Appendix B

Review of Matrices and Linear Algebra

The concept of *matrices*, defined as a set of numbers arranged in various *rows* and *columns*, began with efforts to simultaneously solve a set of linear algebraic equations. Hence, the study of matrices and their properties is referred to as *linear algebra*. For example, consider the following linear equations:

$$\begin{aligned} x + 2y + 3z &= 5 \\ -x + y + 7z &= 0 \\ 3x - 17y + 2z &= 21 \end{aligned} \tag{B.1}$$

An attempt to solve these equations simultaneously for the unknowns x, y, and z results in their being expressed in the following form:

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 1 & 7 \\ 3 & -17 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \\ 21 \end{bmatrix} \tag{B.2}$$

In Eq. (B.2), we denote

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 1 & 7 \\ 3 & -17 & 2 \end{bmatrix}; \quad \mathbf{V} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 5 \\ 0 \\ 21 \end{bmatrix} \tag{B.3}$$

and re-write Eq. (B.2) as $\mathbf{AV} = \mathbf{d}$, where $\mathbf{A}$ is called a *matrix* of size (3×3) (because it has three rows and three columns), and $\mathbf{V}$ and $\mathbf{d}$ are *matrices* of size (3×1). A matrix with only *one column* (such as $\mathbf{V}$ and $\mathbf{d}$) has a special name – *column vector*, while a matrix with only *one row* is called a *row vector*. A matrix with only *one* row and only *one* column consists of only one number, and is called a *scalar*. The numbers of which a matrix is formed are called the *elements* of the matrix. For example, in Eq. (B.3), the matrix $\mathbf{A}$ has nine elements. In general, a matrix with n rows and m columns would have *nm* elements, and is denoted as follows:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \tag{B.4}$$

The element of the matrix, **A**, in Eq. (B.4) located in the ith row and jth column is denoted by a_{ij}. Some elementary matrix operations are defined as follows:

Addition: addition of two matrices of the *same size*, **A** and **B**, is defined as addition of *all the corresponding elements*, a_{ij} and b_{ij}, of the matrices **A** and **B**, i.e.

$$\mathbf{A} + \mathbf{B} = \mathbf{C} \tag{B.5}$$

where the element c_{ij} of the matrix **C** is calculated as $c_{ij} = a_{ij} + b_{ij}$ for all i and j. *Matrix subtraction* is defined in the same manner as the addition, except that the corresponding elements are subtracted rather than added.

Multiplication by a scalar: a matrix, **A**, is said to be *multiplied* by a scalar, a, if all the elements of **A** are multiplied by a.

Multiplication of a row vector with a column vector: a row vector, **r**, of size $(1 \times n)$ is said to be multiplied with a column vector, **c**, of size $(n \times 1)$ if the products of the corresponding elements, $r_{1k}c_{k1}$, are summed as follows, resulting in a *scalar*:

$$\mathbf{rc} = [\, r_{11} \quad r_{12} \quad \dots \quad r_{1n} \,] \begin{bmatrix} c_{11} \\ c_{21} \\ \dots \\ c_{n1} \end{bmatrix} = r_{11}c_{11} + r_{12}c_{21} + \cdots + r_{1n}c_{n1} \tag{B.6}$$

Multiplication of two matrices: a matrix, **A**, of size $(n \times p)$ is said to be *multiplied* with a matrix, **B**, of size $(p \times m)$, resulting in a matrix, **C**, of size $(n \times m)$, expressed as

$$\mathbf{AB} = \mathbf{C} \tag{B.7}$$

such that the element c_{ij} of the matrix, **C**, is the *multiplication* of the ith *row* of the matrix, **A**, with the jth *column* of the matrix, **B**, for all i and j. For example, in Eq. (B.2), a matrix of size (3×3) is multiplied with a column vector of size (3×1) to produce a column vector of size (3×1).

Some special matrices are defined as follows:

Square matrix: a matrix that has *equal* number of rows and columns is called a *square matrix*. The elements, a_{ii}, of a square matrix, **A**, for all i are said to be the *diagonal elements* of the square matrix.

Symmetric matrix: a square matrix, **A**, is said to be *symmetric* if the element in the ith row and jth column is *equal* to the element in the jth row and ith column, i.e. $a_{ij} = a_{ji}$, *for all* i and j.

Diagonal matrix: a square matrix that has all the elements, except the *diagonal elements*, equal to zero is called a *diagonal matrix.*

Identity matrix: a diagonal matrix that has *all* the *diagonal elements* equal to *unity* is called an *identity matrix*, and is denoted by **I**. An identity matrix has the property that if **A** is a square matrix of the same size as the identity matrix, **I**, then

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A} \tag{B.8}$$

Some of the more advanced matrix operations are defined as follows (you may refer to a textbook on linear algebra for details of these operations [1–3]):

Transpose: transpose of a matrix, **A**, is defined as the matrix, $\mathbf{A}^\mathbf{T}$, in which the *rows* and *columns* of the matrix, **A**, have been *interchanged.* Transpose of a symmetric matrix, **S**, is equal to **S**. The transpose of a product of two matrices has the following property:

$$(\mathbf{AB})^T = \mathbf{B}^\mathbf{T}\mathbf{A}^\mathbf{T} \tag{B.9}$$

Trace: *Trace* of a *square* matrix, **A**, of size $(n \times n)$ is defined as the sum of *all* the *diagonal elements* of **A**, i.e.

$$trace(\mathbf{A}) = a_{11} + a_{22} + \cdots + a_{nn} \tag{B.10}$$

Determinant and minor: *Determinant* of a square matrix, **A**, of size $(n \times n)$ is defined as the sum of *all possible* products of n elements, each taken from a *different* column. A more useful (but *recursive*) definition of the determinant of **A**, denoted by $|\mathbf{A}|$, is the following:

$$|\mathbf{A}| = a_{11}D_{11} - a_{12}D_{12} + a_{13}D_{13} + \cdots + (-1)^n a_{1n}D_{1n} \tag{B.11}$$

where D_{ij} is called the *minor* of element, a_{ij}, and defined as the *determinant* of the *square sub-matrix* of size $((n-1) \times (n-1))$ formed out of the matrix **A** by deleting the ith row and the jth column. This procedure can be followed recursively (either by hand or a computer program) until we are left with the minors of the *smallest* size, i.e. a *scalar*. Instead of finding the determinant by taking elements from different columns, we can alternatively take the elements from *different rows* of **A**, and express the determinant as follows:

$$|\mathbf{A}| = a_{11}D_{11} - a_{21}D_{21} + a_{31}D_{31} + \cdots + (-1)^n a_{n1}D_{n1} \tag{B.12}$$

If $|\mathbf{A}| = 0$, then the matrix, **A**, is said to be *singular*. If $|\mathbf{A}| \neq 0$, then the matrix **A** is said to be *non-singular*. Determinant has the following property:

$$|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}| \tag{B.13}$$

Cofactor: the cofactor, c_{ij}, of an element, a_{ij}, of a square matrix, **A**, is defined as

$$c_{ij} = (-1)^{i+j} D_{ij} \tag{B.14}$$

where D_{ij} is the *minor* associated with the element, a_{ij}.

Adjoint: the *transpose* of a matrix whose elements are the *cofactors*, c_{ij}, of a square matrix, **A**, is called the *adjoint* of **A**, denoted by *adj*(**A**), and expressed as follows

$$adj(\mathbf{A}) = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix}^T \tag{B.15}$$

Inverse: the inverse of a square matrix, **A**, is defined as a matrix, $\mathbf{A}^{-1}$, which when multiplied with **A**, produces an identity matrix, **I**,

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I} \tag{B.16}$$

The matrix inverse can be calculated by the *Cramer's rule*, expressed as follows:

$$\mathbf{A}^{-1} = adj(\mathbf{A})/|\mathbf{A}| \tag{B.17}$$

From Eq. (B.17), it is clear that $\mathbf{A}^{-1}$ exists only if **A** is *non-singular*. Some important properties of matrix inverse are the following:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{B.18}$$

$$(\mathbf{A}^{-1})^T = (\mathbf{A}^{\mathbf{T}})^{-1} \tag{B.19}$$

$$|\mathbf{A}^{-1}| = 1/|\mathbf{A}| \tag{B.20}$$

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1} \tag{B.21}$$

Eigenvalues: the *eigenvalues* of a square matrix, **A**, are defined as the roots, λ_i, of the following *characteristic polynomial equation*:

$$|\lambda\mathbf{I} - \mathbf{A}| = 0 \tag{B.22}$$

where **I** is an identity matrix of the same size as **A**. A matrix of size $(n \times n)$ has n eigenvalues. The *product* of all the eigenvalues of a square matrix is equal to the determinant of the matrix. The *sum* of all the eigenvalues of a square matrix is equal to the trace of the matrix.

Eigenvectors: The *eigenvector*, $\mathbf{v_i}$, of a square matrix, **A**, associated with the eigenvalue, λ_i, of **A** is defined as the vector that satisfies the following equation:

$$\mathbf{A}\mathbf{v_i} = \lambda_i \mathbf{v_i} \tag{B.23}$$

Cayley–Hamilton Theorem: a fundamental relation of linear algebra is the theorem that states that if the characteristic polynomial equation (Eq. (B.22)) of a square matrix, **A**, of size $(n \times n)$ is expressed as

$$|\lambda\mathbf{I} - \mathbf{A}| = \lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_{n-1}\lambda + \alpha_n = 0 \tag{B.24}$$

where α_k are *characteristic coefficients*, then

$$\mathbf{A}^n + \alpha_1\mathbf{A}^{n-1} + \cdots + \alpha_{n-1}\mathbf{A} + \alpha_n\mathbf{I} = \mathbf{0} \tag{B.25}$$

where $\mathbf{A^k}$ denotes **A** *multiplied* by **A**, k times (or, **A** raised to the *power* k), and **0** denotes a matrix with *all* zero elements of the same size as **A**.

Hermitian: the *transpose* of the *complex conjugate* of a matrix, **A**, is called the *hermitian* of **A**, and denoted by $\mathbf{A^H}$. A matrix, **U**, whose hermitian equals its inverse (i.e. $\mathbf{U^H} = \mathbf{U}^{-1}$) is called a *unitary matrix*.

Differentiation of a matrix by a scalar: the *derivative* of a matrix, **A**, with respect to a scalar, c, is defined as the matrix, $d\mathbf{A}/dc$, each element of which is the *derivative* of the corresponding element, da_{ij}/dc, of the matrix, **A**.

Integration of a matrix: the *integral* of a matrix, **A**, is defined as the matrix whose elements are *integrals* of the corresponding elements of **A**.

Differentiation of a scalar by a vector: the derivative of a scalar by a vector is defined as the vector, each element of which is the *derivative* of the scalar with the *corresponding* element of the vector. For example, if a scalar z, is related to *two vectors*, **x** and **y** such that $z = \mathbf{x^T y} = \mathbf{y^T x}$, then the *partial derivative* of z with respect to the vector, **x**, is $\partial z/\partial\mathbf{x} = \mathbf{y}$, and the partial derivative of z with respect to the vector **y** is $\partial z/\partial\mathbf{y} = \mathbf{x}$.

Quadratic form: the scalar, z, is called the *quadratic form* of the vector, **x**, if it can be expressed as

$$z = \mathbf{x^T Q x} \tag{B.26}$$

where **Q** is a *square* matrix. It can be shown that the derivative of z with respect to **x** is $\partial z/\partial\mathbf{x} = 2\mathbf{Qx}$.

Bilinear form: the scalar, z, is called the *bilinear form* of the vectors, **x** and **y**, if it can be expressed as

$$z = \mathbf{x^T Q y} \tag{B.27}$$

where $\mathbf{Q}$ is a matrix. It can be shown that the derivative of z with respect to $\mathbf{x}$ is $\partial z/\partial \mathbf{x} = \mathbf{Q}\mathbf{y}$ and the derivative of z with respect to $\mathbf{y}$ is $\partial z/\partial \mathbf{y} = \mathbf{Q}^{\mathbf{T}}\mathbf{x}$.

Positive definiteness: if the quadratic form $\mathbf{x}^{\mathbf{T}}\mathbf{Q}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$, then the quadratic form is said to be *positive definite*. If the quadratic form $\mathbf{x}^{\mathbf{T}}\mathbf{Q}\mathbf{x} \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$, then the quadratic form is said to be *positive semi-definite*.

References

1. Gantmacher, F.R. *Theory of Matrices*. Chelsea, New York, 1959.
2. Strang, G. *Linear Algebra and Its Applications*. Academic Press, New York, 1976.
3. Kreyszig, E. *Advanced Engineering Mathematics*. Wiley, New York, 1972.