F.2.1. Matrix and Vector Algebra

F. L. Markley

Reproduced with permission from *Spacecraft Attitude Determination and Control* [Wertz, 1978], Appendix C.

F.2.1.1 Definitions

A *matrix* is a rectangular array of scalar entries known as the *elements* of the matrix. In this book, the scalars are assumed to be real or complex numbers. If all the elements of a matrix are real numbers, the matrix is a *real matrix.* The matrix

$$A \equiv \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} \equiv \left[A_{ij} \right] \tag{Fweb-1}$$

has m rows and n columns, and is referred to as an $m \times n$ *matrix* or as a *matrix of order* $m \times n$. The equation $A = [A_{ij}]$ should be read as, "A is the matrix whose elements are A_{ij}." The first subscript labels the rows of the matrix and the second labels the columns.

Two matrices are equal if and only if they are of the same order and all of the corresponding elements are equal; i.e.,

$$A = B \text{ if and only if } A_{ij} = B_{ij};\ i = 1,\ldots,m;\ j = 1,\ldots,n \tag{Fweb-2}$$

An $n \times n$ matrix is called *a square matrix* and is usually referred to as being of order n rather than $n \times n$.

The *transpose* of a matrix is the matrix resulting from interchanging rows and columns. The transpose of A is denoted by $A^{\mathbf{T}}$, and its elements are given by

$$A^{\mathbf{T}} \equiv [(A^{\mathbf{T}})_{ij}] \equiv [A_{ji}] \tag{Fweb-3}$$

As an example, the transpose of the matrix in Eq. (Fweb-1) is

$$A^{\mathbf{T}} \equiv \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{m1} \\ A_{12} & A_{22} & \cdots & A_{m2} \\ \vdots & \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{mn} \end{bmatrix}$$

It is clear that the transpose of an $m \times n$ matrix is an $n \times m$ matrix, and that the transpose of a square matrix is square. The transpose of the transpose of a matrix is equal to the original matrix:

$$\left(A^{\mathrm{T}}\right)^{\mathrm{T}} = A \tag{Fweb-4}$$

The *adjoint* of a matrix, denoted by $A^{\dagger}$, is the matrix whose elements are the complex conjugates of the elements of the transpose of the given matrix*, i.e.,

$$A^{\dagger} \equiv [(A^{\dagger})_{ij}] \equiv [A^*_{ji}] \tag{Fweb-5}$$

The adjoint of the adjoint of a matrix is equal to the original matrix:

$$(A^{\dagger})^{\dagger} = A \tag{Fweb-6}$$

The adjoint and the transpose of a real matrix are identical.

The *main diagonal* of a square matrix is the set of elements with row and column indices equal. A *diagonal matrix* is a square matrix with nonzero elements only on the main diagonal, e.g.,

$$D = \begin{bmatrix} D_{11} & 0 & \dots & 0 \\ 0 & D_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & D_{nn} \end{bmatrix} \tag{Fweb-7}$$

The *identity matrix* of a given order is the diagonal matrix with all the elements on the main diagonal equal to unity. It is denoted by $\mathbf{1}$, or by $\mathbf{1}_n$ to indicate the order explicitly.

A matrix with only one column is a *column matrix*. An $n \times 1$ column matrix can be identified with a vector in n-dimensional space, and we shall indicate such matrices by boldface letters, as used for vectors.† A matrix with only one row is a *row matrix*; its transpose is a column matrix, so we denote it as the transpose of a vector. The elements of a row or column matrix will be written with only one subscript; for example,

$$\mathbf{B} = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix}, \qquad \mathbf{C}^{\mathrm{T}} \equiv \left[C_1,\ C_2, \cdots, C_m \right] \tag{Fweb-8}$$

A set of m $n \times 1$ vectors $\mathbf{B}^{(i)}$, $i = 1,2,...,m$, is *linearly independent* if and only if the only coefficients a_i, $i = 1,2,...,m$, satisfying the equation

$$\sum_{i=1}^{m} a_i \mathbf{B}^{(i)} = a_1 \mathbf{B}^{(1)} + a_2 \mathbf{B}^{(2)} + \cdots + a_m \mathbf{B}^{(m)} = \mathbf{0} \tag{Fweb-9}$$

are $a_i = 0$, $i = 1,2,...,m$. There can never be more than n linearly independent $n \times 1$ vectors.

F.2.1.2 Matrix Algebra

Multiplication of a matrix by a scalar is accomplished by multiplying each element of the matrix by the scalar, i.e.,

---

* The word adjoint is sometimes used for a different matrix in the literature.

† Strictly speaking, a vector is an abstract mathematical object, and the column matrix is a concrete realization of it, the matrix elements being the components of the vector in some coordinate system.

$$sA \equiv [sA_{ij}] \tag{Fweb-10}$$

Addition of two matrices is possible only if the matrices have the same order. The elements of the matrix sum are the sums of the corresponding elements of the matrix addends, i.e.,

$$A + B \equiv [A_{ij} + B_{ij}] \tag{Fweb-11}$$

Matrix subtraction follows from the above two rules by

$$A - B \equiv A + (-1)B = [A_{ij} - B_{ij}] \tag{Fweb-12}$$

Multiplication of two matrices is possible only if the number of columns of the matrix on the left side of the product is equal to the number of rows of the matrix on the right. If A is of order $1 \times m$ and B is $m \times n$, the product AB is the $l \times$ n matrix given by

$$AB \equiv \left[(AB)_{ij} \right] \equiv \left[\sum_{k=1}^{m} A_{ik} \ B_{kj} \right] \tag{Fweb-13}$$

Matrix multiplication is *associative*

$$A(BC) = (AB)C \tag{Fweb-14}$$

and *distributive over addition*

$$A(B + C) = AB + AC \tag{Fweb-15}$$

but is ***not*** *commutative*, in general,

$$AB \neq BA \tag{Fweb-16}$$

In fact, the products AB and BA are both defined and have the same order only if A and B are square matrices, and even in this case the products are not necessarily equal. For the square matrices

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix},$$

for example, we have

$$AB = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix} \neq BA = \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix}$$

If $AB = BA$, for two square matrices, A and B, we say that A and B *commute*. One interesting case is diagonal matrices, which always commute.

The adjoint (or transpose) of the product of two matrices is equal to the product of the adjoints (or transposes) of the two matrices taken in the opposite order:

$$(AB)^{\dagger} = B^{\dagger}A^{\dagger} \tag{Fweb-17}$$

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}} \tag{Fweb-18}$$

This result easily generalizes to products of more than two matrices.

Multiplying any matrix by the identity matrix of the appropriate order, on the left or the right, yields a product equal to the original matrix. Thus, if B is of order $m \times n$,

$$\mathbf{1}_m B = B\mathbf{1}_n = B \tag{Fweb-19}$$

The product of an $n \times m$ matrix and an m-dimensional vector (an $m \times 1$ matrix) is an n-dimensional vector; thus,

$$\mathbf{Y} \equiv A\mathbf{X} \equiv \left[\sum_{j=1}^{m} A_{ij} X_j \right] \tag{Fweb-20}$$

A similar result holds if a row matrix is multiplied on the right by a matrix,

$$\mathbf{Y}^{\mathbf{T}} = \mathbf{X}^{\mathbf{T}} A^{\mathbf{T}} = \left[\sum_{j=1}^{m} A_{ji} X_j \right] \tag{Fweb-21}$$

An important special case of the above is the multiplication of a $1 \times n$ row matrix (on the left) by an $n \times 1$ column matrix (on the right) which yields a scalar,

$$s \equiv \mathbf{Y}^{\mathbf{T}}\mathbf{X} \equiv \sum_{i=1}^{n} X_j Y_j \tag{Fweb-22}$$

For real vectors, this scalar is the *inner product*, or *dot product*, or *scalar product* of the vectors **X** and **Y**. For vectors with complex components, it is more convenient to define the inner product by using the adjoint of the left-hand vector rather than the transpose. Thus, in general,

$$\mathbf{Y} \cdot \mathbf{X} \equiv \mathbf{Y}\ \mathbf{X} = \sum_{i=1}^{n} Y_i^* X_i \tag{Fweb-23}$$

Note that, in general,

$$\mathbf{Y} \cdot \mathbf{X} = (\mathbf{X} \cdot \mathbf{Y})^* \tag{Fweb-24}$$

This definition reduces to the usual definition for real vectors, for which the inner product is independent of the order in which the vectors appear. Two vectors are *orthogonal* if their inner product is zero. The inner product of a vector with itself

$$\mathbf{X} \cdot \mathbf{X} = \sum_{i=1}^{n} X_i^* X_i = \sum_{i=1}^{n} |X_i|^2 \tag{Fweb-25}$$

is never negative and is zero if and only if all the elements of **X** are zero, i.e., if $\mathbf{X} = \mathbf{0}$. This product will be denoted by $\mathbf{X}^2$ and its positive square root by $|\mathbf{X}|$ or by X, if no confusion results. The scalar $|\mathbf{X}|$ is called the *norm* or *magnitude* of the vector, **X**, and can be thought of as the length of the vector. Thus, with our definition of the inner product,

$$|\mathbf{X}| = 0 \text{ if and only if } \mathbf{X} = \mathbf{0} \tag{Fweb-26}$$

which would not be true if we defined the inner product using the transpose rather than the adjoint, because the square of a complex number generally is not positive.

If we multiply an $n \times 1$ row matrix (on the left) by a $1 \times m$ matrix (on the right), we obtain an n x m matrix. This leads to the definition of the *outer product* of two vectors

$$\mathbf{X}\mathbf{Y}^{\dagger} \equiv [(\mathbf{X}\mathbf{Y}^{\dagger})_{ij}] \equiv [X_i Y_j^*] \tag{Fweb-27}$$

If the vectors are real, the adjoint of **Y** is the transpose of **Y**, and the *ij*th element of the outer product is $X_i Y_j$.

Matrix division can be defined in terms of matrix inverses, which are discussed in Section F.2.1.4.

F.2.1.3 Trace, Determinant, and Rank

Two useful scalar quantities, the *trace* and the *determinant*, can be defined for any *square* matrix. The *rank* of a matrix is defined for any matrix.

The *trace* of an $n \times n$ matrix is the sum of the diagonal elements of the matrix

$$\mathrm{tr}A \equiv \sum_{i=1}^{n} A_{ii} \tag{Fweb-28}$$

The trace of a product of square matrices is unchanged by a cyclic permutation of the order of the product

$$\mathrm{tr}\left(ABC\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} A_{ij} B_{jk} C_{ki} = \mathrm{tr}\left(CAB\right) \tag{Fweb-29}$$

However, $\mathrm{tr}(ABC) \neq \mathrm{tr}(ACB)$, in general.

The *determinant* of an $n \times n$ matrix is the complex number defined by

$$\det A \equiv \left| A_{ij} \right| \equiv \sum \left(-1\right)^{p} A_{1p_1} A_{2p_2} \cdots A_{np_n} \tag{Fweb-30}$$

where the set of numbers $\{p_1, p_2, \ldots, p_n\}$ is a *permutation*, or rearrangement, of $\{1,2,\ldots,n\}$. Any permutation can be achieved by a sequence of pairwise interchanges. A permutation is uniquely *even* or *odd* if the number of interchanges required is even or odd, respectively. The exponent p in Eq. (Fweb-30) is zero for even permutations and unity for odd ones. The sum is over all the $n!$ distinct permutations of $\{1,2,\ldots,n\}$. It is not difficult to show that Eq. (Fweb-30) is equivalent to

$$\det A = \sum_{j=1}^{n} \left(-1\right)^{i+j} A_{ij} M_{ij} \tag{Fweb-31}$$

for any *fixed* $i = 1,2,\ldots,n$, where M_{ij} is the *minor* of A_{ij}, defined as the determinant of the $(n-1) \times (n-1)$ matrix formed by omitting the ith row and jth column from A. This form provides a convenient method for evaluating determinants by successive reduction to lower orders. For example,

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} &= 1 \times \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \times \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \times \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ &= \left(5 \times 9 - 8 \times 6\right) - 2\left(4 \times 9 - 7 \times 6\right) + 3\left(4 \times 8 - 7 \times 5\right) = 0 \end{aligned} \tag{Fweb-32}$$

The determinant of the product of two square matrices is equal to the product of the determinants

$$\det\,(AB) = (\det A)(\det B) \tag{Fweb-33}$$

The determinant of a scalar multiplied by an $n \times n$ matrix is given by

$$\det(sA) = s^{n} \det A \tag{Fweb-34}$$

The determinants of a matrix and of its transpose are equal:

$$\det A^{\mathrm{T}} = \det A \tag{Fweb-35}$$

Thus, the determinant of the adjoint is

$$\det A^{\dagger} = (\det A)^{*} \tag{Fweb-36}$$

The determinant of a matrix with all zeros on one side of the main diagonal is equal to the product of the diagonal elements.

The *rank* of a matrix is the order of the largest square array in that matrix, formed by deleting rows and columns that has a nonvanishing determinant. Clearly, the rank of an $m \times n$ matrix cannot exceed the smaller of m and n. The matrices A, A^T, $A^\dagger$, $A^\dagger A$, and $AA^\dagger$ all have the same rank.

F.2.1.4 Matrix Inverses and Solutions to Simultaneous Linear Equations

Let A be an $m \times n$ matrix of rank k. An $n \times m$ matrix B is a *left inverse* of A if $BA = \mathbf{1}_n$. An $n \times m$ matrix C is a *right inverse* of A if $AC = \mathbf{1}_m$. There are four possible cases: k is less than both m and n, $k = m = n$, $k = m < n$, and $k = n < m$. If k is less than both m and n, then no left or right inverse of A exists.* If $k = m = n$, then $\boldsymbol{A}$ is *nonsingular* and has a unique *inverse*, A^{-1}, which is both a left and a right inverse:

$$A^{-1}A = AA^{-1} = \mathbf{1} \qquad (k = m = n) \tag{Fweb-37}$$

A nonsingular matrix is a square matrix with nonzero determinant; all other matrices are *singular*. If $k = m < n$, then A has no left inverse but an infinity of right inverses, one of which is given by

$$A^R = A^\dagger (AA^\dagger)^{-1} \qquad (k = m < n) \tag{Fweb-38}$$

If $k = n < m$, then A has no right inverse but an infinity of left inverses, one of which is

$$A^L = (A^\dagger A)^{-1} A^\dagger \qquad (k = n < m) \tag{Fweb-39}$$

A^L or A^R is called the *generalized inverse* or *pseudoinverse* of A.

Consider the set of m simultaneous linear equations in n unknowns; $X_1, X_2, \ldots, X_n$;

$$A\mathbf{X} = \mathbf{Y} \tag{Fweb-40}$$

If $k = m = n$, then $\mathbf{X} = \mathrm{A}^{-1}\,\mathbf{Y}$ is the unique solution to the set of equations. It follows that a nonzero solution to $A\mathbf{X} = \mathbf{0}$ is possible only if A is singular, i.e.,

$$A\mathbf{X} = \mathbf{0} \text{ for } \mathbf{X} \neq \mathbf{0}, \text{ only if } \det A = 0 \tag{Fweb-41}$$

If $k = m < n$, there are more unknowns than equations, so there are an infinite number of solutions. The solution with the smallest norm, $|\mathbf{X}|$, is

$$\mathbf{X} = A^R\mathbf{Y} \tag{Fweb-42}$$

If $k = n < m$, there are more equations than unknowns; therefore, no solution exists, in general. However, the vector $\mathbf{X}$ that comes closest to a solution, in the sense of minimizing $|A\mathbf{X} - \mathbf{Y}|$, is

$$\mathbf{X} = A^L\mathbf{Y} \tag{Fweb-43}$$

Note that although $AA^L \neq \mathbf{1}_m$, it is possible that $AA^L\mathbf{Y} = \mathbf{Y}$ for the particular $\mathbf{Y}$ in Eq. (Fweb-40). In this case, Eq. (Fweb-40) has a unique solution given by Eq. (Fweb-43).

It is easy to see that if A is nonsingular, then A^{-1} is nonsingular also and

$$(A^{-1})^{-1} = A \tag{Fweb-44}$$

Likewise, if A is nonsingular, then A^T and $A^\dagger$ are nonsingular and their inverses are given by

$$(A^T)^{-1} = (A^{-1})^T \tag{Fweb-45}$$

$$(A^\dagger)^{-1} = (A^{-1})^\dagger \tag{Fweb-46}$$

* It is possible to define a *pseudoinverse* for a general matrix, which in this case is neither a left nor a right inverse. In the other three cases, the pseudoinverse is identical with A^{-1}, A^R, and A^L, respectively. The results on solutions of simultaneous linear equations can be generalized with this definition [Wiberg, 1971].

If two matrices, A and B, are nonsingular, their product is nonsingular also; and the inverse of the product is the product of the inverses, taken in the opposite order:

$$(AB)^{-1} = B^{-1}A^{-1} \qquad \text{(Fweb-47)}$$

This result easily generalizes to products of more than two matrices.

Various algorithms exist for calculating matrix inverses; several are described by Carnahan, et al., [1969] and by Forsythe and Moler [1967]. An example of a subroutine for this purpose is INVERT, described in Section 20.3.

F.2.1.5 Special Types of Square Matrices, Matrix Transformations

A *symmetric* matrix is a square matrix that is equal to its transpose:

$$A^{\mathrm{T}} = A, \quad A_{ij} = A_{ji} \qquad \text{(Fweb-48)}$$

A *skew-symmetric* or *antisymmetric* matrix is equal to the negative of its transpose:

$$A^{\mathrm{T}} = -A, \quad A_{ij} = -A_{ji} \qquad \text{(Fweb-49)}$$

Clearly, a skew-symmetric matrix must have zeros on its main diagonal. An example of a skew symmetric matrix is $\mathbf{\Omega}$ in Section 16.1. A *Hermitian* matrix is equal to its adjoint:

$$A^{\dagger} = A, \quad A_{ij} = A^{*}_{ji} \qquad \text{(Fweb-50)}$$

A real symmetric matrix is a special case of a Hermitian matrix. An *orthogonal* matrix is a matrix whose transpose is equal to its inverse:

$$A^{\mathrm{T}} = A^{-1}, \quad AA^{\mathrm{T}} = A^{\mathrm{T}}A = \mathbf{1} \qquad \text{(Fweb-51)}$$

A *unitary* matrix is a matrix whose adjoint is equal to its inverse:

$$A^{\dagger} = A^{-1}, \quad AA^{\dagger} = A^{\dagger}A = \mathbf{1} \qquad \text{(Fweb-52)}$$

A *real orthogonal* matrix is a special case of a unitary matrix. The product of two unitary (or orthogonal) matrices is unitary (or orthogonal). This result generalizes to products of more than two matrices. A similar result generally does **not** hold for Hermitian or symmetric matrices. A *normal* matrix is a matrix that commutes with its adjoint

$$A^{\dagger}A = AA^{\dagger}$$

Thus, both Hermitian matrices and unitary matrices are special cases of normal matrices.

By the rules for determinants of products and adjoints, it is easy to see that if $\boldsymbol{A}$ is unitary

$$|\det A|^2 = 1 \qquad \text{(Fweb-53)}$$

Thus, $\det A$ is a complex number with absolute value unity. Similarly, if A is orthogonal,

$$(\det A)^2 = 1 \quad \det A = \pm 1 \qquad \text{(Fweb-54)}$$

An orthogonal matrix with positive determinant is a *proper* orthogonal matrix; an orthogonal matrix is *improper* if its determinant is negative.

Let $\mathbf{X}$ be an n-dimensional vector and let A be an $n \times n$ matrix. Then $A\mathbf{X}$ is another n-dimensional vector and can be thought of as the transformation of $\mathbf{X}$ by A. If $\mathbf{X}$ and $\mathbf{Y}$ are two vectors, the inner product of $A\mathbf{X}$ and $A\mathbf{Y}$ is

$$\left(A\mathbf{X}\right)\cdot\left(A\mathbf{Y}\right) = \left(A\mathbf{X}\right) \left(A\mathbf{Y}\right) = \mathbf{X} \; A \; A\mathbf{Y}$$

If A is unitary,

$$\left(A\mathbf{X}\right)\cdot\left(A\mathbf{Y}\right) = \mathbf{X}\cdot\mathbf{Y} \qquad \text{(Fweb-55)}$$

The dot product is unchanged if both vectors are transformed by the same unitary matrix. This result with $\mathbf{X} = \mathbf{Y}$ shows that the norm of a vector is unchanged, too, so the unitary matrix can be thought of as performing a *rotation* of the vector in n-dimensional space, thereby preserving its length. If the vectors are real, the rotations correspond to *proper real orthogonal matrices*.

The transformations of a matrix are defined analogously to the transformations of a vector, but they involve multiplying the matrix on both the left and right sides, rather than only on the left side. Several kinds of transformations are defined. If B is nonsingular, then

$$A_S = B^{-1}AB \qquad \text{(Fweb-56)}$$

is a *similarity transformation* on A. We say that A_S is *similar* to A. A special case occurs if B is unitary. In this case we have a *unitary transformation* on A,

$$A_U = B^{\dagger}AB \qquad \text{(Fweb-57)}$$

A second special case occurs if B is orthogonal, in which case

$$A_O = B^{\mathrm{T}}AB \qquad \text{(Fweb-58)}$$

defines an *orthogonal transformation* on A.

It follows directly from the invariance of the trace to cyclic permutations of the order of matrix products, Eq. (Fweb-28), that

$$\mathrm{tr}A_S = \mathrm{tr}A_U = \mathrm{tr}A_O = \mathrm{tr}A \qquad \text{(Fweb-59)}$$

Also, by the rules on determinants,

$$\det A_S = \det A_U = \det A_O = \det A \qquad \text{(Fweb-60)}$$

It is easy to see that

$$A_U = B\ A\ B \qquad \text{(Fweb-61)}$$

and

$$A_O^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}B \qquad \text{(Fweb-62)}$$

Thus, A_U is Hermitian (or unitary) if A is Hermitian (or unitary), and A_O is symmetric (or orthogonal) if A is symmetric (or orthogonal).

F.2.1.6 Eigenvectors and Eigenvalues

If A is an $n \times n$ matrix and if

$$A\mathbf{X} = \lambda\mathbf{X} \qquad \text{(Fweb-63)}$$

for some nonzero vector $\mathbf{X}$ and scalar λ, we say that $\mathbf{X}$ is an *eigenvector* of A and that λ is the corresponding *eigenvalue*. We can rewrite Eq. (Fweb-63) as

$$(A - \lambda\mathbf{1})\,\mathbf{X} = \mathbf{0} \qquad \text{(Fweb-64)}$$

so we see from Eq. (Fweb-41) that λ is an eigenvalue of A if and only if

$$\det(A - \lambda\mathbf{1}) = 0 \qquad \text{(Fweb-65)}$$

This is called the *characteristic equation* for A. It is an nth-order equation for λ and has n roots, counting multiple roots according to their multiplicity.

Because the equation $A\mathbf{X} = \lambda\mathbf{X}$ is unchanged by multiplying both sides by a scalar s, it is clear that $s\mathbf{X}$ is an eigenvector of A if $\mathbf{X}$ is. This freedom can be used to *normalize* the eigenvectors, i.e., to choose the constant so that $\mathbf{X} \cdot \mathbf{X} = 1$. From n eigenvectors of A, $\mathbf{X}^{(i)}$, i = 1,2,...,n, we can construct the matrix

$$P \equiv \begin{bmatrix} X_1^{(1)} & X_1^{(2)} & X_1^{(3)} & \cdots & X_1^{(n)} \\ X_2^{(1)} & X_2^{(2)} & X_2^{(3)} & \cdots & X_2^{(n)} \\ \vdots & \vdots & \vdots & & \vdots \\ X_n^{(1)} & X_n^{(2)} & X_n^{(3)} & \cdots & X_n^{(n)} \end{bmatrix} \tag{Fweb-66}$$

Matrix multiplication and the eigenvalue equation [Eq. (Fweb-63)] give

$$AP = \begin{bmatrix} \lambda_1 X_1^{(1)} & \lambda_2 X_1^{(2)} & \cdots & \lambda_n X_1^{(n)} \\ \lambda_1 X_2^{(1)} & \lambda_2 X_2^{(2)} & \cdots & \lambda_n X_2^{(n)} \\ \vdots & \vdots & & \vdots \\ \lambda_1 X_n^{(1)} & \lambda_2 X_n^{(2)} & \cdots & \lambda_n X_n^{(n)} \end{bmatrix} = P\Lambda \tag{Fweb-67}$$

where Λ is the diagonal matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{Fweb-68}$$

The matrix P is nonsingular if and only if the n eigenvectors are linearly independent. In this case,

$$\Lambda = P^{-1}AP \tag{Fweb-69}$$

and we say that A is *diagonalizable* by the similarity transformation $P^{-1}AP$. If A is a normal matrix, we can choose n eigenvectors that are *orthonormal*, or simultaneously orthogonal and normalized:

$$\mathbf{X}^{(i)} \cdot \mathbf{X}^{(j)} = \delta_i^j \equiv \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{Fweb-70}$$

When the eigenvectors are orthonormal, P is a unitary matrix and A is diagonalizable by the unitary transformation $\Lambda = P^{\dagger}AP$. Any square matrix can be brought into *Jordan canonical form* [Hoffman and Kunze, 1961] by a similarity transformation

$$J = P^{-1}AP \tag{Fweb-71}$$

where the matrix J has the eigenvalues of A on the main diagonal and all zeros below the main diagonal. It follows from Eqs. (Fweb-71), (Fweb-59), and (Fweb-60) that the trace of A is equal to the sum of its eigenvalues, and the determinant of A is equal to the product of its eigenvalues; i.e.,

$$\mathrm{tr}A = \sum_{i=1}^{n} \lambda_i \tag{Fweb-72}$$

$$\det A = \lambda_1 \lambda_2 \ldots \lambda_n \tag{Fweb-73}$$

Many algorithms exist for finding eigenvalues and eigenvectors of matrices, several of which are discussed by Carnahan, et al., [1969] and by Stewart [1973]. Using Eq. (Fweb-61), we can see that the eigenvalues of Hermitian matrices are real numbers and the eigenvalues of unitary matrices are complex numbers with absolute value unity. Because the characteristic equation of a real matrix is a polynomial equation with real coefficients, the eigenvalues of a real matrix must either be real or must occur in complex conjugate pairs.

The case of a real orthogonal matrix deserves special attention. Because such a matrix is both real and unitary, the only possible eigenvalues are +1, –1, and complex conjugate pairs with absolute value unity. It follows that the determinant of a real orthogonal matrix is $(-1)^m$ where m is the multiplicity of the root $\lambda = -1$ of the characteristic equation. A proper real orthogonal matrix must have an even number of roots at $\lambda = -1$, and thus an even number for all $\lambda \neq 1$, because complex roots occur in conjugate pairs. Thus, an $n \times n$ proper real orthogonal matrix with *n odd* must have at least one eigenvector with eigenvalue +1. This is the basis of Euler's Theorem, discussed in Section 12.1.

It is also of interest to establish that the eigenvectors of a real symmetric matrix can be chosen to be real. The complex conjugate of the eigenvector equation, Eq. (Fweb-63), is $A\mathbf{X}^* = \lambda\mathbf{X}^*$, because both A and λ are real. Thus, $\mathbf{X}^*$ is an eigenvector of A with the same eigenvalue as $\mathbf{X}$. Now, either $\mathbf{X} = \mathbf{X}^*$, in which case the desired result is obtained, or $\mathbf{X} \neq \mathbf{X}^*$. In the latter case, we can replace $\mathbf{X}$ and $\mathbf{X}^*$ by the linear combinations $\mathbf{X} + \mathbf{X}^*$ and $i(\mathbf{X} - \mathbf{X}^*)$, which are real eigenvectors corresponding to the eigenvalue λ. Thus, we can always find a real orthogonal matrix P to diagonalize a real symmetric matrix A by Eq. (Fweb-69).

F.2.1.7 Functions of Matrices

Let f(x) be any function of a variable x, for example, $\sin x$ or $\exp x$. We want to give a meaning to f(M), where M is a square matrix. If f (x) has a power series expansion about $x = 0$,

$$\mathrm{f}\left(x\right) = \sum_{n=0}^{\infty} a_n x^n \tag{Fweb-74}$$

then we can formally (i.e., ignoring questions of convergence) define f(M) by

$$\mathrm{f}\left(M\right) = \sum_{n=0}^{\infty} a_n M^n \tag{Fweb-75}$$

with the same coefficients a_n. It is clear that f(M) is a square matrix of the same order as M. If M is a diagonalizable matrix, then by Eq. (Fweb-69),

$$M = P\,\Lambda\,P^{-1} \tag{Fweb-76}$$

where P is the matrix of eigenvectors defined by Eq. (Fweb-66), and Λ is the diagonal matrix of eigenvalues. Then,

$$M^n = \left(P\,\Lambda\,P^{-1}\right)^n = P\,\Lambda^n\,P^{-1} \tag{Fweb-77}$$

and

$$\mathrm{f}\left(M\right) = P\left(\sum_{n=0}^{\infty} a_n \Lambda^n\right)P^{-1} = P\begin{bmatrix} \mathrm{f}\left(\lambda_1\right) & 0 & \cdots & 0 \\ 0 & \mathrm{f}\left(\lambda_2\right) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \mathrm{f}\left(\lambda_n\right) \end{bmatrix}P^{-1} \tag{Fweb-78}$$

If M is a diagonalizable matrix, Eq. (Fweb-78) gives an alternative definition of f(M) that is valid when f(x) does not have a power series expansion, and agrees with Eq. (Fweb-75) when a power series expansion exists.

As an example, consider $\exp\left(\frac{1}{2}\Omega t\right)$, where Ω is the 4×4 matrix introduced in Section 16.1,

$$\Omega = \begin{bmatrix} 0 & \omega_3 & -\omega_2 & \omega_1 \\ -\omega_3 & 0 & \omega_1 & \omega_2 \\ \omega_2 & -\omega_1 & 0 & \omega_3 \\ -\omega_1 & -\omega_2 & -\omega_3 & 0 \end{bmatrix}$$

Matrix multiplication shows that

$$\Omega^2 = -\left(\omega_1^2 + \omega_2^2 + \omega_3^2\right)\mathbf{1} \equiv -\omega^2\,\mathbf{1}\,,$$

so it follows that

$$\Omega^{2k} = (-1)^k\,\omega^{2k}\,\mathbf{1}$$

$$\Omega^{2k+1} = (-1)^k\,\omega^{2k}\,\Omega$$

for all nonnegative k. Now,

$$\exp\left(\frac{1}{2}\Omega t\right) = \sum_{n=0}^{\infty}\frac{\left(\frac{1}{2}\Omega t\right)^n}{n!} = \sum_{k=0}^{\infty}\left[\frac{\left(\frac{1}{2}\Omega t\right)^{2k}}{(2k)!} + \frac{\left(\frac{1}{2}\Omega t\right)^{2k+1}}{(2k+1)!}\right]$$

$$= \mathbf{1}\sum_{k=0}^{\infty}\frac{(-1)\left(\frac{1}{2}\omega t\right)^{2k}}{(2k)!} + \Omega\omega^{-1}\sum_{k=0}^{\infty}\frac{(-1)^k\left(\frac{1}{2}\omega t\right)^{2k+1}}{(2k+1)!} \qquad \text{(Fweb-79)}$$

$$= \mathbf{1}\cos\left(\frac{1}{2}\omega t\right) + \Omega\omega^{-1}\sin\left(\frac{1}{2}\omega t\right)$$

$$= \begin{bmatrix} c & n_3 s & -n_2 s & n_1 s \\ -n_3 s & c & n_1 s & n_2 s \\ n_2 s & -n_1 s & c & n_3 s \\ -n_1 s & -n_2 s & -n_3 s & c \end{bmatrix}$$

where

$$c \equiv \cos\left(\frac{1}{2}\omega t\right)$$

$$s \equiv \sin\left(\frac{1}{2}\omega t\right)$$

$$n_i \equiv \omega_i / \omega \qquad i = 1,2,3$$

This example shows that the matrix elements of f(M) are not the functions f of the matrix elements of M, in general.

F.2.1.8 Vector Calculus

Let ϕ be a scalar function of the n arguments $X_1,X_2,...,X_n$. We consider the arguments to be the components of a column vector

$$\mathbf{X} \equiv [X_1,X_2,...,X_n]^T$$

The n partial derivatives of ϕ with respect to the elements of $\mathbf{X}$ are the components of the *gradient* of ϕ, denoted by

$$\frac{\partial \phi}{\partial \mathbf{X}} \equiv \left[\frac{\partial}{\partial X_1}, \frac{\partial}{\partial X_2}, \cdots, \frac{\partial}{\partial X_n} \right] \tag{Fweb-80}$$

Note that $\partial\phi/\partial\mathbf{X}$ is considered a $1 \times n$ row matrix. If we eliminate the function ϕ from Eq. (Fweb-80), we obtain the *gradient operator*

$$\frac{\partial}{\partial \mathbf{X}} \equiv \left[\frac{\partial}{\partial X_1}, \frac{\partial}{\partial X_2}, \cdots, \frac{\partial}{\partial X_n} \right] \tag{Fweb-81}$$

The matrix product of the $1 \times n$ gradient operator with an $n \times 1$ vector $\mathbf{Y}$ yields a scalar, the *divergence* of $\mathbf{Y}$, which we denote by

$$\frac{\partial}{\partial \mathbf{X}} \cdot \mathbf{Y} \equiv \sum_{i=1}^{n} \frac{\partial Y_i}{\partial X_i} \tag{Fweb-82}$$

The dot is used to emphasize the fact that the divergence is a scalar, although the usage is somewhat different from that in Eq. (Fweb-23).

The mn partial derivatives of an m-dimensional vector $\mathbf{Y}$ with respect to $X_1,X_2,\ldots,X_n$ can be arranged in an $m \times n$ matrix denoted by

$$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}} \equiv \left[\frac{\partial Y_i}{\partial X_j} \right] \tag{Fweb-83}$$

This is like an outer product of $\mathbf{Y}$ and $\partial/\partial\mathbf{X}$; however, the analogy is not perfect because the gradient operator appears on the right in the matrix product sense and on the left in the operator sense.

F.2.1.9 Vectors in Three Dimensions

In this section, we only consider vectors with three real components. For three-component vectors, three products are defined: the dot product, the outer product, and the *cross product*. The cross product, or *vector product*, is a vector defined by

$$\mathbf{U} \times \mathbf{V} \equiv \begin{bmatrix} U_2V_3 - U_3V_2 \\ U_3V_1 - U_1V_3 \\ U_1V_2 - U_2V_1 \end{bmatrix} \tag{Fweb-84}$$

The following identities are often useful:

$$\mathbf{U} \cdot \mathbf{V} \equiv U_1V_1 + U_2V_2 + U_3V_3 = UV\cos\theta \tag{Fweb-85a}$$

$$\left| \mathbf{U} \times \mathbf{V} \right| = UV \sin\theta \tag{Fweb-85b}$$

where $\theta\,(0 \le \theta \le 180°)$ is the angle between **U** and **V**. In addition,

$$\mathbf{U} \times \mathbf{V} = -\mathbf{V} \times \mathbf{U} \tag{Fweb-86}$$

$$\mathbf{U} \cdot (\mathbf{U} \times \mathbf{V}) = 0 \tag{Fweb-87}$$

$$\mathbf{U} \cdot (\mathbf{V} \times \mathbf{W}) = \mathbf{V} \cdot (\mathbf{W} \times \mathbf{U}) = \mathbf{W} \cdot (\mathbf{U} \times \mathbf{V}) = \begin{vmatrix} U_1 & U_2 & U_3 \\ V_1 & V_2 & V_3 \\ W_1 & W_2 & W_3 \end{vmatrix} \tag{Fweb-88}$$

$$\begin{aligned} \left[\mathbf{U} \cdot (\mathbf{V} \times \mathbf{W}) \right]^2 &= (\mathbf{U} \times \mathbf{V}) \cdot \left[(\mathbf{V} \times \mathbf{W}) \times (\mathbf{W} \times \mathbf{U}) \right] \\ &= \mathbf{U}^2 \mathbf{V}^2 \mathbf{W}^2 - \mathbf{U}^2 (\mathbf{V} \cdot \mathbf{W})^2 - \mathbf{V}^2 (\mathbf{U} \cdot \mathbf{W})^2 - \mathbf{W}^2 (\mathbf{U} \cdot \mathbf{V})^2 + 2 (\mathbf{U} \cdot \mathbf{V})(\mathbf{V} \cdot \mathbf{W})(\mathbf{W} \cdot \mathbf{U}) \end{aligned} \tag{Fweb-89}$$

$$\mathbf{U} \times (\mathbf{V} \times \mathbf{W}) = \mathbf{V}(\mathbf{U} \cdot \mathbf{W}) - \mathbf{W}(\mathbf{U} \cdot \mathbf{V}) \tag{Fweb-90}$$

$$\mathbf{0} = \mathbf{U} \times (\mathbf{V} \times \mathbf{W}) + \mathbf{V} \times (\mathbf{W} \times \mathbf{U}) + \mathbf{W} \times (\mathbf{U} \times \mathbf{V}) \tag{Fweb-91}$$

$$(\mathbf{U} \times \mathbf{V}) \cdot (\mathbf{W} \times \mathbf{X}) = (\mathbf{U} \cdot \mathbf{W})(\mathbf{V} \cdot \mathbf{X}) - (\mathbf{U} \cdot \mathbf{X})(\mathbf{V} \cdot \mathbf{W}) \tag{Fweb-92}$$

The following identity provides a means of writing the vector **W** in terms of **U**, **V**, and **U** × **V**, if **U** × **V** ≠ **0**:

$$\begin{aligned} \left[(\mathbf{U} \times \mathbf{V}) \cdot (\mathbf{U} \times \mathbf{V}) \right] \mathbf{W} &= \left[(\mathbf{V} \times \mathbf{U}) \cdot (\mathbf{V} \times \mathbf{W}) \right] \mathbf{U} + \left[(\mathbf{U} \times \mathbf{V}) \cdot (\mathbf{U} \times \mathbf{W}) \right] \mathbf{V} \\ &\quad + \left[\mathbf{W} \cdot (\mathbf{U} \times \mathbf{V}) \right] \mathbf{U} \times \mathbf{V} \end{aligned} \tag{Fweb-93}$$

If A is a real orthogonal matrix,

$$(A\mathbf{U}) \times (A\mathbf{V}) = \pm A(\mathbf{U} \times \mathbf{V}) \tag{Fweb-94}$$

where the positive sign holds if A is proper, and the negative sign if A is improper.

The tangent of the *rotation angle* from **V** to **W** about **U** (the angle of the rotation in the positive sense about **U** that takes **V** × **U** into a vector parallel to **W** × **U**) is

$$\tan\Theta = \frac{\left| \mathbf{U} \right| \mathbf{U} \cdot (\mathbf{V} \times \mathbf{W})}{\mathbf{U}^2 (\mathbf{V} \cdot \mathbf{W}) - (\mathbf{U} \cdot \mathbf{V})(\mathbf{U} \cdot \mathbf{W})} \tag{Fweb-95}$$

The quadrant of Θ is given by the fact that the numerator is a positive constant multiplied by sin Θ, and the denominator is the same positive constant multiplied by cos Θ. If, **U, V,** and **W** are unit vectors, Θ is the same as the rotation angle on the celestial sphere defined in Appendix A. Equation (Fweb-95) is derived in Section 7.3. [See Eqs. (7-57).]

References

1. Bellman, R. E., *Introduction to Matrix Algebra*. New York: McGraw-Hill, Inc., 1960.

2. Carnahan, B., H. A. Luther and J. 0. Wilkes, *Applied Numerical Methods*. New York: John Wiley & Sons, Inc., 1969.
3. Forsythe, George E., and C. Moler, *Computer Solution of Linear Algebraic Systems*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1967.
4. Halmos, P. R., ed., *Finite-Dimension Vector Spaces*, Second Edition. Princeton, NJ: D. Van Nostrand Company, Inc., 1958.
5. Hoffman, Kenneth and Ray Kunze, *Linear Algebra*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1961.
6. Noble, Ben, *Applied Linear Algebra*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1969.
7. Stewart, G. W., *Introduction to Matrix Computations*. New York: Academic Press, Inc., 1973.
8. Wilberg, Donald M., State Space and Linear Systems, *Schaum's Outline Series*, New York: McGraw-Hill, Inc., 1971.