Appendix C

A summary of matrix algebra

In multivariate data analysis, knowledge of the fundamentals of matrix algebra is inevitable even though most biologists are concerned with the methods only at the 'user' level. Correct interpretation of results is impossible without understanding basic theory which, in turn, necessitates some background of mathematics, including matrix algebra. The terms and concepts most useful to us are presented in didactic order which implies that the following presentation becomes more and more difficult to digest towards the end. The examples discussed in the book are not repeated here, reference is made to them where they are relevant.

Basic terms

Matrix: Rectangular ('tabular') array of $n \times m$ quantities of any kind, the *elements* of the matrix. The array is put into brackets (sometimes round parentheses). n is the number of rows, m is the number of columns. A matrix is usually denoted by a boldface capital letter with its size (order) also shown after or under the symbol or – most often – as a subscript. Short reference to a matrix may also be made by writing its general element in brackets. The following denotations are therefore equivalent and choice among them is a matter of taste:

$$\mathbf{A}_{n,m} = \mathbf{A}(n \times m) = \underset{n,m}{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2m} \\ \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{im} \\ \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nj} & \cdots & a_{nm} \end{bmatrix} = [a_{ij}]_{n \times m} \,. \qquad \text{(C1)}$$

The elements of a matrix are real numbers in most cases (as in the present book, see Matrix 2.1), but functions, vectors or even matrices may also be arranged into matrices. Two matrices are of the same order if they agree in the number of rows and columns (be warned that the order is neither 'dimensionality' nor 'rank'!)

Vector: Special matrix with only one row and m columns (row vector) or with n rows and one column (column vector). A vector is conveniently denoted by a boldface lowercase letter. A *column vector* may be presented in the following form

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_i \\ \cdots \\ a_n \end{bmatrix}, \tag{C2}$$

whereas its corresponding row vector (now $n = m$) is given by:

$$\mathbf{a}' = \begin{bmatrix} a_1 \; a_2 \; \dots \; a_j \; \dots \; a_m \end{bmatrix} . \tag{C3}$$

The prime $'$ indicates that a row vector is in fact the transpose of the column vector (see below). To avoid confusions, a prime must always be attached to the symbol of row vectors.

Scalar: A matrix containing only one row and one column is termed a *scalar*, a single number, denoted by italicized lowercase letters, such as a.

Some special matrices

Square matrix: In this, the number of rows agrees with the number of columns ($n = m$). A matrix of order n incorporates n^2 elements. All distance, correlation and similarity matrices are square. The elements from the upper left to the lower right (a_{ii}) constitute the *main (principal) diagonal* of the matrix.

Diagonal matrix: It is a square matrix with all off-diagonal values being zero (zeros may of course appear in the diagonal as well). For example, the following matrix contains eigenvalues in the main diagonal:

$$\mathbf{L}_{n,n} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 & \dots & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ 0 & 0 & \dots & \lambda_i & \dots & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ 0 & 0 & \dots & 0 & \dots & \lambda_n \end{bmatrix}, \tag{C4}$$

and is often abbreviated by diag (λ_{ii}) or $<\lambda_1, \lambda_2, ..., \lambda_i, ..., \lambda_n>$. Many textbooks use the symbol **D** for diagonal matrices, but in the present book this symbol was reserved for distance (dissimilarity) matrices.

Identity matrix: It is a diagonal matrix with one-s in the main diagonal. The *identity vector* has a single 1 value while all others are zero. Examples are

$$\mathbf{I}_{n,n} = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ 0 & 0 & \dots & 1 & \dots & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{i}' = [0 \; 1 \; \dots \; 0 \; \dots \; 0] . \tag{C5a-b}$$

These constructs should not be confused with the unit matrix and unit vector, respectively, because in these cases all elements are 1. For example:

$$\mathbf{1}'_5 = [\; 1 \; 1 \; 1 \; 1 \; 1 \;] . \tag{C6}$$

The role of the identity matrix in matrix algebra is analogous to the role of the number 1 in conventional arithmetics.

Null matrix and null vector: All of their elements are zero. For simplicity, both are usually abbreviated as **0**.

Partitioned matrices: A given matrix may be divided into smaller sub-matrices when, for examaple, its rows and/or columns can be arranged into groups. A partitioned correlation matrix is 7.17.

Transposition: The transpose of matrix **A** is another matrix, **A′**, obtained by interchanging the rows and the columns of **A**. For example,

$$\mathbf{A}_{6,4} = \begin{bmatrix} 3 & 3 & 4 & 12 \\ 0 & 7 & 8 & 7 \\ 4 & 8 & 4 & 3 \\ 2 & 10 & 1 & 9 \\ 6 & 2 & 9 & 5 \\ 10 & 5 & 6 & 11 \end{bmatrix} \quad \text{and its transpose} \quad \mathbf{A}'_{4,6} = \begin{bmatrix} 3 & 0 & 4 & 2 & 6 & 10 \\ 3 & 7 & 8 & 10 & 2 & 5 \\ 4 & 8 & 4 & 1 & 9 & 6 \\ 12 & 7 & 3 & 9 & 5 & 11 \end{bmatrix}. \qquad \text{(C7)}$$

A transposed matrix may also be abbreviated as $\mathbf{A}^*$ or $\mathbf{A}^{\mathrm{T}}$. The transpose of a column vector is a row vector and vice versa. To save space, a column vector is most often presented as its transpose (as in C3).

Symmetric matrices: If a matrix **A** is identical to its transpose, that is, $\mathbf{A} = \mathbf{A}'$ or $a_{ij} = a_{ji}$, then it is called to be *symmetric.* Obviously, only square matrices can be mentioned in this context. Examples are the covariance, correlation and distance matrices, i.e., most resemblance matrices, provided that the resemblance function itself obeys the symmetry axiom. (In *asymmetric matrices,* $a_{ij} \neq a_{ji}$.) When printed or typed, it is sufficient to show the main diagonal and the lower semimatrix (as in matrix 3.2).

Antisymmetric matrices: In these matrices, the relationship $a_{ij} = -a_{ji}$ holds for all values off the principal diagonal.

Trace of a matrix: This is the sum of the values in the main diagonal of an $\mathbf{A}_{n,n}$ square matrix. It is denoted by tr(**A**):

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} . \qquad \text{(C8)}$$

Operations with matrices

Arithmetic operations do not apply directly to matrices and some of them do not 'work' at all. In addition, most of the matrix operations are defined only if the matrices satisfy the *conformity* conditions required by the given operation.

Equality of matrices: Two matrices are equal if and only if they are of the same order and all corresponding elements are equal. In other words, $\mathbf{A} = \mathbf{B}$, if $a_{ij} = b_{ij}$ for all i and j.

Addition and subtraction: These operations apply elementwise to matrices of the same order. That is,

$$\mathbf{C} = \mathbf{A} + \mathbf{B} , \quad \text{where} \quad c_{ij} = a_{ij} + b_{ij} . \qquad \text{(C9)}$$

Multiplication by a scalar: The product of a matrix **A** by a scalar quantity b is obtained by multiplying all values of **A** with b.

Scalar product of two vectors: If a row vector and a column vector have the same number of elements, then the sum of the products of the corresponding elements is the *scalar product* of the vectors. More formally,

$$\mathbf{a}'\,\mathbf{b} = [a_1, a_2, \ldots, a_i, \ldots, a_n] \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_i \\ \cdots \\ b_n \end{bmatrix} = \sum_{i=1}^{n} a_i\, b_i \,. \qquad \text{(C10)}$$

When writing matrix equations, we must bear in mind that the order of terms is always important! In this case, reverse ordering corresponds to the next operation.

Matrix product of two vectors: This is obtained from a column vector post-multiplied by a row vector, resulting in a matrix **C** whose general element c_{ij} is the product of the i-th element of the column vector and the j-th element of the row vector. That is:

$$\mathbf{a}\,\mathbf{b}' = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_i \\ \cdots \\ a_n \end{bmatrix} [b_1, b_2, \ldots, b_j, \ldots, b_m] = \begin{bmatrix} a_1b_1 & a_1b_2 & \ldots & a_1b_j & \ldots & a_1b_m \\ a_2b_1 & a_2b_2 & \ldots & a_2b_j & \ldots & a_2b_m \\ . & . & \ldots & . & \ldots & . \\ a_ib_1 & a_ib_2 & \ldots & a_ib_j & \ldots & a_ib_m \\ . & . & \ldots & . & \ldots & . \\ a_nb_1 & a_nb_2 & \ldots & a_nb_j & \cdots & a_nb_m \end{bmatrix} . \qquad \text{(C11)}$$

The number of elements in the two vectors is not required to be the same. If $n = m$, then **C** is a square matrix.

The product of two matrices: Two matrices can only be multiplied by each other if the number of columns in the left matrix equals the number of rows in the right matrix, i.e., the two matrices are conformable in that order. This was the case above for both types of vectorial products. Assuming that the two matrices are $\mathbf{A}_{n,m}$ and $\mathbf{B}_{m,p}$, the product matrix **C** will be of order n,p. This is illustrated in Figure C1. An element c_{ij} of this matrix is obtained as the scalar product of the i-th row vector of **A** and the j-th column vector of **B**:

$$\mathbf{C}{=}\mathbf{AB} \;\; \text{such that} \;\; c_{ij} = \sum_{k=1}^{m} a_{ik}\, b_{kj} \,. \qquad \text{(C12)}$$

It follows from the above definition that **A** and **B** can be multiplied by each other in both directions if they are of the same order. In general, multiplication of matrices is not a commutative operation: **AB** ≠ **BA.** (There are matrices whose multiplication is commutative, e.g., the diagonal matrices).

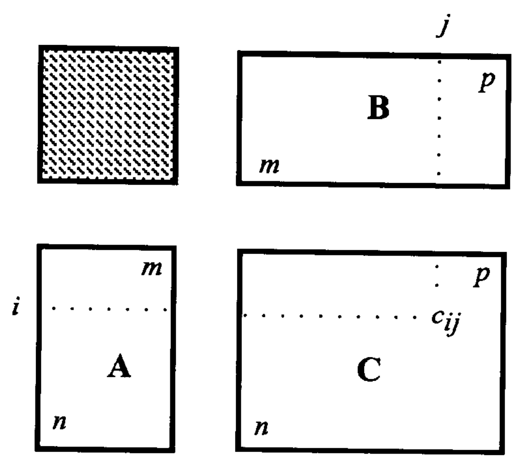

Figure C1. Illustrating matrix multiplication. If the rectangles are proportional to the order of the matrices **A** and **B** to be multiplied, then the upper left area is a square and the product **C** will correspond to the rectangle on the lower right. Dots indicate the rows and columns from which c_{ij} is derived.

Two important relationships are mentioned without further comment:

$$\mathbf{IA} = \mathbf{AI} = \mathbf{A}\;; \qquad \text{(C13a)}$$

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'. \qquad \text{(C13b)}$$

If, for example, $\mathbf{X}_{n,m}$ is a data matrix with n variables and m observations, then matrix $\mathbf{C}_{n,n} = \mathbf{XX}'$ is the cross product matrix of the variables. From centred data, the same operation followed by multiplication with the scalar $1/(m-1)$ yields the variance/covariance matrix.

Multiplication of several matrices: A series of matrices can be multiplied if, from left to right, each pair of matrices is conformable, such as $\mathbf{A}_{n,m}\,\mathbf{B}_{m,p}\,\mathbf{C}_{p,q}$. Since every square matrix is conformable with itself, they can be raised to any power, $\mathbf{A}^k$, where k is a positive integer. By definition, $\mathbf{A}^0 = \mathbf{I}$. Special cases are the projector or *idempotent* matrices: their any power is identical to the matrix itself, i.e., $\mathbf{A} = \mathbf{A}^k$, $k = 1, 2, ...$ The identity matrix is an example of idempotent matrices.

Multiplication of a matrix by a vector: Its result can only be another vector, as obvious from the above definitions. A matrix can be pre-multiplied by a row vector only, and the result is also a row vector (e.g., $\mathbf{a}'_{1,n}\,\mathbf{X}_{n,m} = \mathbf{b}'_{1,m}$). A matrix can only be post-multiplied by a column vector, with the result being another column vector ($\mathbf{X}_{n,m}\,\mathbf{a}_{m,1} = \mathbf{b}_{n,1}$). Pre- and post-multiplication of a square matrix with vectors containing the same elements will produce a scalar. This is called the *quadratic form*: $Q(\mathbf{A}) = \mathbf{x}'\mathbf{Ax}$ (examples in Formulae 3.3 and 7.19).

Geometric interpretations

There is a very close relationship between matrix algebra and coordinate geometry. As mentioned in Chapter 2, a data matrix can be represented in two different coordinate systems. In addition, the following definitions are extremely helpful in understanding the meaning of several formulations presented in this book.

Vector length, normalization of vector: An $\mathbf{x}_j$ column vector of data matrix $\mathbf{X}_{n,m}$ 'points' to the position of object j in the data space. The length of this vector is obtained from the Pythagorean theorem as follows:

$$|\,\mathbf{x}_j\,| = (\mathbf{x}'_j\,\mathbf{x}_j)^{1/2} = \sqrt{\sum_{i=1}^{n} x_{ij}^2}\,. \qquad \text{(C14)}$$

This quantity may also be called the 'absolute value' of the vector. Normalization of a vector implies its standardization to unit length, achieved by dividing each value by vector length. Consequently, the sum of squared elements of a normalized vector, i.e., its scalar product by itself, is 1.

The angle between two vectors: The cosine of the angle between vectors $\mathbf{x}_j$ and $\mathbf{x}_k$ is their scalar product divided by the product of their lengths:

$$\cos\theta_{jk} = \frac{\mathbf{x}'_j\,\mathbf{x}_k}{(\mathbf{x}'_j\,\mathbf{x}_j)^{1/2}\,(\mathbf{x}'_k\,\mathbf{x}_k)^{1/2}}\,. \qquad \text{(C15)}$$

This appears in the formulation of chord distance (3.54). If the cosine of the angle is 1, then the two vectors coincide (collinear) whereas at the other extreme with the cosine being zero, the two vectors are at right angle (orthogonal) to each other. Thus, the condition of orthogonality for two vectors is that their scalar product, the denominator of Formula C15, is zero.

Orthogonal matrices: Matrix $\mathbf{X}_{n,m}$ is orthogonal by columns if all possible pairs of column vectors are orthogonal (Reyment & Jöreskog 1993). As a consequence, the product

$$\mathbf{X'X} = \mathbf{D} \tag{C16}$$

results in a diagonal matrix. For example,

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & -3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 14 \end{bmatrix}. \tag{C17}$$

In the strictest sense, orthogonality is reserved for square matrices. When orthogonality holds, both products of the matrix with its transpose yield the same diagonal matrix:

$$\mathbf{A'A} = \mathbf{AA'} = \mathbf{D} \tag{C18}$$

Orthonormal matrices: An identity matrix resulting on the right side of Equation C16 indicates that all column vectors of matrix **X** are normalized and orthogonal in every possible pair. This special case is called *orthonormality.* For square orthonormal matrices we have

$$\mathbf{A'A} = \mathbf{AA'} = \mathbf{I}\,. \tag{C19}$$

It is confusing that the term orthogonality is sometimes restricted to orthonormal matrices.

A typical orthonormal matrix is the *rotation matrix* which takes the following form for n=2:

$$\mathbf{P}_{2,2} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}. \tag{C20}$$

A data matrix $\mathbf{X}_{2,m}$ premultiplied by the rotation matrix causes the whole coordinate system to be rotated rigidly around the origin by a degree of φ in a clock-wise direction:

$$\mathbf{Y}_{2,m} = \mathbf{P}_{2,2}\,\mathbf{X}_{2,m}\ . \tag{C21}$$

Rotation matrices are used, for example, in Procrustes analysis (Subsection 9.2.4).

Euclidean distance: The distance between points j and k may be calculated as the difference between the two corresponding vectors:

$$d_{jk} = |\,\mathbf{x}_j - \mathbf{x}_k\,| = [(\mathbf{x}_j - \mathbf{x}_k)'(\mathbf{x}_j - \mathbf{x}_k)]^{½}\,. \tag{C22}$$

This is in fact Formula 3.47 written in matrix algebraic terms.

Determinants

A definition: The determinants of square matrices are central in importance for matrix algebra. The determinant is a unique number, a scalar, with a distant analogy to the absolute value of real numbers, hence abbreviated in similar way:

$$|\mathbf{A}| = \det(\mathbf{A})\,. \tag{C23}$$

The determinant of a scalar is itself. The determinant of a 2×2 matrix is determined according to the following relationship:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}\,a_{22} - a_{21}\,a_{12}\,. \tag{C24}$$

For matrices of order three, the determinants can only be calculated using the so-called minors according to the formula:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \tag{C25}$$

Finding the determinant of matrices of higher order requires the computer; manual calculations become extremely cumbersome when the order increases. The geometric meaning of determinants is straightforward. For example, the following quantity

$$\begin{vmatrix} 5 & 0 \\ 0 & 5 \end{vmatrix} = 25 - 0 = 25 \qquad \text{(C26)}$$

corresponds to the area of a square determined by the two column vectors and two line segments drawn parallel to the vectors (dashed lines in Fig. C2a). In general, the absolute value of the determinant of a 2×2 matrix indicates an area of a parallelogram (Fig. C2b), that of a 3×3 matrix corresponds to a volume, and for higher orders we speak of hypervolumes of dimensionality n.

Properties of determinants: The following list describes some fundamental relationships without further comments:

1) The determinant of a matrix is identical to the determinant of its transpose:

$$|\mathbf{A}| = |\mathbf{A}'| \qquad \text{(C27)}$$

2) If two neighboring rows (or columns) are interchanged in the matrix, then the determinant changes its sign, while its absolute value does not change. For example,

$$\begin{vmatrix} 5 & 3 \\ 2 & 5 \end{vmatrix} = 25 - 6 = 19, \quad \text{and} \quad \begin{vmatrix} 3 & 5 \\ 5 & 2 \end{vmatrix} = 6 - 25 = -19 \qquad \text{(C28a-b)}$$

(the case on the left is shown geometrically in Fig. C2b).

3) If a column (or a row) is multiplied by a scalar c, then the determinant of the new matrix is also c times the determinant of the original matrix.

4) The determinant of the product of two matrices (of the same order, of course) equals the product of the original determinants:

$$|\mathbf{AB}| = |\mathbf{A}|\,|\mathbf{B}|.$$

5) The determinant of a diagonal matrix is the product of the diagonal values:

$$|\mathbf{D}| = \prod d_{ii} \qquad \text{(C29)}$$

(such as the generalized variance, cf. Section 7.1).

6) If matrix $\mathbf{A}$ of order n is multiplied by a scalar c, then the determinant will become:

$$|c\mathbf{A}| = c^n|\mathbf{A}|. \qquad \text{(C30)}$$

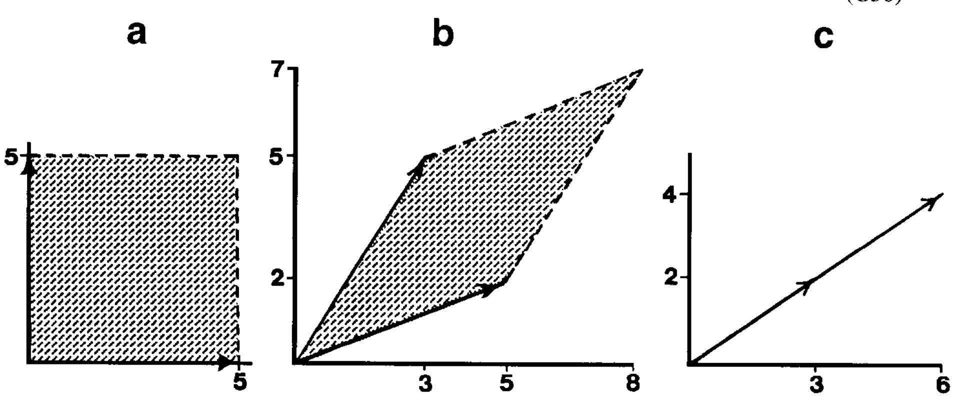

Figure C2. Geometric illustration of determinants in two dimensions. **a:** Positive determinant from Equation C26, corresponding to a square, **b:** the determinant for Equation C28a is the area of a parallelogram, **c:** illustration of a zero determinant (Equation C31), i.e., a zero area.

7) The determinant of orthonormal matrices is $|\mathbf{P}| = \pm 1$.

Singular matrices: The notion of singularity is closely related to the determinants. In multivariate analysis, singularity is perhaps more important than any other property mentioned above. The definition is simple: a matrix is singular if its determinant is zero. Singularity results if one column (or a row) of the matrix is obtained by multiplying another column (or row) of the matrix by a scalar. For example,

$$\begin{vmatrix} 6 & 3 \\ 4 & 2 \end{vmatrix} = 12 - 12 = 0\,. \qquad \text{(C31)}$$

Its illustration is presented in Fig. C2c. As seen, there is zero area because the two column vectors are collinear. For matrices of higher order, singularity means that the hypervolume for n dimensions becomes 0, although it may still be nonzero in a subspace of lower dimensionality. Singularity therefore means something like that the 'matrix does not represent a truly n-dimensional space', but for further clarification, see the discussion of the inverse.

Minor: A minor of square matrix $\mathbf{A}$ is obtained by deleting column j and row i. In Equation C25, we used three possible minors of a matrix, each derived by discarding the first row and one of the columns.

In multivariate analysis, minors obtained from symmetric matrices by the removal of a row and its corresponding column are of central importance. If all such minors of matrix $\mathbf{A}$ have positive determinants, then this matrix is termed to be *positive definite.* If all such determinants are non-negative (i.e., zero is allowed), then we speak of a *positive semi-definite* matrix. These are denoted by $\mathbf{A} > 0$, and $\mathbf{A} \geq 0$, respectively.

Inversion

Division was not mentioned among matrix operations, because it is undefined in matrix algebra. In arithmetics – as we know – division by a non-zero number can also be written as a multiplication by the reciprocal value of the divisor. Also, the reciprocal value multiplied by the divisor itself yields 1. This latter procedure, as we shall see below, can be generalized to some matrices via *matrix inversion.*

The definition of inverse: The inverse of matrix $\mathbf{A}$ is denoted by $\mathbf{A}^{-1}$ and satisfies the criteria:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}\;. \qquad \text{(C32)}$$

This formula immediately shows the analogous relationship with the reciprocal value. As mentioned earlier, the role of the identity matrix $\mathbf{I}$ is the same in matrix algebra as the role of 1 in arithmetics. Whereas all scalars but zero have reciprocal value, there are an infinite number of matrices that do not have inverse. These are the *singular* matrices. (Matrix inversion is applied in this book, for example, in Formula 7.68.)

Inverting matrices is a computer-oriented task, so we do not present any example of calculation here. Instead, some fundamental relationships are summarized.

1) The inverse of diagonal matrices is another diagonal matrix whose diagonal elements are the reciprocal values of the original diagonal elements. That is, $\mathbf{D}^{-1} = \text{diag}\,(1 / d_{ii})$. Inverted diagonal matrices are used to divide the elements of diagonal matrices, as in Equations 7.36-7.40. (Note that $\mathbf{D}^{-1/2} = \text{diag}\,(1 / \sqrt{d_{ii}})$, and $\mathbf{D}^{1/2} = \text{diag}\,(\sqrt{d_{ii}})$).

2) Since the product of an orthonormal matrix and its transpose produces **I** (Formula C19), the transpose of the orthonormal matrices is identical to their inverse: $\mathbf{P}' = \mathbf{P}^{-1}$. This simplifies greatly the calculations.

Generalized inverse: For completeness, it is noted that there is another type of inversion, applicable to any matrix. Singularity and any 'shape' are allowed by the *generalized inverse,* denoted by $\mathbf{U}^{-}$ for matrix **U** and defined as $\mathbf{U}\mathbf{U}^{-}\mathbf{U} = \mathbf{U}$. If **A** is not singular, then $\mathbf{A}^{-1} = \mathbf{A}^{-}$.

'Interior structure' of matrices

Decomposition into the product of two matrices: As we have seen above, the product of conformable matrices $\mathbf{A}_{n,m}$ and $\mathbf{B}_{m,p}$ yields matrix $\mathbf{C}_{n,p}$. Now, let us consider the backwards operation. Scalars can be written as a product of two real numbers in an infinite number of ways; for example, 24 is obtained as 2×12, 4×6, 0.5×48 and so on. On the analogue of this, any matrix can also be generated as a product of two conformable matrices in an infinite number of ways. The importance of this fact will become obvious later.

Linear independence: The relationships between column (or row) vectors of a matrix can be characterized in terms of *linear combinations*. A linear combination of vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_i, ..., \mathbf{v}_m$ is another vector **y** obtained as:

$$\mathbf{y} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + ... + a_i\mathbf{v}_i + ... + a_m\mathbf{v}_m \ . \qquad \text{(C33)}$$

If **y** is obtained as the linear combination of vectors $\mathbf{v}_i$ only if all *scalars* a_i are zero, then **y** is said to be *linearly independent* from vectors $\mathbf{v}_i$. For example, if we have that $\mathbf{y}' = [1\ 0\ 0]$, $\mathbf{v}'1 = [\ 2\ 1\ 1]$ and $\mathbf{v}'2 = [\ 1\ 0\ 2\ ]$, then the condition C33 is satisfied only if $a_1 = a_2 = 0$. If a linear combination can be generated such that one or more scalars are non-zero, then we speak of linear dependence. Consequently, for a set of vectors, e.g., for all column vectors of $\mathbf{X}_{n,m}$, linear independence holds provided that

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + ... + a_i\mathbf{x}_i + ... + a_n\mathbf{x}_m = 0 \qquad \text{(C34)}$$

satisfies only if $a_i = 0$ for all i.

The rank of matrices: Having introduced the notion of linear independence, a fair definition of ranks is also possible. The rank of any matrix $\mathbf{X}_{n,m}$ is the smallest possible integer k for which **X** can be reproduced as the product of any $\mathbf{A}_{n,k}$ and $\mathbf{B}_{k,m}$. In other words, k is the minimum common order for all the possible pairs of matrices that can be used to reproduce **X** as a product. The rank is denoted by r(**X**).

Another useful definition follows. The rank of matrix $\mathbf{X}_{n,m}$ is the maximum number of linearly independent rows (or columns). $0 \leq \mathrm{r}(\mathbf{X}) \leq \min(n, m)$, best illustrated in the subsequent paragraph.

Intuitively, the rank of a matrix is interpreted as its inherent dimensionality. Let us consider the coordinate systems of Figure C3, illustrating vectors in two 'starting' dimensions. Fig. C3a shows column vectors of the following 2×3 data matrix (rows are dimensions):

$$\mathbf{X} = \begin{bmatrix} -2 & 2 & 3 \\ -4 & 4 & 6 \end{bmatrix} . \qquad \text{(C35)}$$

The vectors are collinear, i.e., all the three 'endpoints' fall onto the line running through the origin. Any vector can be written as the linear combination of the other two, for example, $\mathbf{x}_1 = -1/2\ \mathbf{x}_2 - 1/3\ \mathbf{x}_3$. As a consequence, the inherent dimensionality of the system is only 1. The conclusion is similar if the two row vectors are examined in a three-dimensional space, because they also coincide, that is there is only one background dimension (because $\mathbf{x}_1 = 1/2\ \mathbf{x}_2$). We can conclude that the vectors span a one-dimensional subspace. If we now consider the following small matrix:

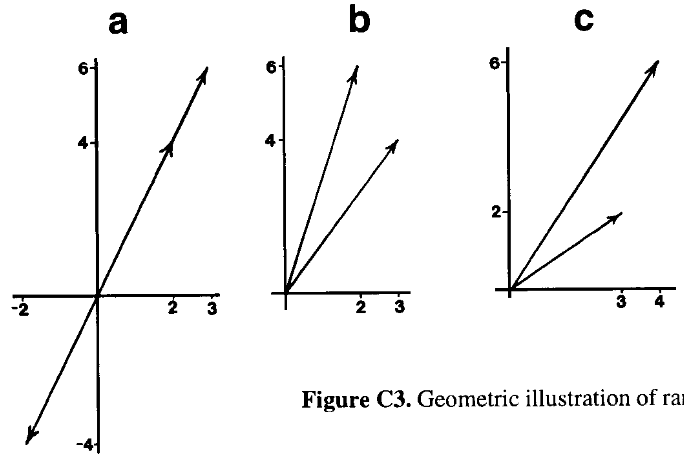

Figure C3. Geometric illustration of rank.

$$\mathbf{X} = \begin{bmatrix} 3 & 2 \\ 4 & 6 \end{bmatrix} \qquad \text{(C36)}$$

and take the rows (Fig. C3b) or the columns (Fig. C3c) as dimensions, then we see that the vectors are not collinear. The vectors can only be drawn on the plane, consequently matrix C36 has a rank of 2. Of course, no examples can be visualized for many dimensions, but we hope that the low-dimensional cases are convincing enough.

There are many important relationships concerning the rank of matrices:

1) $r(\mathbf{X}) = r(\mathbf{X}')$, (C37)

which is clear from the above geometric illustration.

2) $r(\mathbf{X}+\mathbf{Y}) \leq r(\mathbf{X}) + r(\mathbf{Y})$. (C38)

3) $r(\mathbf{XY}) \leq \min \{r(\mathbf{X}), r(\mathbf{Y})\}$. (C39)

4) In multivariate analysis, the following relationship is of central importance:

$r(\mathbf{X}'\mathbf{X}) = r(\mathbf{XX}') = r(\mathbf{X})$, (C40)

which means that the rank of the cross-product matrix obtained from matrix **X** and its transpose is identical to the rank of **X.** As a consequence, the rank of correlation and variance/covariance matrices equals the rank of centred data matrices.

5) The rank of matrix $\mathbf{A}_{n,n}$ is n if and only if the matrix is not singular; otherwise the rank is smaller. Based on minors, we can provide another definition of the rank, at least for square matrices. The rank of matrix $\mathbf{A}_{n,n}$ is the order of its largest non-singular submatrix. There will be yet another definition for symmetric matrices (see eigenvalues).

As an example, consider the cross products matrix $\mathbf{XX}'$ obtained from matrix C35:

$$\mathbf{XX}' = \begin{bmatrix} 17 & 34 \\ 34 & 68 \end{bmatrix} . \qquad \text{(C41)}$$

Its determinant is 17×68–34×34 = 1156–1156 = 0, that is, the matrix is singular, its dimensionality is lower than 2.

6) The rank of diagonal matrices is the number of non-zero elements in the main diagonal. Only null-matrices have a zero rank; all other matrices have positive rank.

Eigenvalues and eigenvectors of symmetric matrices: Our knowledge on the interior structure of symmetric matrices is completed by the discussion of eigenvalues and eigenvectors (the eigenanalysis of non-symmetric square matrices is not treated here, so that in the sequel **A** is understood as a symmetric matrix of order n). Eigenanalysis informs us about the rank of matrices and the relative importance of inherent dimensions, thus providing a deep insight into the hidden structure of data. The outset is that there exists a vector **v** and a scalar λ that satisfy the following equation:

$$\mathbf{Av} = \lambda\mathbf{v} \text{ , that is } \mathbf{Av} - \lambda\mathbf{v} = \mathbf{0}. \qquad \text{(C42a-b)}$$

Using the identity matrix, we have the following alternative expression:

$$(\mathbf{A} - \lambda\mathbf{I})\ \mathbf{v} = \mathbf{0}; \qquad \text{(C43)}$$

which is called the *characteristic equation* of matrix **A**. There are n+1 unknowns in this equation, the elements of the vector and λ. The equation is solved using the *characteristic determinant* given by

$$|\mathbf{A} - \lambda\mathbf{I}| = \mathbf{0} = q(\lambda). \qquad \text{(C44)}$$

This means that the rank of **A**–λ**I** is at least one smaller than r(**A**). q(λ) is a polynomial of order n, with n different roots. These roots are found by expanding the determinant, for which we have seen an example with $n = 2$ in Section 7.1. For large values of n, the solutions are readily determined by computer programs.

The roots of Equation C44 are called the *eigenvalues* (latent roots) of matrix **A**. They can be arranged in descending order, i.e., $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. In this way, the relative importance of the inherent dimensions of **A** is revealed. Some fundamental relationships concerning eigenvalues are as follows:

1) The number of positive eigenvalues is the rank of the matrix. Therefore, if there is at least one zero eigenvalue, then the matrix is singular. In general, the rank of $\mathbf{X}_{n,m}$ is the number of nonzero eigenvalues of the cross products matrix **X′X**.

2) $\sum_{i=1}^{n} \lambda_i = \mathrm{tr}(\mathbf{A})$. (C45)

3) $\prod_{i=1}^{n} \lambda_i = |\mathbf{A}|$. (C46)

The **v** eigenvectors carry the information on inherent dimensions. Using the eigenvalues, the eigenvectors are calculated according to Equation C43. However, there are an infinite number of solutions for a given eigenvalue: if $\mathbf{v}_1$ is an eigenvector, then $c\mathbf{v}_1$ is also an eigenvector, such that c is any scalar. By convention, normalized eigenvectors are determined, so that each will have a length of 1. These are still non-unique, because they can be multiplied by -1 (the direction is arbitrary). Each pair of eigenvectors is orthogonal and, as a consequence, the normalized eigenvectors are orthonormal.

If the eigenvalues are written into the main diagonal of the **L** diagonal matrix and the eigenvectors are the columns of matrix **V**, then Equation C42a is equivalent to writing:

$$\mathbf{AV} = \mathbf{VL} \qquad \text{(C47)}$$

Decomposition of a matrix into the product of three matrices: As mentioned above, any matrix can be written as a product of two conformable matrices in an infinite number of ways. It follows from C47 that matrix **A** may also be expressed as a product of three matrices. Since **VV′=I**, post-multiplication of both sides of Equation C47 by **V′** yields:

$$\mathbf{A\,V\,V'} = \mathbf{A\,I} = \mathbf{A} = \mathbf{V\,L\,V'} \qquad \text{(C48)}$$

In words, matrix **A** is reproduced if **L** is pre-multiplied by **V** and post-multiplied by **V′**. This is called the *spectral decomposition* of matrix **A**. An important consequence of this is that any correlation or covariance matrix can be perfectly reproduced using the eigenvalues and eigenvectors. If C48 is written in the following way:

$$\mathbf{A} = \lambda_1 \mathbf{v}_1\mathbf{v}'_1 + \lambda_2\mathbf{v}_2\mathbf{v}'_2 + \ldots + \lambda_n\mathbf{v}_n\mathbf{v}'_n \,, \qquad \text{(C49)}$$

then we see that matrix **A** is obtained as the weighted sum of n matrices, each with order and rank of one, being derived as the matrix product of an eigenvector by itself.

Any matrix $\mathbf{X}_{n,p}$ $(n \geq p)$ may be written as the product of three matrices according to the *singular value* decomposition (if $n<p$, then the following is valid for the transpose of the matrix). Assume that $r(\mathbf{X}) = p$, which is often the case. This decomposition is unique, i.e., there is only one solution for every matrix:

$$\mathbf{X} = \mathbf{USV}' \,. \qquad \text{(C50)}$$

In this, **U** is of the same order as **X** and its column vectors are orthonormal. **S** and **V** are square matrices of order p. **V** is also orthonormal, whereas **S** is a diagonal matrix whose diagonal elements are termed the *singular values* of matrix **X** and are arranged in descending order. **U** and **V** comprise the so-called singular vectors. $s_{ii}>0$ for all i (if the rank were less than p, then some singular values were zero). It directly follows that **X** can be written as the *linear combination* of p matrices each of rank 1:

$$\mathbf{X} = s_1\mathbf{u}_1\mathbf{v}_1' + s_2\mathbf{u}_2\mathbf{v}_2' + \ldots + s_p\mathbf{u}_p\mathbf{v}_p' \qquad \text{(C51)}$$

According to the Eckart - Young theorem, if the first k terms are added only, then the matrix thus obtained is the best least squares k-rank approximation to **X.** This fact is utilized in principal components analysis, in the Euclidean and Mahalanobis biplots.

There is a close relationship between the spectral and the singular value decomposition. Assume that $\mathbf{A} = \mathbf{X}'\mathbf{X}$. Then, from C50 we have that

$$\mathbf{X}'\mathbf{X} = (\mathbf{USV}')'(\mathbf{USV}') = \mathbf{VS}(\mathbf{U}'\mathbf{U})\mathbf{SV}' \,. \qquad \text{(C52)}$$

Since $\mathbf{U}'\mathbf{U} = \mathbf{I}$, and both **S** and $\mathbf{S}^2$ are diagonal matrices, rewriting the middle term of C52 will produce a formula similar to C48. It has two consequences:

1) The singular values are the square roots of the eigenvalues of **A**, that is $\mathbf{L}^{1/2} = \mathbf{S}$.

2) The eigenvectors of **A** are identical to the right singular vectors of **X**, as also expressed by the denotations (**V**).

It is noted that if s_i is a singular value of matrix **X**, then the corresponding singular value of matrix $c\mathbf{X}$ will be cs_i, where c is a scalar.