1 e_1 \). Similarly, all the eigenvectors are reproduced times their
corresponding eigenvalue. Now consider the second column of the product. This is merely

$$A t_1 = e_1 + \lambda_1 t_1$$  \hspace{1cm} (1.10.14)

For the third column, we have

$$A t_2 = t_1 + \lambda_1 t_2$$  \hspace{1cm} (1.10.15)

and so forth, for the \(k-1\) principal vectors. The matrix formed from the product \(AM\) is then seen to be of the form

$$AM = \begin{bmatrix}
\lambda e_{11} + \lambda_1 t_{11} & t_{11} + \lambda_1 t_{21} & \cdots & \lambda_n e_{n1} \\
\lambda e_{12} + \lambda_1 t_{12} & t_{12} + \lambda_1 t_{22} & \cdots & \lambda_n e_{n2} \\
\vdots & \vdots & \ddots & \vdots \\
\lambda e_{1n} + \lambda_1 t_{1n} & t_{1n} + \lambda_1 t_{2n} & \cdots & \lambda_n e_{nn}
\end{bmatrix}.$$  \hspace{1cm} (1.10.16)

This product may be factored in the form

$$AM = \begin{bmatrix}
e_{11} & t_{11} & \cdots & e_{n1} \\
e_{12} & t_{12} & \cdots & e_{n2} \\
\vdots & \vdots & \ddots & \vdots \\
e_{1n} & t_{1n} & \cdots & e_{nn}
\end{bmatrix} \begin{bmatrix}
\lambda_1 & 1 & 0 & \cdots & 0 \\
0 & \lambda_1 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & \cdots & \cdots & \cdots & \lambda_1 \\
0 & \cdots & \cdots & \cdots & \lambda_k + 1 \\
\end{bmatrix}.$$  \hspace{1cm} (1.10.17)

where the submatrix in \(\lambda_1\) is \(k\) by \(k\). The product is of the form \(MJ\), where

$$J = \begin{bmatrix}
\lambda_1 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \lambda_1 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \lambda_1 & 1 & \cdots & 0 & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & \lambda_1 & 1 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 & \lambda_1 & 0 & 0 & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & 0 & 0 & \lambda_{k+1} & 0 & \cdots \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \lambda_{k+2} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \lambda_n \\
\end{bmatrix}.$$  \hspace{1cm} (1.10.18)
The similarity transformation

\[ M^{-1}AM = J \quad (1.10.19) \]
yields a nearly diagonal matrix \( J \), which is called the Jordan canonical form. Note that the matrix has the form of a diagonal matrix for the eigenvectors, while the submatrix for the repeated root contains the eigenvalue \( \lambda_1 \) along the diagonal and the element unity along the upper subdiagonal.

In the case of a repeated root with more than one eigenvector, the submatrix has a form similar to that below.

\[
J_{11} = \begin{bmatrix}
\lambda_1 & 1 & 0 & \ldots & 0 & 0 \\
0 & \lambda_1 & 1 & \ldots & 0 & 0 \\
0 & 0 & \lambda_1 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & \lambda_1 & 1 \\
0 & 0 & 0 & \ldots & 0 & \lambda_1
\end{bmatrix} \quad (1.10.20)
\]

For more than one repeated root, the Jordan canonical form (also called normal form) is

\[
J = \begin{bmatrix}
J_{11} & 0 & \ldots & 0 \\
0 & J_{22} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & J_{pp}
\end{bmatrix}, \quad (1.10.21)
\]

where each of the \( J_{ii} \) is a submatrix in canonical form (for repeated roots) or diagonal (for distinct roots).

It is important to realize that any real matrix may be reduced to the canonical form as above. If there is a complete set of eigenvectors, the canonical form is diagonal. Otherwise, some of the submatrices contain off-diagonal elements. We shall find this result leads to considerable simplification of the analysis of later problems.

If the eigenvalues of a matrix are all greater than zero, the matrix is said to be positive definite. Conversely, a matrix all of whose eigenvalues are less than zero is said to be negative definite.

1.11 Geometric Interpretation

The eigenvalue problem can be given a very useful and illustrative setting in terms of the geometry of quadratic surfaces. We first
consider the equation of a quadratic surface in \( n \)-dimensional space

\[
\frac{x_1^2}{d_1} + \frac{x_2^2}{d_2} + \ldots + \frac{x_n^2}{d_n} = 1. \tag{1.11.1}
\]

We note that the equation can be written in matrix form as

\[
x^T D x = 1, \tag{1.11.2}
\]

where

\[
x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{1.11.3}
\]

and

\[
D = \begin{bmatrix} 1/d_1 & 0 & \ldots & 0 \\ 0 & 1/d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1/d_n \end{bmatrix}. \tag{1.11.4}
\]

This result suggests that there is some intimate relation between the quadratic equation (1.11.1) and the diagonalization of matrices. To see this relation, consider a real symmetric matrix \( A \) and the quadratic form \( x^T A x \). The quadratic form can be written

\[
x^T A x = a_{11} x_1^2 + a_{12} x_1 x_2 + \ldots + a_{1n} x_1 x_n \\
+ a_{22} x_2^2 + a_{23} x_2 x_3 + \ldots + a_{2n} x_2 x_n \\
+ \ldots \\
+ a_{nn} x_n^2.
\]

(1.11.5)

If we set \( x^T A x = 1 \), then the equation represents a general second-order surface. The normal, \( N \), to the surface

\[
f(x_1, x_2, \ldots, x_n) = 1 \tag{1.11.6}
\]
is given by\(^5\)

\[
N = \begin{bmatrix}
\frac{\partial f}{\partial x_1} \\
\frac{\partial f}{\partial x_2} \\
\vdots \\
\frac{\partial f}{\partial x_n}
\end{bmatrix}.
\] (1.11.7)

The normal to the surface \(x^T A x\) is thus

\[
N = 2A x,
\] (1.11.8)

which follows from the symmetry of \(A\).

The principal axes of a quadratic surface are defined as the directions at which the normal vector is parallel to the radius vector. Thus, a principal \(x\) axis is a direction such that

\[
\beta x = N,
\] (1.11.9)

where \(\beta\) is some constant. Consequently, the principal axes satisfy the equation

\[
A x = \lambda x.
\] (1.11.10)

The principal axes are particularly useful since the equation of the quadratic surface expressed in terms of the principal axes contains only a sum of squares. The eigenvectors of the matrix \(A\) are seen to be just the principal axes of the quadratic surface. If we transform the matrix \(A\) by the modal matrix, say \(M\), then we find

\[
A' = M^{-1} A M = \Lambda,
\] (1.11.11)

\(^5\) This relation may be proved by noting that Eq. (1.11.6) implies that

\[
\sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} = 0,
\]

where the \(x_i\) are assumed to be functions of some parameter \(t\). Since the tangent to the surface is proportional to the vector \([dx_i/dt]\), the normal must be proportional to the vector \([\partial f/\partial x_i]\).
and the quadratic form

\[ x^T A x = x' \Lambda x' = 1, \]

which is just the form of Eq. (1.11.2). Notice that the expanded equation is

\[ x^T A x = \lambda_1 x_1'^2 + \lambda_2 x_2'^2 + \ldots + \lambda_n x_n'^2 = 1. \]

The eigenvalues are equal to the reciprocal of the square of the length of the principal axes.

The occurrence of repeated roots can be interpreted in this geometric view. If two roots are equal, then the quadratic surface has rotational symmetry about the axes orthogonal to the eigenvectors of the repeated root. A zero root implies the quadratic surface lies in a space orthogonal to the given direction.

1.12 Biorthogonal Vectors

For a real symmetric matrix, we have shown that the eigenvectors form a set of mutually orthogonal vectors. The eigenvectors are a convenient basis for the space of the problem. In the case of a nonsymmetric matrix, the eigenvectors may not be mutually orthogonal however. It is convenient, in this case, to generate a second set of vectors which are not orthogonal amongst themselves, but are orthogonal with respect to the original set of vectors. Such relationships are known as biorthogonality relationships.

The importance of such relationships can be seen from the following simple example. Consider a vector \( x \) in two-dimensional space, as shown in Fig. 1.12.1.
We write \( \mathbf{x} \) as the matrix

\[
\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.
\]  

(1.12.1)

The square of the length of \( \mathbf{x} \) is

\[
\mathbf{x}^T \mathbf{x} = x_1^2 + x_2^2.
\]  

(1.12.2)

Now consider the same vector in a nonorthogonal coordinate system of base vectors, \( \mathbf{u}_1, \mathbf{u}_2 \), as shown in Fig. 1.12.2 as an example.

![Vector in nonorthogonal coordinates](image)

**Fig. 1.12.2.** Vector \( \mathbf{x} \) in the nonorthogonal coordinate system \( \mathbf{u}_1, \mathbf{u}_2 \).

We shall assume \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \) are related to \( \mathbf{i}, \mathbf{j} \) as

\[
\mathbf{u}_1 = \mathbf{i},
\]  

(1.12.3)

\[
\mathbf{u}_2 = (-1/\sqrt{2})\mathbf{i} + (1/\sqrt{2})\mathbf{j}.
\]  

(1.12.4)

The vector \( \mathbf{x} \) may be written as

\[
\mathbf{x} = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix},
\]  

(1.12.5)

when the components of \( \mathbf{x} \) are referred to the \( \mathbf{u}_1, \mathbf{u}_2 \) basis. To find the components \( x_1', x_2' \), we take projections of \( \mathbf{x} \) parallel to the \( \mathbf{u}_1, \mathbf{u}_2 \) axes. We have

\[
x_1' = x_1 + x_2,
\]  

(1.12.6)

\[
x_2' = \sqrt{2} x_2.
\]  

(1.12.7)

If we consider the length squared of \( \mathbf{x} \) in this coordinate system as

\[
\mathbf{x}^T \mathbf{x} = (x_1')^2 + (x_2')^2 = x_1^2 + 2x_1x_2 + x_2^2 + 2x_2^2,
\]  

(1.12.8)
the result obviously does not agree with the earlier results. The difficulty is that the length is not given by $x^2 x$ when the components are expressed in nonorthogonal coordinates. In order to find a simple expression for the length of a vector, it is necessary to introduce another coordinate system which is said to be biorthogonal to the $u_1, u_2$ system. Thus, we desire a system, say $w_1, w_2$, with the property

$$w_m^T u_n = \delta_{mn}. \quad (1.12.9)$$

Obviously the system

$$w_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad (1.12.10)$$

$$w_2 = \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix}, \quad (1.12.11)$$

is the desired one, where the components of $w_1, w_2$ are expressed in the $i, j$ system. Notice that the biorthogonality condition (1.12.9) is also a statement relating the normalization of both sets of vectors. Although the lengths of $u_1, u_2$ are unity, the lengths of $w_1, w_2$ are both $\sqrt{2}$. The vector $x$ in the $w$ system is shown in Fig. 1.12.3.

![Fig. 1.12.3. Vector $x$ in the nonorthogonal coordinate system $w_1, w_2$.](image)

The vector $x$ may be written

$$x = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix}, \quad (1.12.12)$$

when the components are referred to the $w_1, w_2$ system. The components are found to be

$$x_1' = \sqrt{2}x_1(1/\sqrt{2}) = x_1, \quad (1.12.13)$$

$$x_2' = (-x_1 + x_2)(1/\sqrt{2}) = -\frac{x_1}{\sqrt{2}} + \frac{x_2}{\sqrt{2}}. \quad (1.12.14)$$
The component of a vector $\mathbf{x}$ along some basis vector, such as $\mathbf{w}_1$ or $\mathbf{w}_2$, is the length of the vector $\mathbf{x}$ when projected on that basis vector. The length is the number of units of the basis vector contained in the component of the projection of the vector $\mathbf{x}$. Consequently, since the basis vector, $\mathbf{w}_1$ or $\mathbf{w}_2$, is not of the same length as the original basis vectors, $\mathbf{i}$ or $\mathbf{j}$, each component $x_1''$ and $x_2''$ must be rescaled according to the ratio of the lengths of the final and original basis vectors. In the present example, the rescaling happens to be identical for each term or component. The length squared

$$(x_1'')^2 + (x_2'')^2 = x_1'^2 + \frac{x_1^2}{2} + \frac{x_2^2}{2} - x_1x_2,$$  

(1.12.15a)

is again incorrect. However, the product

$$(x_1', x_1'') + (x_2', x_2'') = x_1'^2 + x_1x_2 - x_1x_2 + x_2'^2$$  

(1.12.15b)

is correct. The correct length is computed when we use the components expressed in a nonorthogonal system and the coordinate system biorthogonal thereto.

An algebraic proof of this geometric result is derived by expressing coordinate axes in matrix equations (see Eq. (1.6.2) et seq.). Thus, if\(^6\)

$$\mathbf{x} = [\mathbf{i}, \mathbf{j}] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$  

(1.12.16)

then

$$\mathbf{x}^T \mathbf{x} = [x_1, x_2] \begin{bmatrix} \mathbf{i} \\ \mathbf{j} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [x_1, x_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$  

(1.12.17)

since $\mathbf{i}$, $\mathbf{j}$ are orthogonal. For the $\mathbf{u}_1$, $\mathbf{u}_2$ system, we have

$$\mathbf{x} = [\mathbf{u}_1, \mathbf{u}_2] \begin{bmatrix} x_1' \\ x_2' \end{bmatrix},$$  

(1.12.18)

while for the $\mathbf{w}_1$, $\mathbf{w}_2$ system we have

$$\mathbf{x} = [\mathbf{w}_1, \mathbf{w}_2] \begin{bmatrix} x_1'' \\ x_2'' \end{bmatrix},$$  

(1.12.19)

and hence

$$\mathbf{x}^T \mathbf{x} = [x_1'', x_2''] \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} [\mathbf{u}_1, \mathbf{u}_2] \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = [x_1'', x_2''] \begin{bmatrix} x_1' \\ x_2' \end{bmatrix},$$  

(1.12.20)

\(^6\) Note here $\mathbf{i}$ and $\mathbf{j}$ are regarded as submatrices of the matrix $[\mathbf{i}, \mathbf{j}]$. 
since $w_1$, $w_2$ and $u_1$, $u_2$ are biorthogonal. Note further that

$$x^Tx = [x_1, x_2] \begin{bmatrix} i & (i, u_1) \\ j & (j, u_1) \end{bmatrix} \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix} = [x_1, x_2] (i, u_2) \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix} \neq x_1x'_1 + x_2x'_2. \tag{1.12.22}$$

The square matrix in Eq. (1.12.22) is known as the metric tensor. One invariably prefers to use orthogonal systems, or at worst biorthogonal systems, since then the metric is thus unity.

The procedure for constructing a dual system to a set of nonorthogonal eigenvectors is relatively simple. Let the matrix $A$ be nonsymmetric but with distinct roots. Denote the eigenvalues as $\lambda_i$ and the corresponding eigenvectors as $e_i$. The eigenvalues satisfy the determinant

$$|A - \lambda I| = 0. \tag{1.12.23}$$

We now consider the transpose matrix $A^T$. The eigenvalues of $A^T$ are obviously the same $\lambda$'s as above, since interchanging rows and columns does not change the value of the determinant. Let the eigenvectors of the transpose matrix be denoted as $U_j$. We then have

$$Ae_i = \lambda_i e_i, \quad A^Tu_j = \lambda_j u_j, \quad (i \neq j). \tag{1.12.24}$$

We now multiply the first equation by $u_j^T$, the second by $e_i^T$ and subtract. We have

$$u_j^T Ae_i - e_i^T A^Tu_j = 0 = (\lambda_i - \lambda_j)u_j^T e_i. \tag{1.12.26}$$

Since the eigenvalues are not equal, the eigenvectors must be orthogonal. Thus the original set of eigenvectors and the transposed set (sometimes called the adjoint eigenvectors) form a biorthogonal system.

This result may immediately be interpreted in terms of quadratic forms. A vector $x$ has a representation in a coordinate system, say $u_i$, as

$$x = x_1u_1 + x_2u_2 + ... + x_nu_n. \tag{1.12.27}$$

The vector may also be represented in the dual coordinate system, say $w_i$, as

$$x = x'_1w_1 + x'_2w_2 + ... + x'_nw_n. \tag{1.12.28}$$
A general quadratic surface in these coordinates would be represented as

\[(x)^{T}Ax = f(x', x) = \text{constant.} \quad (1.12.29)\]

The expanded form of Eq. (1.12.29) is known as a bilinear form rather than a quadratic form. The normal to the surface is again given by

\[N = \frac{\partial f}{\partial x'}, \quad (1.12.30)\]

and hence the principal axes \(x\) are given by

\[Ax = \lambda x. \quad (1.12.31)\]

The dual problem is then

\[x^{T}A^{T}x = f(x', x) = \text{constant,} \quad (1.12.32)\]

with principal axes \(x\) given by

\[A^{T}x = \lambda x, \quad (1.12.33)\]

since the eigenvalues of the transpose matrix \(A^{T}\) equal the eigenvalues of the matrix \(A\). The eigenvectors of the matrix operator are the principal axes of the associated quadratic form. The principal axes of the surface are skewed in general. Nevertheless, if the eigenvectors are complete, the surface can be transformed to the form

\[\lambda_{1}x_{1}^{2} + \lambda_{2}x_{2}^{2} + \ldots + \lambda_{n}x_{n}^{2} = \text{constant} \quad (1.12.34)\]

and

\[\lambda_{1}(x_{1}')^{2} + \lambda_{2}(x_{2}')^{2} + \ldots + \lambda_{n}(x_{n}')^{2} = \text{constant.} \quad (1.12.35)\]

In this case if a root is repeated, we may not be able to assume rotational symmetry. Instead the two eigenvectors may collapse into only one vector, since there is no orthogonality relationship between eigenvectors of a given skewed system.

1.13 Nonnegative Matrices

Of particular usefulness in the numerical solution of differential equations is the theory of nonnegative matrices. In this section we define
several matrix properties and relate these properties to nonnegative matrices.

Frequently one is interested in estimating the largest eigenvalue of a matrix without actually solving the secular equation. A useful theorem is the Gerschgorin theorem which states that the largest eigenvalue is equal to or less than the maximum value of the sum of the magnitudes of the elements in any row. That is, if $A = [a_{ij}]$

$$|\lambda_{\max}| \leq \max_i \sum_j |a_{ij}|.$$ (1.13.1)

The proof of this theorem is simple. Let $\lambda$ be any eigenvalue of $A$ and $e$ the corresponding eigenvector. We then have

$$\lambda e_i = \sum_j a_{ij} e_j,$$ (1.13.2)

which is true for all $i$. Now choose the element of $e$ of largest amplitude, say $e_k$. Then we have

$$|\lambda| \leq \sum_j |a_{kj}| \left| \frac{e_j}{e_k} \right| \leq \sum_j |a_{kj}|.$$ (1.13.3)

Consequently, the largest eigenvalue is bounded by Eq. (1.13.1).

Frequently the largest eigenvalue is called the spectral radius of a matrix, since all eigenvalues lie within or on a circle of radius $\lambda_{\max}$ in the complex plane. We shall denote the spectral radius of $A$ as $r(A)$. Gerschgorin's theorem is then

$$r(A) \leq \max_i \sum_j |a_{ij}|.$$ (1.13.4)

Any matrix $A$ is said to be reducible if there exists a permutation transformation $P$, i.e., if the rows and columns can be permuted similarly, such that

$$PAP^T = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$ (1.13.5)

where the submatrices $A_{11}, A_{22}$ are square, but not necessarily of the same order. If no permutation transformation exists such that (1.13.5) is true, then $A$ is called irreducible. The property of irreducibility implies a connectedness in the problem as seen by the following example. Consider a vector $x$ and a reducible matrix $A$. The product $Ax$ can be written

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 \\ A_{22}x_2 \end{bmatrix}.$$ (1.13.6)
The result indicates that the transformation of the components of $x_2$ is independent of the components of $x_1$. The solution of the equation

$$Ax = y$$

(1.13.7)

can be accomplished as two separate problems

$$A_{11}x_1 + A_{12}x_2 = y_1,$$  \hspace{1cm} (1.13.8a)

$$A_{22}x_2 = y_2.$$  \hspace{1cm} (1.13.8b)

The values of $x_2$ are independent of $x_1$. Physically this implies that some portion of the solution is independent of certain other values of the solution. Such a case arises in multigroup approximations where the fast flux in the core is "disconnected" from the thermal flux in the reflector. On the other hand, if the matrix $A$ is irreducible, then the components of the solution of Eq. (1.13.7) are related to and dependent upon one another.

A nonnegative matrix $A$ is a matrix such that

$$A = [a_{ij}],$$

(1.13.9)

and

$$a_{ij} \geq 0, \hspace{1cm} \text{all } i, j.$$  \hspace{1cm} (1.13.10)

We denote a nonnegative matrix $A$ as $A \geq 0$. Similarly, if

$$a_{ij} > 0, \hspace{1cm} \text{all } i, j,$$  \hspace{1cm} (1.13.11)

then $A$ is called a positive matrix denoted $A > 0$. A very useful theorem regarding nonnegative matrices is the following. If $A$ is nonnegative, then $A$ has a nonnegative real eigenvalue, and the corresponding eigenvector has nonnegative components, not all zero. The proof of the theorem is involved (see Reference 7, pp. 66–68), and we offer a heuristic justification instead. Since $A$ is nonnegative, the quadratic form associated with $A$ represents an ellipsoid and must have a principal axis somewhere in the first quadrant. Since $A$ is nonnegative, any vector with nonnegative components is transformed by $A$ into a nonnegative vector, hence the eigenvalue is nonnegative.

A sharpened form of the above theorem is the following\(^7\): if $A$ is a nonnegative irreducible matrix, then $A$ has a positive real eigenvalue, and the corresponding eigenvector has positive components. To prove

\(^7\) From Reference 8. Some further results in this section are also from Reference 8, Chapter II.
this we note first that $A$ has an eigenvector $x \geq 0$, $x \neq 0$ by the previous theorem. If the corresponding eigenvalue is zero, then we have

$$Ax = \lambda x = 0.$$  \hspace{1cm} (1.13.12)

Since $x \neq 0$, then $A$ must have at least one column identically zero, which implies $A$ is reducible, contrary to hypothesis. Therefore, $\lambda \neq 0$. Conversely, if the eigenvector has some zero components, then we have, after a permutation of rows of $x$ and corresponding rows and columns of $A$,

$$x = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$$  \hspace{1cm} (1.13.13)

and

$$Ax = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ 0 \end{bmatrix} = \begin{bmatrix} A_{11}x_1 \\ A_{21}x_1 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ 0 \end{bmatrix}.$$  \hspace{1cm} (1.13.14)

But then $A_{21} = 0$ and again $A$ is reducible contrary to the hypothesis. Therefore, $x > 0$.

The above result is contained in a classical theorem by Perron and Frobenius which can be stated: If $A$ is a nonnegative irreducible matrix, then $A$ has a positive simple real eigenvalue $\lambda_0$ equal to the spectral radius of $A$. The corresponding eigenvector has all positive components.

To prove that $\lambda_0$ equals the spectral radius of $A$, we consider the matrix $B$ with $0 \leq B$, and $0 \leq b_{ij} \leq a_{ij}$, all $i$, $j$. Thus every element of $B$ is nonnegative and equal to or less than the corresponding element of $A$. We denote the relationship as $0 \leq B \leq A$. We have

$$Ax = \lambda_0 x,$$  \hspace{1cm} (1.13.15)

where $x$ has positive components. Similarly,

$$A^Ty = \lambda_0 y,$$  \hspace{1cm} (1.13.16)

where $y$ has positive components. Now let

$$Bz = \gamma z,$$  \hspace{1cm} (1.13.17)

where $\gamma$ is any eigenvalue of $B$. We now show $\gamma < \lambda_0$ for $B < A$ and $\gamma = \lambda_0$ for $B = A$, which proves $\lambda_0$ equals the spectral radius.

From Eq. (1.13.17) we have

$$\gamma z_i = \sum_j b_{ij}z_j,$$  \hspace{1cm} (1.13.18)
and
\[ |\gamma| \cdot |z_i| \leq \sum_j b_{ij} |z_j| \leq \sum_j a_{ij} |z_j|, \] (1.13.19)
since all elements of \( A, B \) are nonnegative. We multiply Eq. (1.13.19) by \( y_i \) and sum on \( i \) to obtain
\[ |\gamma| \cdot \sum_i y_i |z_i| \leq \sum_i \sum_j a_{ij} y_i |z_j| = \lambda_0 \sum_i y_i |z_j|, \] (1.13.20)

hence
\[ |\gamma| \leq \lambda_0. \] (1.13.21)

If \( \gamma = \lambda_0 \), then the equality holds in Eq. (1.13.19) and requires that
\[ \gamma |z_i| = \sum_j b_{ij} |z_j| = \sum_j a_{ij} |z_j|, \] (1.13.22)

and then \( B = A \).

To prove that \( \lambda_0 \) is a simple root, we need only show that the determinant
\[ |A - \lambda I| = P(\lambda) \] (1.13.23)
has a zero of multiplicity one when \( \lambda = \lambda_0 \). If any polynomial \( P(\lambda) \) has a repeated root at \( \lambda_0 \)
\[ \left. \frac{dP(\lambda)}{d\lambda} \right|_{\lambda_0} = 0. \] (1.13.24)

From Eq. (1.13.23) we readily see that the derivative of the secular polynomial can be written
\[ \frac{dP(\lambda)}{d\lambda} = -\sum_{i=1}^n |M_{ii} - \lambda I|, \] (1.13.25)

where \( M_{ii} \) is the \( i \)th principal minor of \( A \). From previous results we know \( 0 \leq M_{ii} \leq A \) and hence
\[ -|M_{ii} - \lambda_0 I| > 0 \quad (\text{all } i). \] (1.13.26)

We then have
\[ \frac{dP(\lambda_0)}{d\lambda} > 0 \] (1.13.27)
and hence \( \lambda_0 \) is a simple root.

Thus we have shown that \( \lambda_0 \) equals the spectral radius of \( A \) and further, if any element of \( A \) increases, then the spectral radius increases. Having
established the Perron-Frobenius theorem for nonnegative irreducible matrices, we may immediately sharpen the earlier theorem regarding nonnegative matrices in general. In particular, if $A$ is a nonnegative reducible matrix, then $A$ has a nonnegative real eigenvalue which equals the spectral radius of $A$, and as before the corresponding eigenvector has nonnegative components. To prove that the nonnegative eigenvalue is the spectral radius, we merely write $A$ in reduced form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

and examine the matrices $A_{11}, A_{22}$. If they are also reducible, we continue the reduction until all diagonal submatrices are irreducible or null. If the $A_{ii} = 0$, then all the eigenvalues are zero. If any $A_{ii} \neq 0$ then the largest eigenvalue of the nonzero $A_{ii}$ determines the spectral radius. Also, for two matrices $A, B$ such that $0 \leq B \leq A$, it follows from above that

$$r(B) \leq r(A).$$

We shall have occasion to use these results in Chapters III and IV when we discuss the technique for solving simultaneous equations.

1.14 Special Forms and Matrix Factorization

We now consider a few special matrices of interest in our later work. Consider first the square matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$  \hspace{1cm} (1.14.1)

The matrix $A$ may be factored into the form

$$A = L + D + U$$

with

$$L = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n,n-1} \end{bmatrix}.$$  \hspace{1cm} (1.14.3a)
1.14 SPECIAL FORMS AND MATRIX FACTORIZATION

The matrix $L$ contains elements only beneath the principal diagonal and is called strictly lower triangular. Similarly, $U$, which has elements only above the main diagonal, is called strictly upper triangular. $D$ is obviously diagonal. A matrix of the form $L + D$ is then called lower triangular, whereas a matrix of the form $U + D$ is called upper triangular. Notice that if $B$ is defined as

$$B = L + D,$$  \hspace{1cm} (1.14.4)

then $|B| = |D|$. If $|D| \neq 0$, then $B^{-1}$ exists and is of the form

$$B^{-1} = L' + D'.$$

That is, if $B$ lower triangular and if $B^{-1}$ exists, then $B^{-1}$ is also lower triangular. Similarly, if $C$ upper triangular, and if $C^{-1}$ exists, then $C^{-1}$ is also upper triangular. We apply the terms upper triangular, diagonal, and lower triangular only to matrices whose elements are simple elements, and not submatrices.

Frequently one encounters a matrix of the form

$$A = \begin{bmatrix}
a_{11} & a_{12} & 0 & 0 & \cdots & 0 & 0 \\
a_{21} & a_{22} & a_{23} & 0 & \cdots & 0 & 0 \\
0 & a_{32} & a_{33} & a_{34} & \cdots & 0 & 0 \\
0 & 0 & a_{43} & a_{44} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & a_{n-1,n-1} & a_{n-1,n} \\
0 & 0 & 0 & 0 & \cdots & a_{n,n-1} & a_{n,n}
\end{bmatrix}. \hspace{1cm} (1.14.5)$$

Such a matrix has elements only along the main diagonal and the two nearest adjacent diagonals. A matrix of this form is called a tridiagonal matrix. We shall encounter such matrices in approximating second derivatives of functions.
A generalization of the tridiagonal matrix is a matrix of the form

\[
\begin{bmatrix}
A_{11} & A_{12} & 0 & \cdots & 0 & 0 \\
A_{21} & A_{22} & A_{23} & \cdots & 0 & 0 \\
0 & A_{32} & A_{33} & \cdots & 0 & 0 \\
& \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \cdots & A_{n-1,n-1} & A_{n-1,n} \\
0 & 0 & 0 & \cdots & A_{n,n-1} & A_{nn}
\end{bmatrix},
\]  

(1.14.6)

where the elements \(A_{ij}\) are themselves submatrices. If the submatrices \(A_{ii}\) are tridiagonal and if the submatrices \(A_{i,i\pm1}\) are diagonal, then the matrix \(A\) is called block tridiagonal. Such matrices occur in approximating the Laplacian operator in two dimensions.

The inversion of tridiagonal matrices may be readily accomplished by taking advantage of the large number of zero elements contained in the matrix. Consider the matrix equation

\[
Ax = y,
\]

(1.14.7)

where \(A\) is assumed tridiagonal. We factor \(A\) in the form

\[
A = CB
\]

(1.14.8)

with

\[
C = \begin{bmatrix}
1 & 0 & \cdots & 0 & 0 \\
c_{21} & 1 & \cdots & 0 & 0 \\
0 & c_{32} & \cdots & 0 & 0 \\
& \ddots & \ddots & \ddots & \ddots \\
0 & 0 & \cdots & 1 & 0 \\
0 & 0 & \cdots & c_{n,n-1} & 1
\end{bmatrix},
\]

(1.14.9a)

and

\[
B = \begin{bmatrix}
b_{11} & b_{12} & 0 & \cdots & 0 \\
0 & b_{22} & b_{23} & \cdots & 0 \\
0 & 0 & b_{33} & \cdots & 0 \\
& \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \cdots & b_{n-1,n} \\
0 & 0 & 0 & \cdots & b_{nn}
\end{bmatrix}.
\]

(1.14.9b)
In order for the factorization to be true, we must require

\[ b_{11} = a_{11} \]  
\[ b_{12} = a_{12} \]  
\[ c_{21} b_{11} = a_{21} \]  
\[ c_{21} b_{12} + b_{22} = a_{22} \]  
\[ b_{22} = a_{22} \]  
\[ c_{n,n-1} b_{n-1,n-1} = a_{n,n-1} \]  
\[ c_{n,n-1} b_{n-1,n} + b_{n,n} = a_{n,n} \]  

The above equations may be solved in the order \( b_{11}, b_{12}, c_{21}, b_{22}, b_{23}, \ldots, c_{p,p-1}, b_{pp}, b_{p,p+1}, \ldots, c_{n,n-1}, b_{nn} \). In order for the solution to exist, we must require \(|A| \neq 0\). We now write Eq. (1.14.7) in the form

\[ CBx = y. \]  

We define a vector \( z \) such that

\[ y = Cz. \]  

This definition leads to the requirement

\[ z_1 = y_1, \]  
\[ c_{21} z_1 + z_2 = y_2, \]  
\[ \vdots \]  
\[ c_{p,p-1} z_{p-1} + z_p = y_p, \]  
\[ \vdots \]  
\[ c_{n,n-1} z_{n-1} + z_n = y_n. \]  

Equations (1.14.16) can readily be solved for \( z \). The entire set of equations now becomes

\[ Bx = z. \]
The solution is then given by

\[ b_{11}x_1 + b_{12}x_2 = z_1, \]
\[ b_{22}x_2 + b_{23}x_3 = z_2, \]
\[ \vdots \]
\[ b_{pp}x_p + b_{p,p+1}x_{p+1} = z_p, \]
\[ \vdots \]
\[ b_{nn}x_n = z_n, \]  

(1.14.18)

from which the \( x_i \) are readily found by starting from \( z_n \) and working in sequence back to \( z_1 \).

Let us consider the number of operations\(^8\) involved in this scheme (called matrix factorization) and compare this number with that involved in the Gauss reduction method. To generate the matrices \( B \) and \( C \), we refer to equations (1.14.10, 11, 12, 13). The elements \( b_{11}, b_{12} \) are obtained without algebraic operations. To obtain \( c_{21} \) requires 1 operation (division), \( b_{22} \) requires 2 operations (multiplication and subtraction). Therefore, 3 operations are needed per set of equations. For \( N \) unknowns, a total of \( 3N-3 \approx 3N \) operations are necessary.

To find the vector \( z \), a total of \( 2N \) operations are necessary; likewise, for finding \( x \), \( 3N \) operations are required. A total of \( 8N \) steps are needed to solve the original set of equations.

We now consider solving the same set of equations by the straightforward reduction method. To reduce the first equation of (1.14.7) to the form

\[ x_1 + a'_{12}x_2 = y'_1, \]  

(1.14.19)

requires 2 operations. To eliminate \( x_1 \) from the second equation requires 4 operations. There are, therefore, 6\( N \) steps to reduce the equations to the upper triangular form. An additional \( 2N \) steps are needed in the back substitution to solve. Consequently, a total of \( 9N \) steps are needed. Thus, the factorization and Gauss reduction method involve the same number of operations; indeed, the former is a special case of the latter (see Problem 9). We shall encounter the technique again in Chapter III.

\(^8\) For computing purposes, the important operations are addition and multiplication. For most computing devices, these operations take longer than control operations of various types.
A generalization of the method is applicable to block tridiagonal matrices but requires inversion of the submatrices occurring in Eq. (1.14.6).

References

There are innumerable books devoted to matrix algebra or having chapters concerning matrices. For a very readable discussion, references 1 and 2 are particularly recommended. References dealing with numerical methods for handling matrices (obtaining inverses, finding the eigenvalues and eigenvectors, etc.) include 3, 4, 5, and 6. A very rigorous discussion of matrices, including discussion of the Perron-Frobenius theorem and related topics, is found in 7. An excellent distillation of the important matrix properties useful in the numerical solution of boundary-value problems is reference 8.


Problems

1. Prove the associative law of matrix multiplication, that is

\[(AB)C = A(BC)\].

2. Write out the following matrix products:

(a) \[
\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_n
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{bmatrix}
\]

(b) \[
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{bmatrix}
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix}
\]


3. Prove Theorems 2 and 5 relating to determinants [p. 7].

4. Show that the variable \( x_i \) of a set of \( n \) homogeneous equations of rank \( n - 1 \) is proportional to any one of the cofactors of its coefficients in the coefficient matrix.

5. Prove that for any matrix \( A \) that \( AA^T \) is a square symmetric matrix.

6. Prove that for square matrices \( A \) and \( B \) of order \( n \) that
\[
\text{adj}(AB) = (\text{adj} B)(\text{adj} A).
\]

7. Find the transpose and inverse of the matrix
\[
\begin{bmatrix}
1/2 & 1/\sqrt{2} & -1/2 \\
1/\sqrt{2} & 0 & 1/\sqrt{2} \\
1/2 & -1/\sqrt{2} & -1/2
\end{bmatrix}.
\]

8. Derive the matrix describing the rotation of a vector, in a plane perpendicular to the \( z \) axis, about the \( z \) axis through the angle \( \varphi \). Derive the matrix describing the rotation of a vector through the angle \( \varphi \) about the \( z \) axis and followed by a rotation through the angle \( \theta \) in a plane containing the vector and the \( z \) axis.

9. Consider the solution of the equations \( At = u \), or
\[
\begin{bmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{bmatrix}
\begin{bmatrix}
t_1 \\
t_2 \\
t_3
\end{bmatrix}
= \begin{bmatrix}
1 \\
u_2 \\
u_3
\end{bmatrix}
\]

by Gauss reduction. The reduced set of equations can be written \( A't = u' \), or
\[
\begin{bmatrix}
1 & a'_{12} & a'_{13} \\
0 & 1 & a'_{23} \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
t_1 \\
t_2 \\
t_3
\end{bmatrix}
= \begin{bmatrix}
u'_1 \\
u'_2 \\
u'_3
\end{bmatrix}.
\]

Derive the sequence of matrix operations which when applied successively to \( A \) yield \( A' \), that is, find the transformations \( F \), such that
\[
F_t \ldots F_2 F_1 A t = A't.
\]

10. Modify the Gauss reduction in problem 9 to eliminate all unknowns from the \( i \)th row except \( a_{ii} \), that is derive the reduced matrix in the form
\[
A' = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}.
\]

This modification is sometimes called the Gauss-Jordan reduction.

11. If a set of linearly independent vectors \( u_n \) are orthogonalized by the Schmidt process to yield an orthogonal set \( t_n \), then the sets are related by a transformation in the form
\[
t_n = T u_n.
\]

Derive an expression for the matrix \( T \).
12. Show that, if an array of vectors is linearly dependent, the array formed by transforming each vector similarly is also linearly dependent.

13. Find the eigenvalues and eigenvectors of the matrix

\[
A = \begin{bmatrix} 3 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix}
\]

Diagonalize the above matrix. Express the vector \( x \) (below) in terms of the eigenvectors of \( A \).

\[
x = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}
\]

14. Use the results of the above problem to evaluate \( A^7 x \) where \( A \) and \( x \) are given in problem 13.

15. If \( A \) is a real symmetric matrix, show that solutions of the inhomogeneous equation

\[
A t - \lambda t = u
\]

can be written in the form

\[
t = \sum_n \frac{\alpha_n}{\lambda_n - \lambda} e_n,
\]

where \( \lambda_n, e_n \) are the eigenvalues and eigenvectors of \( A \). (Assume \( \lambda \neq \lambda_n \), all \( n \)). What are the coefficients \( \alpha_n \)?

16. Show that the above problem has no solution if \( \lambda \) equals an eigenvalue of \( A \) unless \( u \) is orthogonal to the eigenvector corresponding to the eigenvalue \( \lambda \).

17. Prove the solution \( t_1 \) of Eq. (1.10.7) may be chosen orthogonal to the vector \( e_1 \).

18. Show that if two matrices \( A \) and \( B \) commute, then \( A \) and \( B \) possess the same eigenvectors, and conversely.

19. Find the eigenvalues and eigenvectors of the following matrix operators

\[
[a] \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\]

\[
[b] \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\]

\[
[c] \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}
\]

20. Show that the inverse of a nonsingular matrix \( A \) can be written

\[
A^{-1} = M \Lambda^{-1} V^T,
\]

where the matrices \( M \) and \( V \) are constructed from the eigenvectors of \( A \) and \( A^T \), respectively, and where \( \Lambda \) is a diagonal matrix whose elements are the eigenvalues of \( A \).
21. Show that the number of operations needed to solve a set of $N$ simultaneous equations in $N$ unknowns by Gaussian reduction is proportional to $N^3$ in general.

22. Solve the matrix differential equation

$$\frac{du(t)}{dt} = Au(t)$$

where

$$A = \begin{bmatrix} 4 & 2 \\ -1 & 1 \end{bmatrix}$$

and $u[0] = i + j$, i.e.,

$$u(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$