A Matrix Algebra

The aim of this appendix is to give a brief overview of matrix algebra, which covers a number of issues referred to in the main text. It includes an introduction to matrix properties, operators, and classes.

A.1 Introduction

Since many MATLAB functions and operators act on matrices and arrays, it is important that MATLAB users feel at ease with matrix notation and matrix algebra. MATLAB is an ideal environment in which to experiment and learn matrix algebra. While it cannot provide a formal proof of any relationship, it does allow users to verify results and rapidly gain experience in matrix manipulation. In this appendix only definitions and results are provided. For proofs and further explanation, it is recommended that the reader consult Golub and Van Loan (1989).

A.2 Matrices and Vectors

A matrix is a rectangular array of elements that in itself cannot be evaluated. An element of a matrix can be a real or complex number, an algebraic expression, or another matrix. Normally matrices are enclosed in square brackets, parentheses, or braces. In this text square brackets are used. A complete matrix is denoted by an emboldened character. For example,

$$\mathbf{A} = \begin{bmatrix} 3 & -2 \\ -2 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{A} & \mathbf{A} & 2\mathbf{A} \\ \mathbf{A} & -\mathbf{A} & \mathbf{A} \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} 11 \\ -3 \\ 7 \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} (2+3\imath) & \left(p^2+q\right) & (-4+7\imath) & (3-4\imath) \end{bmatrix}$$

and hence

$$\mathbf{B} = \begin{bmatrix} 3 & -2 & 3 & -2 & 6 & -4 \\ -2 & 4 & -2 & 4 & -4 & 8 \\ 3 & -2 & -3 & 2 & 3 & -2 \\ -2 & 4 & 2 & -4 & -2 & 4 \end{bmatrix}$$

where $\imath = \sqrt{(-1)}$. In the preceding examples **A** is a 2×2 square matrix with two rows and two columns of real coefficients. It also has the property of being a symmetric matrix (see Section A.7). The matrix **B** is built up from the matrix **A** and so **B** is a 4×6 real matrix. The matrix **x** is a 3×1 matrix and is usually called a column vector, and **e** is a 1×4 complex matrix, usually called a row vector. Note that **e** has the algebraic expression $p^2 + q$ for its second element. In this vector each element is enclosed in parentheses to clarify its structure. Enclosing an element in parentheses is not a requirement.

If we wish to refer to a particular element in a matrix, we use subscript notation: The first subscript denotes the row, the second the column. In the case of the row and column vectors it is conventional to use a single subscript. Thus, in the preceding examples,

$$a_{21} = -2, \;\; b_{25} = -4, \;\; x_2 = -3, \;\; e_4 = 3 - 4\imath$$

Note also that although **A** and **B** are uppercase letters it is conventional to refer to their elements by lowercase letters. In general the element in the ith row and jth column of **A** is denoted by a_{ij}.

A.3 Some Special Matrices

The identity matrix. The identity matrix, denoted by **I**, has unit values along the leading diagonal and zeros elsewhere. The leading diagonal is the diagonal of elements from the top left to the bottom right of the matrix. For example,

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The subscript indicating the size of the matrix is usually omitted. The identity matrix behaves rather like the scalar quantity 1. In particular, pre- or postmultiplying a matrix by **I** does not change it.

The diagonal matrix. This matrix is square and has nonzero elements *only* along the leading diagonal. Thus

$$\mathbf{A} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 12 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -6 \end{bmatrix}$$

The tridiagonal matrix. This matrix is square and has nonzero elements along the leading diagonal and the diagonals immediately above and below it. Thus, using "x" to denote a nonzero element,

$$\mathbf{A} = \begin{bmatrix} x & x & 0 & 0 & 0 \\ x & x & x & 0 & 0 \\ 0 & x & x & x & 0 \\ 0 & 0 & x & x & x \\ 0 & 0 & 0 & x & x \end{bmatrix}$$

Triangular and Hessenberg matrices. A lower triangular matrix has nonzero elements only on and below the leading diagonal. An upper triangular matrix has nonzero elements on and above the leading diagonal. The Hessenberg matrix is similar to the triangular matrix except that in addition it has nonzero elements on the diagonals adjacent to the leading diagonal.

$$\begin{bmatrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & 0 & x \end{bmatrix}, \quad \begin{bmatrix} x & 0 & 0 & 0 & 0 \\ x & x & 0 & 0 & 0 \\ x & x & x & 0 & 0 \\ x & x & x & x & 0 \\ x & x & x & x & x \end{bmatrix}, \quad \begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & 0 & x & x \end{bmatrix}$$

The first matrix is upper triangular, the second is lower triangular, and the last is upper Hessenberg.

A.4 Determinants

The determinant of $\mathbf{A}$ is written $|\mathbf{A}|$ or $\det(\mathbf{A})$. For a 2×2 array we define its determinant as follows:

$$\text{If } \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ then } \det(\mathbf{A}) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12} \tag{A.1}$$

In general for an $n \times n$ array $\mathbf{A}$, cofactors $C_{ij} = (-1)^{i+j}\,\Delta_{ij}$ can be defined. In this definition Δ_{ij} is the determinant formed from $\mathbf{A}$ when the elements of the ith row and jth column are deleted. Δ_{ij} is called the minor of $\mathbf{A}$. Then

$$\det(\mathbf{A}) = \sum_{k=1}^{n} a_{ik}C_{ik} \quad \text{for any} \quad i = 1, 2, \ldots, n \tag{A.2}$$

This is known as an expansion along the ith row. Frequently the first row is used. This equation replaces the problem of evaluating one $n \times n$ determinant $\mathbf{A}$ by the evaluation of n, $(n-1) \times (n-1)$ determinants. The process can be continued until the cofactors are reduced to 2×2 determinants. Then the formula (A.1) is used. This is the formal definition for the determinant of $\mathbf{A}$ but it is not a computationally efficient procedure.

A.5 Matrix Operations

Matrix transposition. In this operation the rows and columns of a matrix are interchanged or transposed. The transposition of a real matrix $\mathbf{A}$ is denoted by $\mathbf{A}^\top$. For example,

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & 4 \\ 2 & 1 & 7 \end{bmatrix}, \quad \mathbf{A}^\top = \begin{bmatrix} 1 & 2 \\ -1 & 1 \\ 4 & 7 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{x}^\top = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

Note that a square matrix remains square when it is transposed and a column vector transposes into a row vector and vice versa.

Matrix addition and subtraction. This is done by adding or subtracting corresponding elements in the matrices. Thus

$$\begin{bmatrix} 1 & 3 \\ -4 & 5 \end{bmatrix} + \begin{bmatrix} 5 & -4 \\ 6 & 6 \end{bmatrix} = \begin{bmatrix} 6 & -1 \\ 2 & 11 \end{bmatrix}, \quad \begin{bmatrix} -4 \\ 6 \\ 11 \end{bmatrix} - \begin{bmatrix} 3 \\ -3 \\ 2 \end{bmatrix} = \begin{bmatrix} -7 \\ 9 \\ 9 \end{bmatrix}$$

It is apparent that only matrices with the same number of rows and the same number of columns can be added and subtracted. In general, if $\mathbf{A} = \mathbf{B} + \mathbf{C}$ then $a_{ij} = b_{ij} + c_{ij}$.

Scalar multiplication. Every element of a matrix is multiplied by a scalar quantity. Thus if $\mathbf{A} = s\mathbf{B}$, where s is a scalar, then $a_{ij} = sb_{ij}$.

Matrix multiplication. We can only multiply two matrices $\mathbf{B}$ and $\mathbf{C}$ together if the number of columns in $\mathbf{B}$ is equal to the number of rows in $\mathbf{C}$. Such matrices are said to be conformable. If $\mathbf{B}$ is a $p \times q$ matrix and $\mathbf{C}$ is a $q \times r$ matrix, then we can determine the product $\mathbf{A} = \mathbf{BC}$ and the result will be a $p \times r$ matrix. Because the order of matrix multiplication is important, we say that $\mathbf{B}$ premultiplies $\mathbf{C}$ or $\mathbf{C}$ postmultiplies $\mathbf{B}$. If $\mathbf{A} = \mathbf{BC}$, the elements of $\mathbf{A}$ are determined from the following relationship:

$$a_{ij} = \sum_{k=1}^{q} b_{ik} c_{kj} \quad \text{for} \quad i = 1, 2, \ldots, p; \quad j = 1, 2, \ldots, r$$

For example,

$$\begin{aligned} &\begin{bmatrix} 2 & -3 & 1 \\ -5 & 4 & 3 \end{bmatrix} \begin{bmatrix} -6 & 4 & 1 \\ -4 & 2 & 3 \\ 3 & -7 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 2(-6) + (-3)(-4) + 1(3) & 2(4) + (-3)2 + 1(-7) & 2(1) + (-3)3 + 1(-1) \\ (-5)(-6) + 4(-4) + 3(3) & (-5)4 + 4(2) + 3(-7) & (-5)1 + 4(3) + 3(-1) \end{bmatrix} \\ &= \begin{bmatrix} 3 & -5 & -8 \\ 23 & -33 & 4 \end{bmatrix} \end{aligned}$$

Note that the product of a 2×3 and a 3×3 matrix is a 2×3 matrix. Consider four further examples of matrix multiplication:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 5 & 6 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 11 & 10 \\ 27 & 26 \end{bmatrix}, \quad \begin{bmatrix} 5 & 6 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 23 & 34 \\ 9 & 14 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} -4 \\ 3 \\ 3 \end{bmatrix} = 11, \quad \begin{bmatrix} -4 \\ 3 \\ 3 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} -4 & -8 & -12 \\ 3 & 6 & 9 \\ 3 & 6 & 9 \end{bmatrix}$$

In the preceding examples note that while the 2×2 matrices can be multiplied in either order, the product is different. This is an important observation and in general $\mathbf{BC} \neq \mathbf{CB}$. Note also that multiplying a row by a column vector gives a scalar whereas multiplying a column by a row results in a matrix.

Matrix inversion. The inverse of a square matrix $\mathbf{A}$ is written $\mathbf{A}^{-1}$ and is defined by

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

The formal definition of $\mathbf{A}^{-1}$ is

$$\mathbf{A}^{-1} = \mathrm{adj}(\mathbf{A}) / \det(\mathbf{A}) \tag{A.3}$$

where $\mathrm{adj}(\mathbf{A})$ is the adjoint of $\mathbf{A}$. The adjoint of $\mathbf{A}$ is given by

$$\mathrm{adj}(\mathbf{A}) = \mathbf{C}^\top$$

where $\mathbf{C}$ is a matrix composed of the cofactors of $\mathbf{A}$. Using (A.3) is not an efficient way to compute an inverse.

A.6 Complex Matrices

A matrix can have elements that are complex and such a matrix can be expressed in terms of two real matrices. Thus

$$\mathbf{A} = \mathbf{B} + \imath\mathbf{C} \quad \text{where} \quad \imath = \sqrt{(-1)}$$

Here $\mathbf{A}$ is complex and $\mathbf{B}$ and $\mathbf{C}$ are real matrices. The complex conjugate of $\mathbf{A}$ is normally denoted by $\mathbf{A}^*$ and is equal to

$$\mathbf{A}^* = \mathbf{B} - \imath\mathbf{C}$$

Matrix $\mathbf{A}$ can be transposed so that

$$\mathbf{A}^\top = \mathbf{B}^\top + \imath\mathbf{C}^\top$$

Matrix $\mathbf{A}$ can be transposed *and* conjugated at the same time and this is denoted by $\mathbf{A}^{\mathrm{H}}$ and called the Hermitian transpose. Thus

$$\mathbf{A}^{\mathrm{H}} = \mathbf{B}^{\top} - \imath\mathbf{C}^{\top}$$

For example,

$$\mathbf{A} = \begin{bmatrix} 1-\imath & -2-3\imath & 4\imath \\ 2 & 1+2\imath & 7+5\imath \end{bmatrix}, \quad \mathbf{A}^{*} = \begin{bmatrix} 1+\imath & -2+3\imath & -4\imath \\ 2 & 1-2\imath & 7-5\imath \end{bmatrix}$$

$$\mathbf{A}^{\top} = \begin{bmatrix} 1-\imath & 2 \\ -2-3\imath & 1+2\imath \\ 4\imath & 7+5\imath \end{bmatrix}, \quad \mathbf{A}^{\mathrm{H}} = \begin{bmatrix} 1+\imath & 2 \\ -2+3\imath & 1-2\imath \\ -4\imath & 7-5\imath \end{bmatrix}$$

It is important to note that the MATLAB expression `A'` gives the conjugation and transposition of `A` when applied to a complex matrix; that is, it is equivalent to $\mathbf{A}^{\mathrm{H}}$. However, `A.'` gives ordinary transposition, which corresponds to $\mathbf{A}^{\top}$.

A.7 Matrix Properties

The real square matrix $\mathbf{A}$ is

$$\text{symmetric if } \mathbf{A}^{\top} = \mathbf{A}$$

$$\text{skew-symmetric if } \mathbf{A}^{\top} = -\mathbf{A}$$

$$\text{orthogonal if } \mathbf{A}^{\top} = \mathbf{A}^{-1}$$

$$\text{nilpotent if } \mathbf{A}^{p} = \mathbf{0}, \quad \text{where } p \text{ is a positive integer and } \mathbf{0} \text{ is the matrix of zeros}$$

$$\text{idempotent if } \mathbf{A}^{2} = \mathbf{A}$$

The complex square matrix $\mathbf{A} = \mathbf{B} + \imath\mathbf{C}$ is

$$\text{Hermitian if } \mathbf{A}^{\mathrm{H}} = \mathbf{A}$$

$$\text{unitary if } \mathbf{A}^{\mathrm{H}} = \mathbf{A}^{-1}$$

A.8 Some Matrix Relationships

If $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{R}$ are matrices such that

$$\mathbf{W} = \mathbf{P}\,\mathbf{Q}\,\mathbf{R}$$

then

$$\mathbf{W}^{\top} = \mathbf{R}^{\top}\mathbf{Q}^{\top}\mathbf{P}^{\top} \tag{A.4}$$

and

$$\mathbf{W}^{-1} = \mathbf{R}^{-1}\mathbf{Q}^{-1}\mathbf{P}^{-1} \tag{A.5}$$

If $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{R}$ are complex, then (A.5) is still valid and (A.4) becomes

$$\mathbf{W}^{\mathrm{H}} = \mathbf{R}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}}\mathbf{P}^{\mathrm{H}} \tag{A.6}$$

A.9 Eigenvalues

Consider the eigenvalue problem

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

If $\mathbf{A}$ is an $n \times n$ symmetric matrix, then there are n real eigenvalues, λ_i, and n real eigenvectors, $\mathbf{x}_i$, that satisfy this equation. If $\mathbf{A}$ is an $n \times n$ Hermitian matrix, then there are n real eigenvalues, λ_i, and n complex eigenvectors, $\mathbf{x}_i$, that satisfy the eigenvalue problem. The polynomial in λ given by $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ is called the characteristic equation. The roots of this polynomial are the eigenvalues of $\mathbf{A}$. The sum of the eigenvalues of $\mathbf{A}$ equals trace($\mathbf{A}$) where trace($\mathbf{A}$) is defined as the sum of the elements on the leading diagonal of $\mathbf{A}$. The product of the eigenvalues of $\mathbf{A}$ equals det($\mathbf{A}$).

It is interesting to note that if we define $\mathbf{C}$ as

$$\mathbf{C} = \begin{bmatrix} -p_1/p_0 & -p_2/p_0 & \dots & -p_{n-1}/p_0 & -p_n/p_0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

then the eigenvalues of $\mathbf{C}$ are the roots of the polynomial

$$p_0 x^n + p_1 x^{n-1} + \cdots + p_{n-1} x + p_n = 0$$

The matrix $\mathbf{C}$ is called the companion matrix.

A.10 Definition of Norms

The p-norm for the vector $\mathbf{v}$ is defined as follows:

$$||\mathbf{v}||_p = \left(|v_1|^p + |v_2|^p + \cdots + |v_n|^p\right)^{1/p} \tag{A.7}$$

The parameter p can take any value but only three values are commonly used. If $p = 1$ in (A.7), we have the 1-norm, $||\mathbf{v}||_1$:

$$||\mathbf{v}||_1 = |v_1| + |v_2| + \cdots + |v_n| \tag{A.8}$$

If $p = 2$ in (A.7), we have the 2-norm or Euclidean norm of the vector $\mathbf{v}$, which is written $||\mathbf{v}||$ or $||\mathbf{v}||_2$ and is defined as follows:

$$||\mathbf{v}||_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \tag{A.9}$$

Note that it is not necessary to take the modulus of the elements because in this case each element value is squared. The Euclidean norm is also called the length of the vector. These names arise from the fact that in two- or three-dimensional Euclidean space a vector of two or three elements is used to specify a position in space. The distance from the origin to the specified position is identical to the Euclidean norm of the vector.

If p tends to infinity in (A.7), we have $||\mathbf{v}||_\infty = \max(|v_1|, |v_2|, \ldots, |v_n|)$, the infinity norm. At first sight this might appear inconsistent with (A.7). However, when p tends to infinity, the modulus of each element is raised to a very large power and the largest element will dominate the summation.

These functions are implemented in MATLAB; `norm(v,1)`, `norm(v,2)` (or `norm(v)`), and `norm(v,inf)` return the 1, 2, and infinity norms of the vector `v`, respectively.

A.11 Reduced Row Echelon Form

The reduced row echelon form (RREF) of a matrix also has an important role to play in the theoretical understanding of linear algebra. A matrix is transformed into its RREF when the following conditions have been met:

1. All zero rows, if they exist, are at the bottom of the matrix.
2. The first nonzero element in every nonzero row is unity.
3. For each nonzero row, the first nonzero element appears to the right of the first nonzero element of the preceding row.
4. For any column in which the first nonzero element of a row appears, all other elements are zero.

The RREF is determined by using a finite sequence of elementary row operations. It is a standard form and the most fundamental form of a matrix that can be achieved using elementary row operations alone.

For a system of equations $\mathbf{Ax} = \mathbf{b}$ we can define the augmented matrix $[\mathbf{A}\,\mathbf{b}]$. If this matrix is transformed into its RREF, the following may be deduced:

1. If $[\mathbf{A}\,\mathbf{b}]$ is derived from an inconsistent system (i.e., no solution exists) the RREF has a row of the form $[0 \ldots 0\ 1]$.

2. If [**A b**] is derived from a consistent system with an infinity of solutions, then the number of columns of the coefficient matrix is greater than the number of nonzero rows in the RREF; otherwise there is a unique solution and it appears in the last (augmented) column of the RREF.
3. A zero row in the RREF indicates that the original set of equations contained equations with redundant information, that is, information contained in other equations of the system.

In computing the RREF, numerical problems can arise that are common to other procedures that use elementary row operations (see Section 2.6).

A.12 Differentiating Matrices

The rules for matrix differentiation are essentially the same as those for scalars, but care must be taken to ensure that the order of the matrix operations is maintained. The process is illustrated by the following example: differentiate $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$ with respect to each element of $\mathbf{x}$, where $\mathbf{x}$ is a column vector with n elements, $(x_1, x_2, x_3, \ldots, x_n)^\top$, and $\mathbf{A}$ has elements a_{ij} for $i, j = 1, 2, \ldots, n$. We note first that any matrix associated with a quadratic form must be symmetric. Hence the matrix $\mathbf{A}$ is symmetric. We require the gradient of $f(\mathbf{x})$ (i.e., $\nabla f(\mathbf{x})$). The gradient consists of all the first-order partial derivatives of $f(\mathbf{x})$ with respect to each component of the vector $\mathbf{x}$. Now multiplying out the terms of $f(\mathbf{x})$ we have $f(\mathbf{x})$ expressed in component terms as

$$f(\mathbf{x}) = \sum_{i=1}^{n} a_{ii} x_i^2 + \sum_{i=1}^{n} \sum_{\substack{j=1, \\ j \neq i}}^{n} a_{ij} x_i x_j$$

However, we note that since $\mathbf{A}$ is symmetric $a_{ij} = a_{ji}$ and consequently the terms $a_{ij}x_i x_j + a_{ji}x_i x_j$ can be written as $2a_{ij}x_i x_j$. Hence

$$\frac{\partial f(\mathbf{x})}{\partial x_k} = 2a_{kk}x_k + 2\sum_{\substack{j=1, \\ j \neq i}}^{n} a_{kj}x_j \quad \text{for } k = 1, 2, \ldots, n$$

This is of course equivalent to the matrix form

$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x}$$

and this provides the standard matrix result where $\mathbf{x}$ is a column vector.

A.13 Square Root of a Matrix

In order to have a square root, a matrix must be square. If $\mathbf{A}$ is a square matrix and $\mathbf{B}\mathbf{B} = \mathbf{A}$, then $\mathbf{B}$ is the square root of $\mathbf{A}$. If $\mathbf{A}$ is singular, it may not have a square root.

The square matrix $\mathbf{A}$ can be factorized to give $\mathbf{A} = \mathbf{X}\mathbf{D}\mathbf{X}^{-1}$ where $\mathbf{D}$ is a diagonal matrix comprising the n eigenvalues of $\mathbf{A}$, and $\mathbf{X}$ is an $n \times n$ array of the eigenvectors of $\mathbf{A}$. We can expand this expression for $\mathbf{A}$ to give

$$\mathbf{A} = (\mathbf{X}\mathbf{D}^{1/2}\mathbf{X}^{-1})(\mathbf{X}\mathbf{D}^{1/2}\mathbf{X}^{-1})$$

Since

$$\mathbf{A} = \mathbf{B}\mathbf{B}$$

then

$$\mathbf{B} = \mathbf{X}\mathbf{D}^{1/2}\mathbf{X}^{-1}$$

The square root of the diagonal matrix of eigenvalues, $\mathbf{D}$, is determined by taking the square root of each diagonal element, that is, each eigenvalue. Any number, real or complex, will have one positive and one negative square root. Thus to determine the square root of $\mathbf{D}$ (and hence $\mathbf{A}$) we must consider every combination of the positive and negative square roots of the eigenvalues. This gives 2^n possible combinations and hence there are 2^n expressions for $\mathbf{D}^{1/2}$. This will lead to 2^n different square root matrices, $\mathbf{B}$. If $\mathbf{D}^{1/2}$ comprises all the positive roots then the resulting square root matrix is called the principal square root. This matrix is unique.

Consider the following example. If

$$\mathbf{A} = \begin{bmatrix} 31 & 37 & 34 \\ 55 & 67 & 64 \\ 91 & 115 & 118 \end{bmatrix}$$

then, taking the $2^3 = 8$ combinations of square roots, we obtain the following square roots of $\mathbf{A}$. Note that $\mathbf{B}_0$ is the principal square root.

$$\mathbf{B}_0 = \begin{bmatrix} 2.9798 & 2.9296 & 1.8721 \\ 4.3357 & 5.0865 & 3.9804 \\ 5.0313 & 7.1413 & 8.9530 \end{bmatrix} \quad \mathbf{B}_1 = \begin{bmatrix} 1.0000 & 2.0000 & 3.0000 \\ 3.0000 & 4.0000 & 5.0000 \\ 8.0000 & 9.0000 & 7.0000 \end{bmatrix}$$

$$\mathbf{B}_2 = \begin{bmatrix} 2.8115 & 3.0713 & 1.8437 \\ 4.5426 & 4.9123 & 4.0153 \\ 4.9594 & 7.2019 & 8.9408 \end{bmatrix} \quad \mathbf{B}_3 = \begin{bmatrix} 1.1683 & 1.8583 & 3.0284 \\ 2.7931 & 4.1742 & 4.9651 \\ 8.0719 & 8.9395 & 7.0121 \end{bmatrix}$$

The negative of these matrices give a further four square roots of $\mathbf{A}$. Multiplying any one of these matrices by itself will result in the original matrix $\mathbf{A}$.