

---

## Chapter

# 2

# *Matrix algebra: a summary*

## 2.0 Matrix algebra

Matrix language is the algebraic form best suited to the present book. The following chapters will systematically use the flexible and synthetic formulation of *matrix algebra*, with which many ecologists are already acquainted.

There are many reasons why matrix algebra is especially well suited for ecology. The format of computer files, including spreadsheets, in which ecological *data sets* are now most often recorded, is a *matrix* format. The use of *matrix notation* thus provides an elegant and compact representation of ecological information and *matrix algebra* allows operations on whole data sets to be performed. Last but not least, *multidimensional methods*, discussed in following chapters, are nearly impossible to conceptualise and explain without resorting to matrix algebra.

Matrix algebra goes back more than one century: “After Sylvester had introduced matrices [...], it is Cayley who created their algebra [in 1851]” (translated from Bourbaki, 1960). Matrices are of great conceptual interest for theoretical formulations, but it is only with the increased use of *computers* that matrix algebra became truly popular with ecologists. The use of computers naturally enhances the use of matrix notation. Most scientific programming languages are adapted to matrix logic. All matrix operations described in this chapter can be carried out using advanced statistical languages such as R, S-PLUS<sup>®</sup> and MATLAB<sup>®</sup>.

Ecologists who are familiar with matrix algebra could read Sections 2.1 and 2.2 only, where the vocabulary and symbols used in the remainder of this book are defined. Other sections of Chapter 2 may then be consulted whenever necessary.

The present chapter is only a *summary* of matrix algebra. Readers looking for more complete presentations should consult Bronson (2011), where numerous exercises are found. Graybill (2001) and Gentle (2007) provide numerous applications in general

Table 2.1 Ecological data matrix.

Objects	Descriptors						
	$y_1$	$y_2$	$y_3$	...	$y_j$	...	$y_p$
$x_1$	$y_{11}$	$y_{12}$	$y_{13}$	...	$y_{1j}$	...	$y_{1p}$
$x_2$	$y_{21}$	$y_{22}$	$y_{23}$	...	$y_{2j}$	...	$y_{2p}$
$x_3$	$y_{31}$	$y_{32}$	$y_{33}$	...	$y_{3j}$	...	$y_{3p}$
.	.	.	.		.		.
.	.	.	.		.		.
.	.	.	.		.		.
$x_i$	$y_{i1}$	$y_{i2}$	$y_{i3}$	...	$y_{ij}$	...	$y_{ip}$
.	.	.	.		.		.
.	.	.	.		.		.
.	.	.	.		.		.
$x_n$	$y_{n1}$	$y_{n2}$	$y_{n3}$	...	$y_{nj}$	...	$y_{np}$

statistics. There are also a number of recent books, such as Vinod (2011), explaining how to use matrix algebra in R. The older book of Green & Carroll (1976) stresses the geometric interpretation of matrix operations commonly used in statistics.

2.1 The ecological data matrix

Descriptor  
Object

As explained in Section 1.4, ecological data are obtained as object-observations or sampling units, which are described by a set of state values corresponding to as many descriptors, or variables. Ecological data are generally recorded in a table where each column  $j$  corresponds to a descriptor  $y_j$  (species present in the sampling unit, physical or chemical variable, etc.) and each object  $i$  (sampling site, sampling unit, locality, observation) occupies one row. In each cell  $(i,j)$  of the table is found the state taken by object  $i$  for descriptor  $j$  (Table 2.1). Objects will be denoted by a boldface, lower-case letter  $x$ , with a subscript  $i$  varying form 1 to  $n$ , referring to object  $x_i$ . Similarly, descriptors will be denoted by a boldface, lower case letter  $y$  subscripted  $j$ , with  $j$  taking values from 1 to  $p$ , referring to descriptor  $y_j$ . When considering two sets of descriptors, members of the second set will generally have subscripts  $k$  from 1 to  $m$ .

---

Following the same logic, the different values in a data matrix will be denoted by a doubly-subscripted  $y$ , the first subscript designating the object being described and the second subscript the descriptor. For example,  $y_{83}$  is the value taken by object 8 for descriptor 3.

As mentioned in Section 1.4, it is not always obvious which are the objects and which are the descriptors. In ecology, for example, the different sampling sites (objects) may be studied with respect to the species found therein. In contrast, when studying the behaviour or taxonomy of organisms belonging to a given taxonomic group, the objects are the organisms themselves, whereas one of the descriptors could be the types of habitat found at different sampling sites. To unambiguously identify objects and descriptors, one must decide which is the variable defined *a priori* (i.e. the objects). When conducting field or laboratory observations, the variable defined *a priori* is totally left to the researcher, who decides how many observations will be included in the study. Thus, in the first example above, the researcher could choose the number of sampling sites needed to study their species composition. What is observed, then, are the descriptors, namely the different species present and possibly their abundances. Another approach to the same problem would be to ask which of the two sets of variables the researcher could theoretically increase to infinity; this identifies the variable defined *a priori*, or the objects. In the first example, it is the number of samples that could be increased at will — the samples are therefore the objects — whereas the number of species is limited and depends strictly on the ecological characteristics of the sampling sites. In the second example, the variable defined *a priori* corresponds to the organisms themselves, and one of their descriptors could be their different habitats (states).

The distinction between objects and descriptors is not only theoretical. One may analyse either the relationships among descriptors for the set of objects in the study (R mode analysis), or the relationships among objects given the set of descriptors (Q mode study). It will be shown that the mathematical techniques that are appropriate for studying relationships among objects are not the same as those for descriptors. For example, correlation coefficients can only be used for studying relationships among descriptors, which are vectors of data observed on samples extracted from populations with a theoretically infinite number of elements; vector lengths are actually limited by the sampling effort. It would be incorrect to use a correlation coefficient to study the relationship between two objects across the set of descriptors; other measures of association are available for that purpose (Section 7.3). Similarly, when using the methods of multidimensional analysis that will be described in this book, it is important to know which are the descriptors and which are the objects, in order to avoid methodological errors. The results of incorrectly conducted analyses — and there are unfortunately many in the literature — are not necessarily wrong because, in ecology, phenomena that are easily identified are usually sturdy enough to withstand considerable distortion. What is a pity, however, is that the more subtle phenomena, i.e. the very ones for which advanced numerical techniques are used, could very well not emerge at all from a study based on inappropriate methodology.

Linear  
algebra

The table of ecological data described above is an array of numbers known as a *matrix*. The branch of mathematics dealing with matrices is *linear algebra*.

Matrix **Y** is a rectangular, ordered array of numbers  $y_{ij}$ , set out in rows and columns as in Table 2.1:

$$\mathbf{Y} = [y_{ij}] = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1p} \\ y_{21} & y_{22} & \dots & y_{2p} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ y_{n1} & y_{n2} & \dots & y_{np} \end{bmatrix} \quad (2.1)$$

Order

There are  $n$  rows and  $p$  columns. When the *order* of the matrix (also known as its *dimensions* or *format*) must be stated, a matrix of order  $(n \times p)$ , which contains  $n \times p$  elements, is written  $\mathbf{Y}_{np}$ . As above, any given element of **Y** is denoted  $y_{ij}$ , where subscripts  $i$  and  $j$  identify the row and column, respectively (always in that conventional order).

In linear algebra, ordinary numbers are called *scalars*, to distinguish them from *matrices*.

The following notation will be used hereinafter: a matrix will be symbolised by a capital letter in boldface, such as **Y**. The same matrix could also be represented by its general element in italics and in brackets, such as  $[y_{ij}]$ , or alternatively by an enumeration of all its elements, also in italics and in brackets, as in eq. 2.1. Italics will only be used for algebraic symbols, not for actual numbers. Occasionally, other notations than brackets may be found in the literature, i.e.  $(y_{ij})$ ,  $(y_i^j)$ ,  $\{y_{ij}\}$ ,  $\|y_i^j\|$ , or  $\langle ij \rangle$ .

Any subset of a matrix can be explicitly recognized. In the above matrix (eq. 2.1), for example, the following submatrices could be considered:

Square  
matrix

$$\text{a square matrix} \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix}$$

$$\text{a row matrix} [y_{11} \ y_{12} \ \dots \ y_{1p}] , \text{ or a column matrix} \begin{bmatrix} y_{12} \\ y_{22} \\ \cdot \\ \cdot \\ \cdot \\ y_{n2} \end{bmatrix}$$

Matrix notation simplifies the writing of data sets. It also corresponds to the way computers work. Indeed, most programming languages are designed to input data as matrices (arrays) and manipulate them either directly or through a simple system of subscripts. This greatly simplifies programming the calculations. Accordingly, computer packages generally input data as matrices. In addition, many of the statistical models used in multidimensional analysis are based on linear algebra, as will be seen later. So, it is convenient to approach them with data already set in matrix format.

## 2.2 Association matrices

Two important matrices may be derived from the ecological data matrix: the association matrix among objects and the association matrix among descriptors. An association matrix is denoted  $\mathbf{A}$ , and its general element  $a_{ij}$ . Although Chapter 7 is entirely devoted to association matrices, it is important to mention them here in order to better understand the purpose of methods presented in the remainder of the present chapter.

Using data from matrix  $\mathbf{Y}$  (eq. 2.1), one may examine the relationship between the first two objects  $\mathbf{x}_1$  and  $\mathbf{x}_2$ . In order to do so, the first and second rows of matrix  $\mathbf{Y}$

$$\begin{bmatrix} y_{11} & y_{12} & \dots & y_{1p} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} y_{21} & y_{22} & \dots & y_{2p} \end{bmatrix}$$

are used to calculate a measure of association (similarity or distance: Chapter 7), to assess the degree of resemblance between the two objects. This measure, which quantifies the strength of the association between the two rows, is denoted  $a_{12}$ . In the same way, the association of  $\mathbf{x}_1$  with  $\mathbf{x}_3, \mathbf{x}_4, \dots, \mathbf{x}_p$ , can be calculated, as can also be calculated the association of  $\mathbf{x}_2$  with all other objects, and so on for all pairs of objects. The coefficients of association for all pairs of objects are then recorded in a table, ordered in such a way that they could be retrieved for further calculations. This table is the association matrix  $\mathbf{A}$  among objects:

$$\mathbf{A}_{nn} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \quad (2.2)$$

A most important characteristic of any association matrix is that it has a number of rows equal to its number of columns, this number being equal here to the number of objects  $n$ . The number of elements in the above square matrix is therefore  $n^2$ .

Similarly, one may wish to examine the relationships among descriptors. For the first two descriptors,  $\mathbf{y}_1$  and  $\mathbf{y}_2$ , the first and second columns of matrix  $\mathbf{Y}$

$$\begin{bmatrix} y_{11} \\ y_{21} \\ \cdot \\ \cdot \\ \cdot \\ y_{n1} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} y_{12} \\ y_{22} \\ \cdot \\ \cdot \\ \cdot \\ y_{n2} \end{bmatrix}$$

are used to calculate a measure of dependence (Chapter 7) which assesses the degree of association between the two descriptors. In the same way as for the objects,  $p \times p$  measures of association can be calculated among all pairs of descriptors and recorded in the following association matrix:

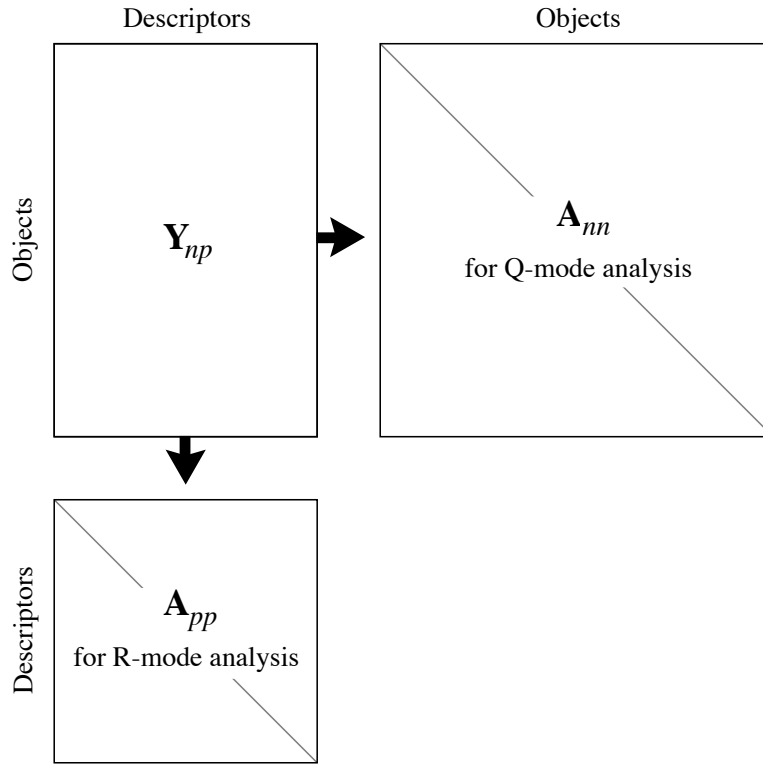
$$\mathbf{A}_{pp} = \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & \cdot & a_{1p} \\ a_{21} & a_{22} & \cdot & \cdot & \cdot & a_{2p} \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ a_{p1} & a_{p2} & \cdot & \cdot & \cdot & a_{pp} \end{bmatrix} \quad (2.3)$$

Association matrices are most often (but not always, see Section 2.3) symmetric, with elements in the upper right triangle being equal to those in the lower left triangle ( $a_{ij} = a_{ji}$ ). Elements  $a_{ii}$  on the diagonal measure the association of a row or a column of matrix  $\mathbf{Y}$  with itself. In the case of objects, the measure of association  $a_{ii}$  of an *object* with itself usually takes a value of either 1 (similarity coefficients) or 0 (distance coefficients). Concerning the association between *descriptors* (columns), the correlation  $a_{ii}$  of a descriptor with itself is 1, whereas the (co)variance provides an estimate  $a_{ii}$  of the variability among the values of descriptor  $i$ .

At this point of the discussion, it should thus be noted that the data, to which the models of multidimensional analysis are applied, are not only matrix  $\mathbf{Y}_{np}$  = [objects  $\times$  descriptors] (eq. 2.1), but also the two association matrices  $\mathbf{A}_{nn}$  = [objects  $\times$  objects] (eq. 2.2) and  $\mathbf{A}_{pp}$  = [descriptors  $\times$  descriptors] (eq. 2.3), as shown in Fig. 2.1.

## 2.3 Special matrices

Matrices with an equal number of rows and columns are called *square* matrices (Section 2.1). These, as will be seen in Sections 2.6 *et seq.*, are the only matrices for



**Figure 2.1** Data analysed in numerical ecology include matrix  $\mathbf{Y}_{np}$  = [objects  $\times$  descriptors] (eq. 2.1) as well as the two association matrices  $\mathbf{A}_{nn}$  = [objects  $\times$  objects] (eq. 2.2) and  $\mathbf{A}_{pp}$  = [descriptors  $\times$  descriptors] (eq. 2.3). The Q and R modes of analysis are defined in Section 7.1.

which it is possible to compute a determinant, an inverse, and eigenvalues and eigenvectors. As a corollary, these operations can be carried out on association matrices, which are square matrices.

Some definitions pertaining to square matrices now follow. In matrix  $\mathbf{B}_{nn}$ , of order  $(n \times n)$  (often called “square matrix of order  $n$ ” or “matrix of order  $n$ ”),

$$\mathbf{B}_{nn} = [b_{ij}] = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} \quad (2.4)$$

the *diagonal elements* are those with identical subscripts for the rows and columns ( $b_{ii}$ ). They are located on the *main diagonal* (simply called *the diagonal*) which, by convention, goes from the upper left to the lower right corners. The sum of the diagonal elements is called the *trace* of the matrix.

Trace

Diagonal matrix      A *diagonal matrix* is a square matrix where all non-diagonal elements are *zero*. Thus,

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

is a diagonal matrix. Diagonal matrices that contain on their diagonal values coming from a vector  $[x_i]$  are noted  $\mathbf{D}(x)$ . Special examples used later in the book are the diagonal matrix of standard deviations  $\mathbf{D}(\sigma)$ , the diagonal matrix of eigenvalues  $\mathbf{D}(\lambda_i)$ , also noted  $\mathbf{\Lambda}$ , and the diagonal matrix of singular values  $\mathbf{D}(w_i)$  also noted  $\mathbf{W}$ .

Identity matrix      A diagonal matrix where all diagonal elements are equal to unity is called a *unit matrix* or *identity matrix*. It is denoted  $\mathbf{D}(1)$  or  $\mathbf{I}$ :

$$\mathbf{D}(1) = \mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad (2.5)$$

This matrix plays the same role, in matrix algebra, as the number 1 in ordinary algebra, i.e. it is the neutral element in multiplication (e.g.  $\mathbf{IB} = \mathbf{B}$ , or  $\mathbf{BI} = \mathbf{B}$ ).

Scalar matrix      Similarly, a *scalar matrix* is a diagonal matrix of the form

$$\begin{bmatrix} 7 & 0 & \dots & 0 \\ 0 & 7 & \dots & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \dots & 7 \end{bmatrix} = 7\mathbf{I}$$

All the diagonal elements are identical since a scalar matrix is the unit matrix multiplied by a scalar (here, of value 7).



Null matrix      A matrix, square or rectangular, whose elements are all zero is called a *null matrix* or *zero matrix*. It is denoted **0** or [0].\*

Triangular matrix      A square matrix with all elements above (or below) the diagonal being zero is called a *lower* (or *upper*) *triangular matrix*. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$$

is an upper triangular matrix. These matrices are very important in matrix algebra because their determinant (Section 2.6) is equal to the product of all terms on the main diagonal (i.e. 24 in this example). Diagonal matrices are also triangular matrices.

Transpose      The *transpose* of a matrix **B** with format  $(n \times p)$  is denoted **B'** and is a new matrix of format  $(p \times n)$  in which  $b'_{ij} = b_{ji}$ . In other words, the rows of one matrix are the columns of the other. Thus, the transpose of matrix

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$$

is matrix

$$\mathbf{B}' = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix}$$

Transposition is an important operation in linear algebra, and also in ecology where a data matrix **Y** (eq. 2.1) may be transposed to study the relationships among descriptors after the relationships among objects have been analysed (or conversely).

---

\* Although the concept of zero was known to Babylonian and Mayan astronomers, inclusion of the zero in a decimal system of numeration finds its origin in India, in the eighth century A.D. at least (Ifrah, 1981). The ten Western-world numerals are also derived from the symbols used by ancient Indian mathematicians. The word *zero* comes from the Arabs, however. They used the word *sifr*, meaning “empty”, to refer to a symbol designating nothingness. The term turned into *cipher*, and came to denote not only zero, but all 10 numerals. *Sifr* is at the root of the medieval latin *zephirum*, which became *zefiro* in Italian and was then abbreviated to *zero*. It is also the root of the medieval latin *cifra*, which became *chiffre* in French where it designates any of the 10 numerals.

**Symmetric matrix** A square matrix that is identical to its transpose is *symmetric*. This is the case when corresponding terms  $b_{ij}$  and  $b_{ji}$ , on either side of the diagonal, are equal. For example,

$$\begin{bmatrix} 1 & 4 & 6 \\ 4 & 2 & 5 \\ 6 & 5 & 3 \end{bmatrix}$$

is symmetric since  $\mathbf{B}' = \mathbf{B}$ . All symmetric matrices are square.

**Non-symmetric matrix** It was mentioned in Section 2.2 that association matrices are generally symmetric. *Non-symmetric* (or *asymmetric*) matrices may be encountered, however. This happens, for example, when each coefficient in the matrix measures the ecological influence of an organism or a species on another, these influences being asymmetrical (e.g. A is a predator of B, B is a prey of A). Asymmetric matrices are also found in behaviour studies, serology, DNA pairing analysis, etc.

**Skew-symmetric matrix** Matrix algebra tells us that any *non-symmetric* matrix may be expressed as the sum of two other matrices, one *symmetric* and one *skew-symmetric*, without loss of information. Consider for instance the two numbers 1 and 3, found in opposite positions (1,2) and (2,1) of the first matrix in the following numerical example:

$$\begin{bmatrix} 1 & 1 & 2 & 2 \\ 3 & 1 & 0 & -1 \\ 1 & 2 & 1 & 0 \\ 0 & -4 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2.0 & 1.5 & 1.0 \\ 2.0 & 1 & 1.0 & -2.5 \\ 1.5 & 1.0 & 1 & 1.5 \\ 1.0 & -2.5 & 1.5 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -1.0 & 0.5 & 1.0 \\ 1.0 & 0 & -1.0 & 1.5 \\ -0.5 & 1.0 & 0 & -1.5 \\ -1.0 & -1.5 & 1.5 & 0 \end{bmatrix}$$

Non-symmetric      Symmetric (average)      Skew-symmetric

The *symmetric* part is obtained by averaging these two numbers:  $(1 + 3)/2 = 2.0$ . The *skew-symmetric* part is obtained by subtracting one from the other and dividing by 2:  $(1 - 3)/2 = -1.0$  and  $(3 - 1)/2 = +1.0$  so that, in the skew-symmetric matrix, corresponding elements on either side of the diagonal have the same absolute values but opposite signs. When the symmetric and skew-symmetric components are added, the result is the original matrix:  $2 - 1 = 1$  for the upper original number, and  $2 + 1 = 3$  for the lower one. Using letters instead of numbers, one can derive a simple algebraic proof of the additivity of the symmetric and skew-symmetric components. The symmetric component can be analysed using the methods applicable to symmetric matrices (for instance, metric or non-metric scaling, Sections 9.3 and 9.4), while analysis of the skew-symmetric component requires methods especially developed to assess asymmetric relationships. Basic references are Coleman (1964) in the field of sociometry and Digby & Kempton (1987, Ch. 6) in numerical ecology. An application to biological evolution is found in Casgrain *et al.* (1996). Relevant biological or ecological information may be found in the symmetric portion only and, in other instances, in the skew-symmetric component only.

## 2.4 Vectors and scaling

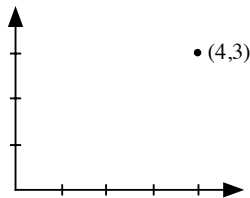
Another matrix of special interest is the *column matrix*, with format  $(n \times 1)$ , which is also known as a *vector*. Some textbooks restrict the term “vector” to *column matrices*, but the expression *row vector* (or simply *vector*, as used in some instances in Chapter 4) may also designate *row matrices*, with format  $(1 \times p)$ .

A (column) vector is noted as follows:

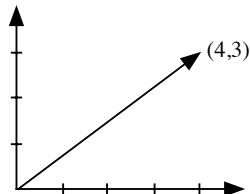
$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_n \end{bmatrix} \quad (2.6)$$

A vector graphically refers to a directed line segment. It also forms a mathematical entity on which operations can be performed. More formally, a vector is defined as an ordered  $n$ -tuple of real numbers, i.e. a set of  $n$  numbers with a specified order. The  $n$  numbers are the coordinates of a point in a  $n$ -dimensional Euclidean space, which may be seen as the end-point of a line segment starting at the origin.

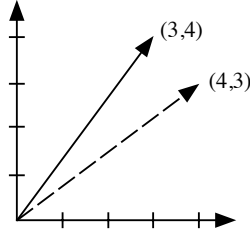
For example, (column) vector  $[4 \ 3]'$  is an ordered doublet (or 2-tuple) of two real numbers  $(4, 3)$ , which may be represented in a two-dimensional Euclidean space:



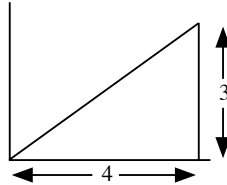
This same point  $(4, 3)$  may also be seen as the end-point of a line segment starting at the origin:



These figures illustrate the two possible representations of a vector; they also stress the *ordered* nature of vectors, since vector  $[3\ 4]'$  is different from vector  $[4\ 3]'$ .



Using the Pythagorean theorem, it is easy to calculate the length of any vector. For example, the length of vector  $[4\ 3]'$  is that of the hypotenuse of a right triangle with base 4 and height 3:



**Length Norm** The length (or *norm*) of vector  $[4\ 3]'$  is therefore  $\sqrt{4^2 + 3^2} = 5$ ; it is also the length (norm) of vector  $[3\ 4]'$ . The norm of vector  $b$  is noted  $\|b\|$ .

**Scaling Normalization** The comparison of different vectors, as to their directions, often requires an operation called *scaling*. In the scaled vector, all elements are divided by the same characteristic value. A special type of scaling is called *normalization*. In the normalized vector, each element is divided by the length of the vector:

normalization

$$\begin{bmatrix} 4 \\ 3 \end{bmatrix} \rightarrow \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix}$$

**Normalized vector** The importance of normalization lies in the fact that the length of a *normalized vector* is equal to unity. Indeed, the length of vector  $[4/5\ 3/5]'$ , calculated by means of the Pythagorean formula, is  $\sqrt{(4/5)^2 + (3/5)^2} = 1$ .

The example of doublet  $(4, 3)$  may be generalized to any  $n$ -tuple  $(b_1, b_2, \dots, b_n)$ , which specifies a vector in  $n$ -dimensional space. The length of the vector is  $\sqrt{b_1^2 + b_2^2 + \dots + b_n^2}$ , so that the corresponding normalized vector is:

$$\begin{bmatrix} b_1/\sqrt{b_1^2 + b_2^2 + \dots + b_n^2} \\ b_2/\sqrt{b_1^2 + b_2^2 + \dots + b_n^2} \\ \cdot \\ \cdot \\ \cdot \\ b_n/\sqrt{b_1^2 + b_2^2 + \dots + b_n^2} \end{bmatrix} = \frac{1}{\sqrt{b_1^2 + b_2^2 + \dots + b_n^2}} \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix} \quad (2.7)$$

The length of any normalized vector, in  $n$ -dimensional space, is 1.

## 2.5 Matrix addition and multiplication

Recording the data in table form, as is usually the case in ecology, opens the possibility of performing operations on these tables. The basic operations of matrix algebra (*algebra*, from the Arabic “al-jabr” which means *reduction*, is the theory of addition and multiplication) are very natural and familiar to ecologists.

**Numerical example.** Fish (3 species) were sampled at five sites in a lake, once a month during the summer (northern hemisphere). In order to get a general idea of the differences among sites, total numbers of fish caught at each site are calculated over the whole summer:

	July	August	September	Whole summer
Site 1	$\begin{bmatrix} 1 & 5 & 35 \end{bmatrix}$	$\begin{bmatrix} 15 & 23 & 10 \end{bmatrix}$	$\begin{bmatrix} 48 & 78 & 170 \end{bmatrix}$	$\begin{bmatrix} 64 & 106 & 215 \end{bmatrix}$
Site 2	$\begin{bmatrix} 14 & 2 & 0 \end{bmatrix}$	$\begin{bmatrix} 54 & 96 & 240 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 70 & 98 & 240 \end{bmatrix}$
Site 3	$\begin{bmatrix} 0 & 31 & 67 \end{bmatrix}$	$\begin{bmatrix} 0 & 3 & 9 \end{bmatrix}$	$\begin{bmatrix} 0 & 11 & 14 \end{bmatrix}$	$\begin{bmatrix} 0 & 45 & 90 \end{bmatrix}$
Site 4	$\begin{bmatrix} 96 & 110 & 78 \end{bmatrix}$	$\begin{bmatrix} 12 & 31 & 27 \end{bmatrix}$	$\begin{bmatrix} 25 & 13 & 12 \end{bmatrix}$	$\begin{bmatrix} 133 & 154 & 117 \end{bmatrix}$
Site 5	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 8 & 14 & 6 \end{bmatrix}$	$\begin{bmatrix} 131 & 96 & 43 \end{bmatrix}$	$\begin{bmatrix} 139 & 110 & 49 \end{bmatrix}$
	$sp1 \ sp2 \ sp3$	$sp1 \ sp2 \ sp3$	$sp1 \ sp2 \ sp3$	$sp1 \ sp2 \ sp3$

This operation is known as *matrix addition*. Note that only matrices of the same order can be added together. This is why, in the first matrix, site 5 was included with abundances of 0 to indicate that no fish had been caught there in July although site 5 had been sampled. Adding two matrices consists in a term-by-term addition. Matrix addition is associative and commutative; its neutral element is the null matrix **0**.

To study seasonal changes in fish productivity at each site, one possible approach would be to add together the terms in each row of each monthly matrix. However, this makes sense only if the selectivity of the fishing gear (say, a net) is comparable for the three species. Let us imagine that the efficiency of the net was 50% for species 2 and 25% for species 3 of what it was for species 1. In such a case, values in each row must be corrected before being added. Correction factors would be as follows: 1 for species 1, 2 for species 2, and 4 for species 3. To obtain

estimates of total fish abundances, correction vector  $[1 \ 2 \ 4]'$  is first multiplied by each row of each matrix, after which the resulting values are added. Thus, for the first site in July:

Site 1 July	Correction factors	Total fish abundance Site 1, July
$\begin{bmatrix} 1 & 5 & 35 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$	$(1 \times 1) + (5 \times 2) + (35 \times 4) = 1 + 10 + 140 = 151$

Scalar  
product

This operation is known in linear algebra as a *scalar product* because this product of two vectors produces a scalar.

In physics, there is another product of two vectors, called the *external* or *vector product*, where the multiplication of two vectors results in a third one, which is perpendicular to the plane formed by the first two. This product is not used in multidimensional analysis. It is however important to know that, in the literature, the expression “vector product” may be used for either that product or the scalar product of linear algebra, and that the scalar product is also called “inner product” or “dot product”. The vector product (of physics) is sometimes called “cross product”. This last expression is also used in linear algebra, for example in “matrix of sum of squares and cross products” (SSCP matrix), which refers to the product of a matrix with its transpose.

In matrix algebra, and unless otherwise specified, multiplication follows a convention that is illustrated by the scalar product above: in this *product* of a column vector *by* a row vector, the row vector *multiplies* the column vector or, which is equivalent, the column vector *is multiplied by* the row vector. This convention, which should be kept in mind, will be followed in the remainder of the book.

The result of a scalar product is a number, which is equal to the sum of the products of those elements with corresponding order numbers. The scalar product is designated by a dot, or is written  $\langle \mathbf{a}, \mathbf{b} \rangle$ , or else there is no sign between the two terms. For example:

$$\mathbf{b}'\mathbf{c} = \mathbf{b}' \cdot \mathbf{c} = \begin{bmatrix} b_1 & b_2 & \dots & b_p \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} = b_1c_1 + b_2c_2 + \dots + b_pc_p = \text{a scalar.} \quad (2.8)$$

The rules for computing scalar products are such that only vectors with the same numbers of elements can be multiplied.

In analytic geometry, it can be shown that the scalar product of two vectors obeys the relationship:

$$\mathbf{b}' \bullet \mathbf{c} = (\text{length of } \mathbf{b}) \times (\text{length of } \mathbf{c}) \times \cos \theta \quad (2.9)$$

Orthogonal  
vectors

When the angle between two vectors is  $\theta = 90^\circ$ , then  $\cos \theta = 0$  and the scalar product  $\mathbf{b}' \bullet \mathbf{c} = 0$ . As a consequence, two vectors whose scalar product is zero are *orthogonal* (i.e. at right angle). This property will be used in Section 2.9 to compute eigenvectors. A matrix whose (column) vectors are all orthogonal to one another is called *orthogonal*. For any pair of vectors  $\mathbf{b}$  and  $\mathbf{c}$  with values centred on their respective mean,  $\cos \theta = r(\mathbf{b}, \mathbf{c})$  where  $r$  is the correlation coefficient (eq. 4.7).

Gram-Schmidt orthogonalization is a procedure to make a vector  $\mathbf{c}$  orthogonal to a vector  $\mathbf{b}$  that has first been normalized (eq. 2.7);  $\mathbf{c}$  may have been normalized or not. The procedure consists of two steps: (1) compute the scalar product  $sp = \mathbf{b}'\mathbf{c}$ . (2) Make  $\mathbf{c}$  orthogonal to  $\mathbf{b}$  by computing  $\mathbf{c}_{\text{ortho}} = \mathbf{c} - sp\mathbf{b}$ . Proof that  $\mathbf{c}_{\text{ortho}}$  is orthogonal to  $\mathbf{b}$  is obtained by showing that  $\mathbf{b}'\mathbf{c}_{\text{ortho}} = 0$ :  $\mathbf{b}'\mathbf{c}_{\text{ortho}} = \mathbf{b}'(\mathbf{c} - sp\mathbf{b}) = \mathbf{b}'\mathbf{c} - sp\mathbf{b}'\mathbf{b}$ . Since  $\mathbf{b}'\mathbf{c} = sp$  and  $\mathbf{b}'\mathbf{b} = 1$  because  $\mathbf{b}$  has been normalized, one obtains  $sp - (sp \times 1) = 0$ . In this book, in the iterative procedures for ordination algorithms (Tables 9.5 and 9.8), Gram-Schmidt orthogonalization will be used in the step where the vectors of new ordination object scores are made orthogonal to previously found vectors.

**Numerical example.** Returning to the above example, it is possible to multiply each row of each monthly matrix with the correction vector (scalar product) in order to compare total monthly fish abundances. This operation, which is the *product of a vector by a matrix*, is a simple extension of the scalar product (eq. 2.8). The product of the July matrix  $\mathbf{B}$  with the correction vector  $\mathbf{c}$  is written as follows:

$$\begin{bmatrix} 1 & 5 & 35 \\ 14 & 2 & 0 \\ 0 & 31 & 67 \\ 96 & 110 & 78 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 1(1) + 5(2) + 35(4) \\ 14(1) + 2(2) + 0(4) \\ 0(1) + 31(2) + 67(4) \\ 96(1) + 110(2) + 78(4) \\ 0(1) + 0(2) + 0(4) \end{bmatrix} = \begin{bmatrix} 151 \\ 18 \\ 330 \\ 628 \\ 0 \end{bmatrix}$$

The product of a vector by a matrix involves calculating, for each row of matrix  $\mathbf{B}$ , a scalar product with vector  $\mathbf{c}$ . Such a product of a vector by a matrix is only possible if the number of *elements in the vector* is the same as the number of *columns in the matrix*. The result is no longer a scalar, but a column vector with dimension equal to the number of rows in the matrix on the left. The general formula for this product is:

$$\mathbf{B}_{pq} \bullet \mathbf{c}_q = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ b_{21} & b_{22} & \dots & b_{2q} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ b_{p1} & b_{p2} & \dots & b_{pq} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_q \end{bmatrix} = \begin{bmatrix} b_{11}c_1 + b_{12}c_2 + \dots + b_{1q}c_q \\ b_{21}c_1 + b_{22}c_2 + \dots + b_{2q}c_q \\ \cdot \\ \cdot \\ \cdot \\ b_{p1}c_1 + b_{p2}c_2 + \dots + b_{pq}c_q \end{bmatrix}$$

Using summation notation, this equation may be rewritten as:

$$\mathbf{B}_{pq} \bullet \mathbf{c}_q = \begin{bmatrix} \sum_{k=1}^q b_{1k} c_k \\ \cdot \\ \cdot \\ \sum_{k=1}^q b_{pk} c_k \end{bmatrix} \quad (2.10)$$

The product of two matrices is the logical extension of the product of a vector by a matrix. Matrix  $\mathbf{C}$ , to be multiplied by  $\mathbf{B}$ , is simply considered as a set of column vectors  $\mathbf{c}_1, \mathbf{c}_2, \dots$ ; eq. 2.10 is repeated for each column. Following the same logic, the resulting column vectors are juxtaposed to form the result matrix. Matrices to be multiplied must be *conformable*, which means that the number of columns in the matrix on the left must be the same as the number of rows in the matrix on the right. For example, given

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 2 \\ 3 & 1 & 1 \\ 1 & 2 & 1 \\ -1 & 3 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & -1 \end{bmatrix}$$

$$\mathbf{C} = [\mathbf{d} \quad \mathbf{e}]$$

the product of  $\mathbf{B}$  with each of the two columns of  $\mathbf{C}$  is:

$$\mathbf{Bd} = \begin{bmatrix} 1(1) + 0(2) + 2(3) \\ 3(1) + 1(2) + 1(3) \\ 1(1) + 2(2) + 1(3) \\ -1(1) + 3(2) + 2(3) \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \\ 8 \\ 11 \end{bmatrix} \quad \text{and} \quad \mathbf{Be} = \begin{bmatrix} 1(2) + 0(1) + 2(-1) \\ 3(2) + 1(1) + 1(-1) \\ 1(2) + 2(1) + 1(-1) \\ -1(2) + 3(1) + 2(-1) \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 3 \\ -1 \end{bmatrix}$$

so that the product matrix is:

$$\mathbf{BC} = \begin{bmatrix} 7 & 0 \\ 8 & 6 \\ 8 & 3 \\ 11 & -1 \end{bmatrix}$$



Thus, the product of two conformable matrices  $\mathbf{B}$  and  $\mathbf{C}$  is a new matrix with the same number of rows as  $\mathbf{B}$  and the same number of columns as  $\mathbf{C}$ . Element  $d_{ij}$ , in row  $i$  and column  $j$  of the resulting matrix, is the scalar product of row  $i$  of  $\mathbf{B}$  with column  $j$  of  $\mathbf{C}$ .

The only way to master the mechanism of matrix products is to go through some numerical examples. As an exercise, readers could apply the above method to two cases which have not been discussed so far, i.e. the *product (bc) of a row vector c by a column vector b*, which gives a *matrix* and not a scalar, and the *product (bC) of a matrix C by a row vector b*, which results in a *row vector*. This exercise would help to better understand the rule of conformability.

As supplementary exercises, readers could calculate numerical examples of the eight following properties of matrix products, which will be used later in the book:

(1)  $\mathbf{B}_{pq} \mathbf{C}_{qr} \mathbf{D}_{rs} = \mathbf{E}_{ps}$ , of order  $(p \times s)$ .

(2) The existence of product  $\mathbf{BC}$  does not imply that product  $\mathbf{CB}$  exists, because matrices are not necessarily conformable in the reverse order; however,  $\mathbf{C}'\mathbf{C}$  and  $\mathbf{CC}'$  always exist.

(3)  $\mathbf{BC}$  is generally not equal to  $\mathbf{CB}$ , i.e. matrix products are not commutative.

(4)  $\mathbf{B}^2 = \mathbf{B} \times \mathbf{B}$  exists only if  $\mathbf{B}$  is a square matrix.

(5)  $[\mathbf{AB}]' = \mathbf{B}'\mathbf{A}'$  and, more generally,  $[\mathbf{ABCD}\dots]' = \dots\mathbf{D}'\mathbf{C}'\mathbf{B}'\mathbf{A}'$ .

(6) The products  $\mathbf{XX}'$  and  $\mathbf{X}'\mathbf{X}$  always give rise to symmetric matrices.

(7) In general, the product of two symmetric but different matrices  $\mathbf{A}$  and  $\mathbf{B}$  is not a symmetric matrix.

(8) If  $\mathbf{B}$  is an orthogonal matrix (i.e. a rectangular matrix whose column vectors are orthogonal to one another), then  $\mathbf{B}'\mathbf{B} = \mathbf{D}$ , where  $\mathbf{D}$  is a diagonal matrix. All non-diagonal terms are zero because of the property of orthogonality, while the diagonal terms are the squares of the lengths of the column vectors. That  $\mathbf{B}'\mathbf{B}$  is diagonal does not imply that  $\mathbf{BB}'$  is also diagonal.  $\mathbf{BB}' = \mathbf{B}'\mathbf{B}$  only when  $\mathbf{B}$  is square and symmetric.

Hadamard  
product

The *Hadamard* or *elementwise* product of two matrices of the same order  $(n \times p)$  is the cell-by-cell product of these two matrices. For example,

$$\text{for } \mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}, \quad \mathbf{A} * \mathbf{B} = \begin{bmatrix} 7 & 16 \\ 27 & 40 \\ 55 & 72 \end{bmatrix}$$

The Hadamard product may be noted by different operator signs, depending on the author. The sign used in this book is  $*$ , as in the R language.

The last type of product to be considered is that of a matrix or vector *by a scalar*. It is carried out according to the usual algebraic rules of multiplication and factoring, i.e. for matrix  $\mathbf{B} = [b_{jk}]$  or vector  $\mathbf{c} = [c_j]$ ,  $d\mathbf{B} = [db_{jk}]$  and  $d\mathbf{c} = [dc_j]$ . For example:

$$3 \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 9 & 12 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 5 \\ 6 \end{bmatrix} 2 = \begin{bmatrix} 10 \\ 12 \end{bmatrix}$$

The terms *premultiplication* and *postmultiplication* may be encountered in the literature. Product  $\mathbf{BC}$  corresponds to *premultiplication* of  $\mathbf{C}$  by  $\mathbf{B}$ , or to *postmultiplication* of  $\mathbf{B}$  by  $\mathbf{C}$ . Unless otherwise specified, it is always premultiplication which is implied and  $\mathbf{BC}$  simply reads:  $\mathbf{B}$  multiplies  $\mathbf{C}$ , or  $\mathbf{C}$  is multiplied by  $\mathbf{B}$ .

## 2.6 Determinant

It is often necessary to transform a matrix into a new one, in such a way that the information of the original matrix is preserved, while new properties that are essential for subsequent calculations are acquired. Such new matrices, which are linearly derived from the original matrix, will be studied in following sections under the names *inverse matrix*, *canonical form*, etc.

The new matrix must have a minimum number of characteristics in common with the matrix from which it is linearly derived. The connection between the two matrices is a matrix function  $f(\mathbf{B})$ , whose properties are the following:

(1) The determinant function must be *multilinear*, which means that it should respond linearly to any change taking place in the rows or columns of matrix  $\mathbf{B}$ .

(2) Since the order of the rows and columns of a matrix is specified, the function should be able to detect, through *alternation of signs*, any change in the positions of rows or columns. As a corollary, if two columns (or rows) are identical,  $f(\mathbf{B}) = 0$ ; indeed, if two identical columns (or rows) are interchanged,  $f(\mathbf{B})$  must change sign but it must also remain identical, which is possible only if  $f(\mathbf{B}) = 0$ .

(3) Finally, there is a scalar associated with this function; it is called its *norm* or *value* of the determinant function. For convenience, the norm is calibrated in such a way that the value associated with the unit matrix  $\mathbf{I}$  is 1, i.e.  $f(\mathbf{I}) = 1$ .

It can be shown that the determinant, as defined below, is the only function that has the above three properties, and that it only exists for square matrices. Therefore, it is not possible to calculate a determinant for a rectangular matrix. The determinant of matrix  $\mathbf{B}$  is denoted  $\det \mathbf{B}$ ,  $\det(\mathbf{B})$ , or, more often,  $|\mathbf{B}|$ :

$$|\mathbf{B}| = \begin{vmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{vmatrix}$$

The value of function  $|\mathbf{B}|$  is a scalar, i.e. a number.

What follows is the formal definition of the value of a determinant. The way to compute it in practice is explained later. The *value of a determinant* is calculated as the sum of all possible products containing one, and only one, element from each row and each column; these products receive a sign according to a well-defined rule:

$$|\mathbf{B}| = \sum \pm (b_{1j_1} b_{2j_2} \dots b_{nj_n})$$

where indices  $j_1, j_2, \dots, j_n$ , go through the  $n!$  permutations of the numbers  $1, 2, \dots, n$ . The sign depends on the number of inversions, in the permutation considered, relative to the sequence  $1, 2, \dots, n$ : if the number of inversions is even, the sign is (+) and, if the number is odd, the sign is (-).

The determinant of a matrix of order 2 is calculated as follows:

$$|\mathbf{B}| = \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} = b_{11}b_{22} - b_{12}b_{21} \quad (2.11)$$

In accordance with the formal definition above, the scalar so obtained is composed of  $2! = 2$  products, each product containing one, and only one, element from each row and each column.

The determinant of a matrix of order higher than 2 may be calculated using different methods, among which is the *expansion by minors*. When looking for a determinant of order 3, a determinant of order  $3 - 1 = 2$  may be obtained by crossing out one row ( $i$ ) and one column ( $j$ ). This lower-order determinant is the *minor* associated with  $b_{ij}$ :

*crossing out row 1 and column 2*

$$\begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \rightarrow \begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix} \quad \text{minor of } b_{12} \quad (2.12)$$

The minor being here a determinant of order 2, its value is calculated using eq. 2.11. When multiplied by  $(-1)^{i+j}$ , the minor becomes a *cofactor*. Thus, the cofactor of  $b_{12}$  is:

$$\text{cof } b_{12} = (-1)^{1+2} \begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix} = - \begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix} \quad (2.13)$$

The expansion by minors of a determinant of order  $n$  is:

$$|\mathbf{B}| = \sum_{i=1}^n b_{ij} \text{cof } b_{ij} \quad \text{for any column } j \quad (2.14)$$

$$|\mathbf{B}| = \sum_{j=1}^n b_{ij} \text{cof } b_{ij} \quad \text{for any row } i$$

The expansion may involve the elements of any row or any column, the result being always the same. Thus, going back to the determinant of the matrix on the left in eq. 2.12, expansion by the elements of the first row gives:

$$|\mathbf{B}| = b_{11} \text{cof } b_{11} + b_{12} \text{cof } b_{12} + b_{13} \text{cof } b_{13} \quad (2.15)$$

$$|\mathbf{B}| = b_{11} (-1)^{1+1} \begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix} + b_{12} (-1)^{1+2} \begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix} + b_{13} (-1)^{1+3} \begin{vmatrix} b_{21} & b_{22} \\ b_{31} & b_{32} \end{vmatrix}$$

**Numerical example.** Equation 2.15 is applied to a simple numerical example:

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix} = 1(-1)^{1+1} \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} + 2(-1)^{1+2} \begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix} + 3(-1)^{1+3} \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}$$

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix} = 1(5 \times 10 - 6 \times 8) - 2(4 \times 10 - 6 \times 7) + 3(4 \times 8 - 5 \times 7) = -3$$

The amount of calculations required to expand a determinant increases very quickly with increasing order  $n$ . This is because the minor of each cofactor must be expanded, the latter producing new cofactors whose minors are in turn expanded, and so forth until cofactors of order 2 are reached. Another, faster method is normally used to calculate determinants by computer. Before describing this method, however, some properties of determinants must be examined; in all cases, *column* may be substituted for *row*.

- (1) The determinant of a matrix is equal to that of its transpose since a determinant may be computed from either the rows or columns of the matrix:  $|\mathbf{A}'| = |\mathbf{A}|$ .
- (2) If two rows are interchanged, the sign of the determinant is reversed.
- (3) If two rows are identical, the determinant is null (corollary of the second property; see beginning of the present section).
- (4) If a scalar is a factor of *one* row, it becomes a factor of the determinant (since it appears *once* in each product).
- (5) If a row is a multiple of another row, the determinant is null (corollary of properties 4 and 3, i.e. factoring out the multiplier produces two identical rows).
- (6) If all elements of a row are 0, the determinant is null (corollary of property 4).
- (7) If a scalar  $c$  is a factor of *all* rows, it becomes a factor  $c^n$  of the determinant (corollary of property 4), i.e.  $|c\mathbf{B}| = c^n |\mathbf{B}|$ .
- (8) If a multiple of a row is added to another row, the value of the determinant remains unchanged.
- (9) The determinant of a triangular matrix (and therefore also of a diagonal matrix) is the product of its diagonal elements.
- (10) The sum of the products of the elements of a row with the corresponding cofactors of a *different* row is equal to zero.
- (11) For two square matrices of order  $n$ ,  $|\mathbf{A}| \cdot |\mathbf{B}| = |\mathbf{AB}|$ .

Pivotal  
condensation

Properties 8 and 9 can be used for rapid computer calculation of the value of a determinant; the method is called *pivotal condensation*. The matrix is first reduced to triangular form using property 8. This property allows the stepwise elimination of all terms on one side of the diagonal through combinations of multiplications by a scalar, and addition and subtraction of rows or columns. Pivotal condensation may be performed in either the upper or the lower triangular parts of a square matrix. If the lower triangular part is chosen, the upper left-hand diagonal element is used as the first pivot to modify the other rows in such a way that their left-hand terms become zero. The technique consists in calculating by how much the pivot must be multiplied to cancel out the terms in the rows below it; when this value is found, property 8 is used with this value as multiplier. When all terms under the diagonal element in the first column are zero, the procedure is repeated with the other diagonal terms as pivots, to cancel out the elements located under them in the same column. Working on the pivots from left to right insures that when values have been changed to 0, they remain so. When the whole lower triangular portion of the matrix is zero, property 9 is used to compute the determinant which is then the product of the modified diagonal elements.

**Numerical example.** The same numerical example as above illustrates the method:

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 10 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{vmatrix}$$

$a \qquad \qquad \qquad b \qquad \qquad \qquad c$

$a$ : (row 2  $- 4 \times$  row 1)       $b$ : (row 3  $- 7 \times$  row 1)       $c$ : (row 3  $- 2 \times$  row 2)

The determinant is the product of the diagonal elements:  $1 \times (-3) \times 1 = (-3)$ .

## 2.7 Rank of a matrix

A square matrix contains  $n$  vectors (rows or columns), which may be *linearly independent* or not (for the various meanings of “independence”, see Box 1.1). Two vectors are *linearly dependent* when the elements of one are proportional to the elements of the other. For example:

$$\begin{bmatrix} -4 \\ -6 \\ -8 \end{bmatrix} \text{ and } \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \text{ are linearly dependent, since } \begin{bmatrix} -4 \\ -6 \\ -8 \end{bmatrix} = -2 \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

Similarly, a vector is linearly dependent on two others, which are themselves linearly independent, when its elements are a linear combination of the elements of the other two. For example:

$$\begin{bmatrix} -1 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ -2 \\ -3 \end{bmatrix}$$

illustrate a case where a vector is linearly dependent on two others, which are themselves linearly independent, since

$$(-2) \begin{bmatrix} -1 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ -2 \\ -3 \end{bmatrix}$$

Rank of  
a square  
matrix

The *rank* of a *square matrix* is defined as the number of linearly independent row vectors (or column vectors) in the matrix. For example:

$$\begin{bmatrix} -1 & -1 & 1 \\ 3 & 0 & -2 \\ 4 & 1 & -3 \end{bmatrix} \quad \begin{array}{l} (-2 \times \text{column } 1) = \text{column } 2 + (3 \times \text{column } 3) \\ \text{or: row } 1 = \text{row } 2 - \text{row } 3 \\ \text{rank} = 2 \end{array}$$

$$\begin{bmatrix} -2 & 1 & 4 \\ -2 & 1 & 4 \\ -2 & 1 & 4 \end{bmatrix} \quad \begin{array}{l} (-2 \times \text{column } 1) = (4 \times \text{column } 2) = \text{column } 3 \\ \text{or: row } 1 = \text{row } 2 = \text{row } 3 \\ \text{rank} = 1 \end{array}$$

According to property 5 of determinants (Section 2.6), a matrix whose *rank* is lower than its *order* has a determinant equal to zero. Finding the rank of a matrix may therefore be based on the determinant of the lower-order submatrices it contains. The *rank* of a square matrix is the order of the largest square submatrix with non-zero determinant that it contains; this is also the maximum number of linearly independent vectors found among the rows or the columns.

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix} = -3 \neq 0, \text{ so that the } \text{rank} = 3$$

$$\begin{vmatrix} -1 & -1 & 1 \\ 3 & 0 & -2 \\ 4 & 1 & -3 \end{vmatrix} = 0 \quad \begin{vmatrix} -1 & -1 \\ 3 & 0 \end{vmatrix} = 3$$

*rank* = 2

The determinant can be used to diagnose the independence of the vectors forming a matrix **X**. For a *square* matrix **X** (symmetric or not), all row and column vectors are linearly independent if  $\det(\mathbf{X}) \neq 0$ .

Linear independence of the vectors in a *rectangular* matrix **X** with more rows than columns ( $n > p$ ) can be determined from the covariance matrix **S** computed from **X** (eq. 4.6): if  $\det(\mathbf{S}) \neq 0$ , all column vectors of **X** are linearly independent. This method of diagnosis of the linear independence of the column vectors requires, however, a matrix **X** with  $n > p$ ; if  $n \leq p$ ,  $\det(\mathbf{S}) = 0$ .

Rank of a  
rectangular  
matrix

**Numerical example 1.** It is possible to determine the rank of a *rectangular* matrix. Several *square* submatrices may be extracted from a *rectangular* matrix, by eliminating rows or/and columns from the matrix. The *rank* of a rectangular matrix is the highest rank of all the square submatrices that can be extracted from it. A first

example illustrates the case where the rank of a rectangular matrix is equal to the number of rows:

$$\begin{vmatrix} 2 & 0 & 1 & 0 & -1 & -2 & 3 \\ 1 & 2 & 2 & 0 & 0 & 1 & -1 \\ 0 & 1 & 2 & 3 & 1 & -1 & 0 \end{vmatrix} \rightarrow \begin{vmatrix} 2 & 0 & 1 \\ 1 & 2 & 2 \\ 0 & 1 & 2 \end{vmatrix} = 5 \quad \text{rank} = 3$$

**Numerical example 2.** In this example, the rank is lower than the number of rows:

$$\begin{vmatrix} 2 & 1 & 3 & 4 \\ -1 & 6 & -3 & 0 \\ 1 & 20 & -3 & 8 \end{vmatrix} \rightarrow \begin{vmatrix} 2 & 1 & 3 \\ -1 & 6 & -3 \\ 1 & 20 & -3 \end{vmatrix} = \begin{vmatrix} 2 & 1 & 4 \\ -1 & 6 & 0 \\ 1 & 20 & 8 \end{vmatrix} = \begin{vmatrix} 2 & 3 & 4 \\ -1 & -3 & 0 \\ 1 & -3 & 8 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 4 \\ 6 & -3 & 0 \\ 20 & -3 & 8 \end{vmatrix} = 0$$

$$\text{rank} < 3 \rightarrow \begin{vmatrix} 2 & 1 \\ -1 & 6 \end{vmatrix} = 13 \quad \text{rank} = 2$$

In this case, the three rows are clearly linearly dependent:  $(2 \times \text{row } 1) + (3 \times \text{row } 2) = \text{row } 3$ . Since it is possible to find a square matrix of order 2 that has a non-null determinant, the rank of the rectangular matrix is 2.

In practice, singular value decomposition (SVD, Section 2.11) can be used to determine the rank of a square or rectangular matrix: the rank is equal to the number of singular values larger than zero. Numerical example 2 will be analysed again in Application 1 of Section 2.11. For square symmetric matrices like covariance matrices, the number of nonzero eigenvalues can also be used to determine the rank of the matrix; see Section 2.10, Second property.

## 2.8 Matrix inversion

In algebra, division is expressed as either  $c \div b$ , or  $c/b$ , or  $c (1/b)$ , or  $c b^{-1}$ . In the last two expressions, division as such is replaced by multiplication with a reciprocal or inverse quantity. In matrix algebra, the division operation of  $\mathbf{C}$  by  $\mathbf{B}$  does not exist. The equivalent operation is multiplication of  $\mathbf{C}$  with the *inverse* or *reciprocal* of matrix  $\mathbf{B}$ . The inverse of matrix  $\mathbf{B}$  is denoted  $\mathbf{B}^{-1}$ ; the operation through which it is computed is called the *inversion* of matrix  $\mathbf{B}$ .

To serve its purpose, matrix  $\mathbf{B}^{-1}$  must be unique and the relation  $\mathbf{B}\mathbf{B}^{-1} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$  must be satisfied. It can be shown that only *square matrices* have unique inverses. It is also only for square matrices that the relation  $\mathbf{B}\mathbf{B}^{-1} = \mathbf{B}^{-1}\mathbf{B}$  is satisfied. Indeed, there are *rectangular* matrices  $\mathbf{B}$  for which several matrices  $\mathbf{C}$  can be found, satisfying for example  $\mathbf{C}\mathbf{B} = \mathbf{I}$  but not  $\mathbf{B}\mathbf{C} = \mathbf{I}$ . There are also rectangular matrices for which no



matrix  $\mathbf{C}$  can be found such that  $\mathbf{CB} = \mathbf{I}$ , whereas an infinite number of matrices  $\mathbf{C}$  may exist that satisfy  $\mathbf{BC} = \mathbf{I}$ . For example:

$$\mathbf{B} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 3 & -1 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 5 & 1 \end{bmatrix} \quad \mathbf{CB} = \mathbf{I} \quad \mathbf{BC} \neq \mathbf{I}$$

$$\mathbf{C} = \begin{bmatrix} 4 & 15 & 4 \\ 7 & 25 & 6 \end{bmatrix} \quad \mathbf{CB} = \mathbf{I} \quad \mathbf{BC} \neq \mathbf{I}$$

*Generalized inverses* can be computed for rectangular matrices by singular value decomposition (Section 2.11, Application 3). Note that several types of generalized inverses, described in textbooks of advanced linear algebra, are not unique.

Inverse of  
a square  
matrix

To calculate the inverse of a square matrix  $\mathbf{B}$ , the *adjugate* or *adjoint matrix* of  $\mathbf{B}$  is first defined. In the *matrix of cofactors* of  $\mathbf{B}$ , each element  $b_{ij}$  is replaced by its cofactor ( $\text{cof } b_{ij}$ ; see Section 2.6). The adjugate matrix of  $\mathbf{B}$  is the *transpose* of the matrix of cofactors:

$$\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix} \rightarrow \begin{bmatrix} \text{cof } b_{11} & \text{cof } b_{21} & \dots & \text{cof } b_{n1} \\ \text{cof } b_{12} & \text{cof } b_{22} & \dots & \text{cof } b_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \text{cof } b_{1n} & \text{cof } b_{2n} & \dots & \text{cof } b_{nn} \end{bmatrix} \quad (2.16)$$

matrix  $\mathbf{B}$                       adjugate matrix of  $\mathbf{B}$

In the case of second order matrices, cofactors are scalar values, e.g.  $\text{cof } b_{11} = b_{22}$ ,  $\text{cof } b_{12} = -b_{21}$ , etc.

The *inverse* of matrix  $\mathbf{B}$  is the adjugate matrix of  $\mathbf{B}$  divided by the determinant  $|\mathbf{B}|$ . The product of the matrix with its inverse gives the unit matrix:

$$\underbrace{\frac{1}{|\mathbf{B}|} \begin{bmatrix} \text{cof } b_{11} & \text{cof } b_{21} & \dots & \text{cof } b_{n1} \\ \text{cof } b_{12} & \text{cof } b_{22} & \dots & \text{cof } b_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \text{cof } b_{1n} & \text{cof } b_{2n} & \dots & \text{cof } b_{nn} \end{bmatrix}}_{\mathbf{B}^{-1}} \underbrace{\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{bmatrix}}_{\mathbf{B}} = \mathbf{I} \quad (2.17)$$

All diagonal terms resulting from the multiplication  $\mathbf{B}^{-1}\mathbf{B}$  (or  $\mathbf{B}\mathbf{B}^{-1}$ ) are of the form  $\sum b_{ij}\text{cof } b_{ij}$ , which is the expansion by minors of a determinant (not taking into account, at this stage, the division of each element of the matrix by  $|\mathbf{B}|$ ). Each diagonal element consequently has the value of the determinant  $|\mathbf{B}|$  (eq. 2.14). All other elements of matrix  $\mathbf{B}^{-1}\mathbf{B}$  are sums of the products of the elements of a row with the corresponding cofactors of a different row. According to property 10 of determinants (Section 2.6), each non-diagonal element is therefore null. It follows that:

$$\mathbf{B}^{-1}\mathbf{B} = \frac{1}{|\mathbf{B}|} \begin{bmatrix} |\mathbf{B}| & 0 & \dots & 0 \\ 0 & |\mathbf{B}| & \dots & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \dots & |\mathbf{B}| \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix} = \mathbf{I} \quad (2.18)$$

Singular  
matrix

An important point is that  $\mathbf{B}^{-1}$  exists only if  $|\mathbf{B}| \neq 0$ . A square matrix with a null determinant is called a *singular* matrix; it has no ordinary inverse (but see *singular value decomposition*, Section 2.11). Matrices that can be inverted are called *nonsingular*.

**Numerical example.** The numerical example of Sections 2.6 and 2.7 is used again to illustrate the calculations:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$$

The determinant is already known (Section 2.6); its value is  $-3$ . The matrix of cofactors is computed, and its *transpose* (adjugate matrix) is divided by the determinant to give the inverse matrix:

$$\begin{array}{ccc} \begin{bmatrix} 2 & 2 & -3 \\ 4 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix} & \begin{bmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix} & -\frac{1}{3} \begin{bmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix} \\ \text{matrix of cofactors} & \text{adjugate matrix} & \text{inverse of matrix} \end{array}$$

As for the determinant (Section 2.6), various methods exist for quickly inverting matrices using computers; they are especially useful for matrices of higher ranks. Description of these methods, which are available in computer packages, is beyond the scope of the present book. A popular method is briefly explained here; it is somewhat similar to the pivotal condensation presented above for determinants.

Gauss-  
Jordan

Inversion of matrix  $\mathbf{B}$  may be conducted using the *method of Gauss-Jordan*. To do so, matrix  $\mathbf{B}_{(n \times n)}$  is first augmented to the right with a same-size identity matrix  $\mathbf{I}$ , thus creating a  $n \times 2n$  matrix. This is illustrated for  $n = 3$ :

$$\begin{bmatrix} b_{11} & b_{12} & b_{13} & 1 & 0 & 0 \\ b_{21} & b_{22} & b_{23} & 0 & 1 & 0 \\ b_{31} & b_{32} & b_{33} & 0 & 0 & 1 \end{bmatrix}$$

If the augmented matrix is multiplied by matrix  $\mathbf{C}_{(n \times n)}$ , and if  $\mathbf{C} = \mathbf{B}^{-1}$ , then the resulting matrix ( $n \times 2n$ ) has an identity matrix in its first  $n$  columns and matrix  $\mathbf{C} = \mathbf{B}^{-1}$  in the last  $n$  columns.

$$[\mathbf{C} = \mathbf{B}^{-1}] [\mathbf{B} \quad , \quad \mathbf{I}] = [\mathbf{I} \quad , \mathbf{C} = \mathbf{B}^{-1}]$$

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & 1 & 0 & 0 \\ b_{21} & b_{22} & b_{23} & 0 & 1 & 0 \\ b_{31} & b_{32} & b_{33} & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & c_{11} & c_{12} & c_{13} \\ 0 & 1 & 0 & c_{21} & c_{22} & c_{23} \\ 0 & 0 & 1 & c_{31} & c_{32} & c_{33} \end{bmatrix}$$

This shows that, if matrix  $[\mathbf{B}, \mathbf{I}]$  is transformed into an equivalent matrix  $[\mathbf{I}, \mathbf{C}]$ , then  $\mathbf{C} = \mathbf{B}^{-1}$ .

The Gauss-Jordan transformation proceeds in two steps.

- In the first step, the diagonal terms are used, one after the other and from left to right, as pivots to make all the off-diagonal terms equal to zero. This is done in exactly the same way as for the determinant: a factor is calculated to cancel out the target term, using the pivot, and property 8 of the determinants is applied using this factor as multiplier. The difference with determinants is that the whole row of the augmented matrix is modified, not only the part belonging to matrix  $\mathbf{B}$ . If an off-diagonal zero value is encountered, then of course it is left as is, no cancellation by a multiple of the pivot being necessary or even possible. If a zero is found on the diagonal, this pivot has to be left aside for the time being (in actual programs, rows and columns are interchanged in a process called pivoting); this zero will be changed to a non-zero value during the next cycle unless the matrix is singular. Pivoting makes programming of this method a bit complex.
- Second step. When all the off-diagonal terms are zero, the diagonal terms of the former matrix  $\mathbf{B}$  are brought to 1. This is accomplished by dividing each row of the augmented matrix by the value now present in the diagonal term of the former  $\mathbf{B}$  (left) portion. If the changes introduced during the first step have made one of the diagonal elements equal to zero, then of course no division can bring it back to 1 and the matrix is singular (i.e. it cannot be inverted).

A Gauss-Jordan algorithm with pivoting is described in the book *Numerical recipes* (Press *et al.*, 2007).

**Numerical example.** To illustrate the Gauss-Jordan method, the same square matrix as above is first augmented, then transformed so that its left-hand portion becomes the identity matrix:

$$(a) \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 10 & 0 & 0 & 1 \end{bmatrix}$$

$$(b) \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & -6 & -11 & -7 & 0 & 1 \end{bmatrix}$$

New row 2  $\leftarrow$  row 2  $-$  4row 1

New row 3  $\leftarrow$  row 3  $-$  7row 1

$$(c) \begin{bmatrix} 3 & 0 & -3 & -5 & 2 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{bmatrix}$$

New row 1  $\leftarrow$  3row 1 + 2row 2

New row 3  $\leftarrow$  row 3  $-$  2row 2

$$(d) \begin{bmatrix} 3 & 0 & 0 & -2 & -4 & 3 \\ 0 & -3 & 0 & 2 & -11 & 6 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{bmatrix}$$

New row 1  $\leftarrow$  row 1 + 3row 3

New row 2  $\leftarrow$  row 2 + 6row 3

$$(e) \begin{bmatrix} 1 & 0 & 0 & -2/3 & -4/3 & 1 \\ 0 & 1 & 0 & -2/3 & 11/3 & -2 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{bmatrix}$$

New row 1  $\leftarrow$  (1/3) row 1

New row 2  $\leftarrow$   $-(1/3)$  row 2

New row 3  $\leftarrow$  row 3

$$(f) -\frac{1}{3} \begin{bmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix}$$

inverse of matrix **B**

The inverse of matrix **B** is the same as calculated above.

The inverse of a matrix has several interesting properties, including:

$$(1) \mathbf{B}^{-1}\mathbf{B} = \mathbf{B}\mathbf{B}^{-1} = \mathbf{I}.$$

$$(2) |\mathbf{B}^{-1}| = 1/|\mathbf{B}|.$$

$$(3) [\mathbf{B}^{-1}]^{-1} = \mathbf{B}.$$

$$(4) [\mathbf{B}']^{-1} = [\mathbf{B}^{-1}]'.$$

$$(5) \text{ If } \mathbf{B} \text{ and } \mathbf{C} \text{ are nonsingular square matrices, } [\mathbf{BC}]^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}.$$

$$(6) \text{ In the case of a symmetric matrix, since } \mathbf{B}' = \mathbf{B}, \text{ then } [\mathbf{B}^{-1}]' = \mathbf{B}^{-1}.$$

Orthonormal matrix

(7) An orthogonal matrix (Section 2.5) whose column vectors are normalized (scaled to length 1: Section 2.4) is called *orthonormal*. A square orthonormal matrix **B** has the property that  $\mathbf{B}' = \mathbf{B}^{-1}$ . This may be shown as follows: on the one hand,  $\mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$  by definition of the inverse of a square matrix. On the other hand, property 8 of matrix products (Section 2.5) shows that  $\mathbf{B}'\mathbf{B} = \mathbf{D}(1)$  when the column vectors in **B** are normalized (which is the case for an orthonormal matrix);  $\mathbf{D}(1)$  is a diagonal matrix of 1's, which is the identity matrix **I** (eq. 2.5). Given that  $\mathbf{B}'\mathbf{B} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$ , then

$\mathbf{B}' = \mathbf{B}^{-1}$ . Furthermore, combining the properties  $\mathbf{B}\mathbf{B}^{-1} = \mathbf{I}$  (which is true for any square matrix) and  $\mathbf{B}' = \mathbf{B}^{-1}$  shows that  $\mathbf{B}\mathbf{B}' = \mathbf{I}$ . For example, the matrix of normalized eigenvectors of a symmetric matrix, which is square and orthonormal (Section 2.9), has these properties.

(8) The inverse of a diagonal matrix is a diagonal matrix whose elements are the reciprocals of the original elements:  $[\mathbf{D}(x_i)]^{-1} = \mathbf{D}(1/x_i)$ .

Inversion is used in many types of applications, as will be seen in the remainder of this book. Classical examples of the role of inverse matrices are solving systems of linear equations and the calculation of regression coefficients.

System of  
linear  
equations

A system of linear equations can be represented in matrix form; for example:

$$\begin{aligned} b_1 + 2b_2 + 3b_3 &= 2 \\ 4b_1 + 5b_2 + 6b_3 &= 2 \\ 7b_1 + 8b_2 + 10b_3 &= 3 \end{aligned} \rightarrow \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$$

which may be written  $\mathbf{A}\mathbf{b} = \mathbf{c}$ . To find the values of the unknowns  $b_1, b_2$  and  $b_3$ , vector  $\mathbf{b}$  must be isolated to the left, which necessitates an inversion of the square matrix  $\mathbf{A}$ :

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$$

The inverse of  $\mathbf{A}$  has been calculated above. Multiplication with vector  $\mathbf{c}$  provides the solution for the three unknowns:

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = -\frac{1}{3} \begin{bmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} = -\frac{1}{3} \begin{bmatrix} 3 \\ 0 \\ -3 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad \begin{aligned} b_1 &= -1 \\ b_2 &= 0 \\ b_3 &= 1 \end{aligned}$$

Systems of linear equations are solved in that way in Subsections 13.2.2 and 13.3.3.

Simple  
linear  
regression

Linear regression analysis is reviewed in Section 10.3. *Regression coefficients* are easily calculated for several models using matrix inversion; the approach is briefly discussed here. The mathematical model for *simple linear regression* (model I, Subsection 10.3.1) is:

$$\hat{y} = b_0 + b_1x$$

The regression coefficients  $b_0$  and  $b_1$  are estimated from the observed data  $\mathbf{x}$  and  $\mathbf{y}$ . This is equivalent to resolving the following system of equations:

$$\begin{array}{l} y_1 = b_0 + b_1 x_1 \\ y_2 = b_0 + b_1 x_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n = b_0 + b_1 x_n \end{array} \rightarrow \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

Least  
squares

Matrix  $\mathbf{X}$  was augmented with a column of 1's in order to estimate the intercept of the regression equation,  $b_0$ . Coefficients  $b$  are estimated by the *method of least squares* (Subsection 10.3.1), which minimizes the sum of squares of the differences between observed values  $y$  and values  $\hat{y}$  calculated using the regression equation. In order to obtain a least-squares best fit, each member (left and right) of matrix equation  $\mathbf{y} = \mathbf{X}\mathbf{b}$  is multiplied by the transpose of matrix  $\mathbf{X}$ , i.e.  $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\mathbf{b}$ . By doing so, the rectangular matrix  $\mathbf{X}$  produces a square matrix  $\mathbf{X}'\mathbf{X}$ , which can be inverted. The values of coefficients  $b_0$  and  $b_1$  forming vector  $\mathbf{b}$  are computed directly, after inverting the square matrix  $[\mathbf{X}'\mathbf{X}]$ :

$$\mathbf{b} = [\mathbf{X}'\mathbf{X}]^{-1} [\mathbf{X}'\mathbf{y}] \quad (2.19)$$

Multiple  
linear  
regression

Using the same approach, it is easy to compute coefficients  $b_0, b_1, \dots, b_m$  of a *multiple linear regression* (Subsection 10.3.3). In this type of regression, variable  $y$  is a linear function of several ( $m$ ) variables  $x_j$ , so that one can write:

$$\hat{y} = b_0 + b_1 x_1 + \dots + b_m x_m$$

Vectors  $\mathbf{y}$  and  $\mathbf{b}$  and matrix  $\mathbf{X}$  are defined as follows:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & \dots & x_{nm} \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{bmatrix}$$

Again, matrix  $\mathbf{X}$  was augmented with a column of 1's in order to estimate the intercept of the equation,  $b_0$ . The least-squares solution is found by computing eq. 2.19. Readers can consult Section 10.3 for computational and variable selection methods to be used in multiple linear regression when the variables  $x_j$  are strongly intercorrelated, as is often the case in ecology.

Polynomial  
regression

In *polynomial regression* (Subsection 10.3.4), several regression parameters  $b$ , corresponding to powers of a single variable  $x$ , are fitted to the observed data. The general regression model is:

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + \dots + b_k x^k$$

The vector of parameters,  $\mathbf{b}$ , is computed in the same way. Vectors  $\mathbf{y}$  and  $\mathbf{b}$ , and matrix  $\mathbf{X}$ , are defined as follows:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^k \\ 1 & x_2 & x_2^2 & \dots & x_2^k \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 1 & x_n & x_n^2 & \dots & x_n^k \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_k \end{bmatrix}$$

The least-squares solution is computed using eq. 2.19. Readers should consult Subsection 10.3.4 where practical considerations concerning the calculation of polynomial regression with ecological data are discussed.

## 2.9 Eigenvalues and eigenvectors

There are other problems, in addition to those examined above, where the determinant and the inverse of a matrix are used to provide simple and elegant solutions. An important one in data analysis is the derivation of an orthogonal form (i.e. a matrix whose vectors are at right angles; Sections 2.5 and 2.8) for a non-orthogonal symmetric matrix. This will provide the algebraic basis for most of the methods studied in Chapters 9 and 11. In ecology, data sets generally include a large number of variables, which are associated to one another (e.g. linearly correlated; Section 4.2). The basic idea underlying several methods of data analysis is to reduce this large number of intercorrelated variables to a smaller number of composite, but linearly independent (Box 1.1) variables, each explaining a different fraction of the observed variation. One of the main goals of numerical data analysis is indeed to generate a small number of variables, each explaining a large portion of the variation, and to ascertain that these new variables explain different aspects of the phenomena under study. The present section only deals with the mathematics of the computation of *eigenvalues* and *eigenvectors*. Applications to the analysis of multidimensional ecological data are discussed in Chapters 4, 9 and 11.

Mathematically, the problem may be formulated as follows. Given a square matrix  $\mathbf{A}$ , one wishes to find a diagonal matrix that is equivalent to  $\mathbf{A}$ . To fix ideas,  $\mathbf{A}$  is a covariance matrix  $\mathbf{S}$  in principal component analysis. Other types of square, symmetric

association matrices (Section 2.2) are used in numerical ecology, hence the use of the symbol **A**:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

In **A**, the terms above and below the diagonal characterize the degree of association of either the objects, or the ecological variables, with one another (Fig. 2.1). In the new matrix **Λ** (capital lambda) being sought, all elements outside the diagonal are null:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & 0 & \dots & 0 \\ 0 & \lambda_{22} & \dots & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \dots & \lambda_{nn} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \quad (2.20)$$

This new matrix is called the *matrix of eigenvalues*<sup>\*</sup>. It has the same trace and the same determinant as **A**. The new variables (*eigenvectors*; see below) whose association is described by this matrix **Λ** are thus linearly independent of one another. The use of the Greek letter λ (lower-case lambda) to represent eigenvalues stems from the fact that eigenvalues are actually *Lagrangian multipliers* λ, as will be shown in Section 4.4. Matrix **Λ** is known as the *canonical form* of matrix **A**; for the exact meaning of *canonical* in mathematics, see Section 11.0.

Canonical  
form

## 1 — Computation

The eigenvalues and eigenvectors of matrix **A** are found from equation

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i \quad (2.21)$$

which allows one to compute the different eigenvalues λ<sub>i</sub> and their associated eigenvectors **u**<sub>i</sub>. First, the validity of eq. 2.21 must be demonstrated.

\* In the literature, the following expressions are synonymous:

eigenvalue	eigenvector
characteristic root	characteristic vector
latent root	latent vector

*Eigen* is the German word for *characteristic*.



To do so, one uses any pair  $h$  and  $i$  of eigenvalues and eigenvectors computed from matrix  $\mathbf{A}$ . Equation 2.21 becomes

$$\mathbf{A}\mathbf{u}_h = \lambda_h \mathbf{u}_h \quad \text{and} \quad \mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i, \quad \text{respectively.}$$

Multiplying these equations by row vectors  $\mathbf{u}_i'$  and  $\mathbf{u}_h'$ , respectively, gives:

$$\mathbf{u}_i' \mathbf{A} \mathbf{u}_h = \lambda_h \mathbf{u}_i' \mathbf{u}_h \quad \text{and} \quad \mathbf{u}_h' \mathbf{A} \mathbf{u}_i = \lambda_i \mathbf{u}_h' \mathbf{u}_i$$

It can be shown that, in the case of a symmetric matrix, the left-hand members of these two equations are equal:  $\mathbf{u}_i' \mathbf{A} \mathbf{u}_h = \mathbf{u}_h' \mathbf{A} \mathbf{u}_i$ ; this would not be true for an asymmetric matrix, however. Using a  $(2 \times 2)$  matrix  $\mathbf{A}$  like the one of Numerical example 1 below, readers can easily check that the equality holds only when  $a_{12} = a_{21}$ , i.e. when  $\mathbf{A}$  is symmetric. So, in the case of a symmetric matrix, the right-hand members are also equal:

$$\lambda_h \mathbf{u}_i' \mathbf{u}_h = \lambda_i \mathbf{u}_h' \mathbf{u}_i$$

Since we are talking about two distinct values for  $\lambda_h$  and  $\lambda_i$ , the only possibility for the above equality to be true is that the product of vectors  $\mathbf{u}_h$  and  $\mathbf{u}_i$  be 0 (i.e.  $\mathbf{u}_i' \mathbf{u}_h = \mathbf{u}_h' \mathbf{u}_i = 0$ ), which is the condition of orthogonality for two vectors (Section 2.5). It is therefore concluded that eq. 2.21

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

can be used to compute vectors  $\mathbf{u}_i$  that are orthogonal when matrix  $\mathbf{A}$  is symmetric. In the case of a non-symmetric matrix, eigenvectors can also be calculated, but they are not orthogonal.

If the scalars  $\lambda_i$  and their associated vectors  $\mathbf{u}_i$  exist, then eq. 2.21 can be transformed as follows:

$$\mathbf{A}\mathbf{u}_i - \lambda_i \mathbf{u}_i = \mathbf{0} \quad (\text{difference between two vectors})$$

and vector  $\mathbf{u}_i$  can be factorized:

$$(\mathbf{A} - \lambda_i \mathbf{I}) \mathbf{u}_i = \mathbf{0} \tag{2.22}$$

Because of the nature of the elements in eq. 2.22, it is necessary to introduce a unit matrix  $\mathbf{I}$  inside the parentheses, where one now finds a difference between two square matrices. According to eq. 2.22, multiplication of the square matrix  $(\mathbf{A} - \lambda_i \mathbf{I})$  with the column vector  $\mathbf{u}_i$  must result in a null column vector ( $\mathbf{0}$ ).

Besides the trivial solution where  $\mathbf{u}_i$  is a null vector, eq. 2.22 has the following solution:

$$|\mathbf{A} - \lambda_i \mathbf{I}| = 0 \tag{2.23}$$

Character-  
istic equation

That is, the determinant of the difference between matrices  $\mathbf{A}$  and  $\lambda_i \mathbf{I}$  must be equal to 0 for each  $\lambda_i$ . Resolving eq. 2.23 provides the eigenvalues  $\lambda_i$  associated with matrix  $\mathbf{A}$ . Equation 2.23 is known as the *characteristic* or *determinantal equation*.

Demonstration of eq. 2.23 goes as follows:

1) One solution to  $(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{u}_i = \mathbf{0}$  is that  $\mathbf{u}_i$  is the null vector:  $\mathbf{u} = [0]$ . This solution is trivial, since it corresponds to the centroid of the scatter of data points. A non-trivial solution must thus involve  $(\mathbf{A} - \lambda_i \mathbf{I})$ .

2) Solution  $(\mathbf{A} - \lambda_i \mathbf{I}) = [0]$  is not acceptable either, since it implies that  $\mathbf{A} = \lambda_i \mathbf{I}$  and thus that  $\mathbf{A}$  be a scalar matrix, which is generally not true.

3) The solution thus requires that  $\lambda_i$  and  $\mathbf{u}_i$  be such that the *scalar product*  $(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{u}_i$  is a null vector. In other words, vector  $\mathbf{u}_i$  must be orthogonal to the space corresponding to  $\mathbf{A}$  after  $\lambda_i \mathbf{I}$  has been subtracted from it; orthogonality of two vectors or matrices is obtained when their scalar product is zero (Section 2.5). The solution  $|\mathbf{A} - \lambda_i \mathbf{I}| = 0$  (eq. 2.23) means that, for each value  $\lambda_i$ , the rank of  $(\mathbf{A} - \lambda_i \mathbf{I})$  is lower than its order, which makes the determinant equal to zero (Section 2.7). Each  $\lambda_i$  is the variance corresponding to one dimension of matrix  $\mathbf{A}$  (Section 4.4). It is then easy to calculate the eigenvector  $\mathbf{u}_i$  that is orthogonal to the space  $(\mathbf{A} - \lambda_i \mathbf{I})$  of lower dimension than  $\mathbf{A}$ . That eigenvector is the solution to eq. 2.22, which specifies orthogonality of  $\mathbf{u}_i$  with respect to  $(\mathbf{A} - \lambda_i \mathbf{I})$ .

For a matrix  $\mathbf{A}$  of order  $n$ , the characteristic equation is a polynomial of degree  $n$ , whose solutions are the eigenvalues  $\lambda_i$ . When these values are found, it is easy to use eq. 2.22 to calculate the eigenvector  $\mathbf{u}_i$  corresponding to each eigenvalue  $\lambda_i$ . There are therefore as many eigenvectors as there are eigenvalues.

There are methods that enable the quick and efficient calculation of eigenvalues and eigenvectors by computer. Three of these are described in Subsection 9.1.9.

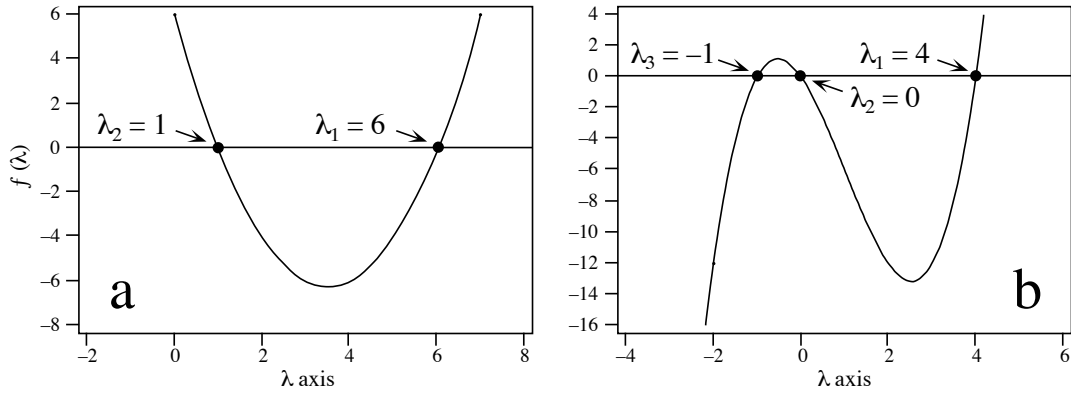
Ecologists, who are more concerned with shedding light on natural phenomena than on mathematical entities, may find unduly technical this discussion of the computation of eigenvalues and eigenvectors. The same subject will be considered again in Section 4.4 in the context of the multidimensional normal distribution. Mastering the bases of this algebraic operation is essential to understand the methods based on *eigenanalysis* (Chapters 9 and 11), which are of prime importance to the analysis of ecological data.

## 2 — Numerical examples

This subsection contains two examples of eigen-decomposition.

**Numerical example 1.** The characteristic equation of the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}$$



**Figure 2.2** (a) The eigenvalues of Numerical example 1 are the values along the  $\lambda$  axis where the function  $\lambda^2 - 7\lambda + 6$  is zero. (b) Similarly for Numerical example 2, the eigenvalues are the values along the  $\lambda$  axis where the function  $\lambda^3 - 3\lambda^2 - 4\lambda$  is zero.

is (eq. 2.23)

$$\left| \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

therefore

$$\left| \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0$$

and thus

$$\begin{vmatrix} 2 - \lambda & 2 \\ 2 & 5 - \lambda \end{vmatrix} = 0$$

The *characteristic polynomial* is found by expanding the determinant (Section 2.6):

$$(2 - \lambda)(5 - \lambda) - 4 = 0$$

which gives

$$\lambda^2 - 7\lambda + 6 = 0$$

from which it is easy to calculate the two values of  $\lambda$  that satisfy the equation (Fig. 2.2a). The two eigenvalues of **A** are:

$$\lambda_1 = 6 \quad \text{and} \quad \lambda_2 = 1$$

The sum of the eigenvalues is equal to the trace (i.e. the sum of the diagonal elements) of **A**.

The ordering of eigenvalues is arbitrary. It would have been equally correct to write that  $\lambda_1 = 1$  and  $\lambda_2 = 6$ , but the convention is to sort the eigenvalues in decreasing order.

Equation 2.22 is used to calculate the eigenvectors  $\mathbf{u}_1$  and  $\mathbf{u}_2$  corresponding to eigenvalues  $\lambda_1$  and  $\lambda_2$ :

for  $\lambda_1 = 6$

$$\left( \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix} - 6 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} u_{11} \\ u_{21} \end{bmatrix} = \mathbf{0}$$

$$\begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{21} \end{bmatrix} = \mathbf{0}$$

for  $\lambda_2 = 1$

$$\left( \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix} - 1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} u_{12} \\ u_{22} \end{bmatrix} = \mathbf{0}$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} u_{12} \\ u_{22} \end{bmatrix} = \mathbf{0}$$

which is equivalent to the following pairs of linear equations:

$$-4u_{11} + 2u_{21} = 0$$

$$2u_{11} - 1u_{21} = 0$$

$$1u_{12} + 2u_{22} = 0$$

$$2u_{12} + 4u_{22} = 0$$

These sets of linear equations are always indeterminate. The solution is given by any point (vector) in the direction of the eigenvector being sought. To remove the indetermination, an arbitrary value is assigned to one of the elements  $u$ , which specifies a particular vector. For example, value  $u = 1$  may be arbitrarily assigned to the first element  $u$  in each set:

given that  $u_{11} = 1$

$u_{12} = 1$

it follows that  $-4u_{11} + 2u_{21} = 0$

$1u_{12} + 2u_{22} = 0$

become  $-4 + 2u_{21} = 0$

$1 + 2u_{22} = 0$

so that  $u_{21} = 2$

$u_{22} = -1/2$

Eigenvectors  $\mathbf{u}_1$  and  $\mathbf{u}_2$  are therefore:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 \\ -1/2 \end{bmatrix}$$

Values other than 1 could have been arbitrarily assigned to  $u_{11}$  and  $u_{12}$  (or, for that matter, to any other term in each vector). For example, the following vectors also satisfy the two pairs of linear equations, since these eigenvectors differ only by a scalar multiplier:

$$\begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

or

$$\begin{bmatrix} -3 \\ -6 \end{bmatrix}$$

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

or

$$\begin{bmatrix} -4 \\ 2 \end{bmatrix}$$

This is the reason why eigenvectors are generally standardized. One method is to assign value 1 to the *largest element* of each vector, and adjust the other elements accordingly. Another standardization method, used for instance in principal component and principal coordinate analyses (Sections 9.1 and 9.3), is to make the length of each eigenvector  $\mathbf{u}_i$  equal to the square root of its eigenvalue (eigenvector *scaled to*  $\sqrt{\lambda_i}$ ).

The most common and practical method is to normalize eigenvectors, i.e. to make their lengths equal to 1. Thus, a *normalized* eigenvector is in fact *scaled to 1*, i.e.  $\mathbf{u}'\mathbf{u} = 1$ . As explained in Section 2.4, normalization is achieved by dividing each element of a vector by the length of this vector, i.e. the square root of the sum of squares of all elements in the vector. Like most other computer packages, the R function **eigen()** outputs normalized eigenvectors.

In the numerical example, the two eigenvectors

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

are normalized to

$$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}$$

Since the eigenvectors are both orthogonal and normalized, they are *orthonormal* (property 7 in Section 2.8).

Had the eigenvectors been multiplied by a negative scalar, their normalized forms would now be the following:

$$\begin{bmatrix} -1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

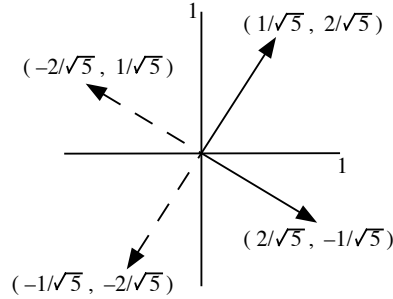
These forms are strictly equivalent to those above.

Since matrix  $\mathbf{A}$  is symmetric, its eigenvectors must be orthogonal. This is easily verified as their product is equal to zero, which is the condition for two vectors to be orthogonal (Section 2.5):

$$\mathbf{u}'_1 \mathbf{u}_2 = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix} = 2/5 - 2/5 = 0$$

The normalized eigenvectors can be plotted in the original system of coordinates, i.e. the Cartesian plane whose axes are the two original descriptors; the association between these

descriptors is given by matrix **A**. This plot (full arrows) shows that the angle between the eigenvectors is indeed  $90^\circ$  ( $\cos 90^\circ = 0$ ) and that their lengths are 1:



The dashed arrows illustrate the same eigenvectors with inverted signs. The eigenvectors with dashed arrows are equivalent to those with full arrows.

Resolving the system of linear equations used to compute eigenvectors is greatly facilitated by matrix inversion. Defining matrix  $\mathbf{C}_{nn} = (\mathbf{A} - \lambda_n \mathbf{I})$  allows eq. 2.22 to be written in a simplified form:

$$\mathbf{C}_{nn} \mathbf{u}_n = \mathbf{0}_n \quad (2.24)$$

Indices  $n$  designate here the dimensions of matrix **C** and vector **u**. Matrix  $\mathbf{C}_{nn}$  contains all the coefficients by which a given eigenvector  $\mathbf{u}_n$  is multiplied. The system of equations is indeterminate, which prevents the inversion of **C** and calculation of **u**. To remove the indetermination, it is sufficient to determine any one element of vector **u**. For example, one may arbitrarily decide that  $u_1 = \alpha$  ( $\alpha \neq 0$ ). Then,

$$\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} \begin{bmatrix} \alpha \\ u_2 \\ \cdot \\ \cdot \\ \cdot \\ u_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

can be written

$$\begin{bmatrix} c_{11}\alpha + c_{12}u_2 + \dots + c_{1n}u_n \\ c_{21}\alpha + c_{22}u_2 + \dots + c_{2n}u_n \\ \cdot \\ \cdot \\ \cdot \\ c_{n1}\alpha + c_{n2}u_2 + \dots + c_{nn}u_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

so that

$$\begin{bmatrix} c_{12}u_2 + \dots + c_{1n}u_n \\ c_{22}u_2 + \dots + c_{2n}u_n \\ \cdot \\ \cdot \\ \cdot \\ c_{n2}u_2 + \dots + c_{nn}u_n \end{bmatrix} = -\alpha \begin{bmatrix} c_{11} \\ c_{21} \\ \cdot \\ \cdot \\ \cdot \\ c_{n1} \end{bmatrix}$$

After setting  $u_1 = \alpha$ , the first column of matrix  $\mathbf{C}$  is transferred to the right. The last  $n - 1$  rows of  $\mathbf{C}$  are then sufficient to define a completely determined system. The first row is removed from  $\mathbf{C}$  in order to obtain a square matrix of order  $n - 1$ , which can be inverted. The determined system thus obtained is:

$$\begin{bmatrix} c_{22}u_2 + \dots + c_{2n}u_n \\ \cdot \\ \cdot \\ \cdot \\ c_{n2}u_2 + \dots + c_{nn}u_n \end{bmatrix} = -\alpha \begin{bmatrix} c_{21} \\ \cdot \\ \cdot \\ \cdot \\ c_{n1} \end{bmatrix}$$

$$\text{which can be written} \quad \mathbf{C}_{(n-1)(n-1)} \mathbf{u}_{(n-1)} = -\alpha \mathbf{c}_{(n-1)} \quad (2.25)$$

This system can be resolved by inversion of  $\mathbf{C}$ , as in Section 2.8:

$$\mathbf{u}_{(n-1)} = -\alpha \mathbf{C}_{(n-1)(n-1)}^{-1} \mathbf{c}_{(n-1)} \quad (2.26)$$

This method of computing the eigenvectors may not work, however, in the case of multiple eigenvalues (see Third property in Section 2.10, below). The following example provides an illustration of the computation through matrix inversion.

**Numerical example 2.** For the asymmetric matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & -1 \\ 0 & 1 & 2 \\ 1 & 4 & 1 \end{bmatrix},$$

the characteristic polynomial, computed from eq. 2.23, is  $\lambda^3 - 3\lambda^2 - 4\lambda = 0$ , from which the three eigenvalues 4, 0 and  $-1$  can be calculated (Fig. 2.2b). The sum of the eigenvalues has to be equal to the trace of  $\mathbf{A}$ , which is 3.

The eigenvectors are computed by inserting each eigenvalue, in turn, into eq. 2.22. For  $\lambda_1 = 4$ :

$$\begin{bmatrix} (1-4) & 3 & -1 \\ 0 & (1-4) & 2 \\ 1 & 4 & (1-4) \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{21} \\ u_{31} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The above system is determined by setting  $u_{11} = 1$ . Using eq. 2.25 gives:

$$\begin{bmatrix} (1-4) & 2 \\ 4 & (1-4) \end{bmatrix} \begin{bmatrix} u_{21} \\ u_{31} \end{bmatrix} = -1 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

from which it follows (eq. 2.26) that

$$\begin{bmatrix} u_{21} \\ u_{31} \end{bmatrix} = \begin{bmatrix} (1-4) & 2 \\ 4 & (1-4) \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

The inverse of matrix  $\begin{bmatrix} -3 & 2 \\ 4 & -3 \end{bmatrix}$  is  $\begin{bmatrix} -3 & -2 \\ -4 & -3 \end{bmatrix}$  so that

$$\begin{bmatrix} u_{21} \\ u_{31} \end{bmatrix} = \begin{bmatrix} -3 & -2 \\ -4 & -3 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

The two other eigenvectors are computed in the same fashion, from eigenvalues  $\lambda_2 = 0$  and  $\lambda_3 = -1$ . The resulting matrix of eigenvectors (columns) is:

$$\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3] = \begin{bmatrix} 1 & 1 & 1 \\ 2 & -2/7 & -1/2 \\ 3 & 1/7 & 1/2 \end{bmatrix} \quad \text{or else} \quad \begin{bmatrix} 1 & 7 & 2 \\ 2 & -2 & -1 \\ 3 & 1 & 1 \end{bmatrix}$$



which is normalized to

$$\mathbf{U} = \begin{bmatrix} 0.27 & 0.95 & 0.82 \\ 0.53 & -0.27 & -0.41 \\ 0.80 & 0.14 & 0.41 \end{bmatrix}$$

Readers can easily check that these eigenvectors, which were extracted from a non-symmetric matrix, are indeed not orthogonal; none of the scalar products between pairs of columns is equal to zero. The eigenanalysis of non-symmetric (or *asymmetric*) matrices will be encountered in linear discriminant analysis and canonical correlation analysis, Sections 11.3 and 11.4.

## 2.10 Some properties of eigenvalues and eigenvectors

**First property.** — A simple rearrangement of eq. 2.21 shows that matrix  $\mathbf{U}$  of the eigenvectors is a transform matrix, allowing one to go from system  $\mathbf{A}$  to system  $\mathbf{\Lambda}$ . Indeed, the equation can be rewritten so as to include all eigenvalues and eigenvectors:

$$\mathbf{AU} = \mathbf{U}\mathbf{\Lambda} \quad (2.27)$$

**Numerical example.** Equation 2.27 can be verified using Numerical example 2 of Section 2.9:

$$\begin{bmatrix} 1 & 3 & -1 \\ 0 & 1 & 2 \\ 1 & 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 7 & 2 \\ 2 & -2 & -1 \\ 3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 7 & 2 \\ 2 & -2 & -1 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

The left and right-hand sides of the equation are identical:

$$\begin{bmatrix} 4 & 0 & -2 \\ 8 & 0 & 1 \\ 12 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & -2 \\ 8 & 0 & 1 \\ 12 & 0 & -1 \end{bmatrix}$$

On the left-hand side of the equation, matrix  $\mathbf{A}$  is postmultiplied by matrix  $\mathbf{U}$  of the eigenvectors whereas, on the right-hand side, the matrix of eigenvalues  $\mathbf{\Lambda}$  is premultiplied by  $\mathbf{U}$ . It follows that  $\mathbf{U}$  achieves a two-way transformation (rows, columns), from the reference system  $\mathbf{A}$  to the system  $\mathbf{\Lambda}$ . This transformation can go both ways, as shown by the following equations which are both derived from eq. 2.27:

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1} \quad \text{and} \quad \mathbf{\Lambda} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U} \quad (2.28)$$

A simple formula may be derived from  $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$ , which can be used to raise matrix  $\mathbf{A}$  to any power  $x$ :

$$\mathbf{A}^x = (\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1})\mathbf{U}\mathbf{\Lambda} \dots \mathbf{U}^{-1}(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1})$$

$$\mathbf{A}^x = \mathbf{U}\mathbf{\Lambda}(\mathbf{U}^{-1}\mathbf{U})\mathbf{\Lambda} \dots (\mathbf{U}^{-1}\mathbf{U})\mathbf{\Lambda}\mathbf{U}^{-1}$$

$$\mathbf{A}^x = \mathbf{U}\mathbf{A}^x\mathbf{U}^{-1}, \text{ because } \mathbf{U}^{-1}\mathbf{U} = \mathbf{I}$$

Raising a matrix to some high power is greatly facilitated by the fact that  $\mathbf{A}^x$  is the matrix of eigenvalues, which is diagonal. Indeed, a diagonal matrix can be raised to any power  $x$  by raising each of its diagonal elements to power  $x$ . It follows that the last equation can be rewritten as:

$$\mathbf{A}^x = \mathbf{U}[\lambda_i^x]\mathbf{U}^{-1} \quad (2.29)$$

This may be verified using the above example. Note 1: this calculation cannot be done if there are negative eigenvalues, as in non-symmetric matrices, and the exponent is not an integer. The reason is that a fractional exponent of a negative number is undefined. Note 2: if  $\mathbf{U}$  is orthonormal,  $\mathbf{U}^{-1} = \mathbf{U}'$ , so that  $\mathbf{A}^x = \mathbf{U}[\lambda_i^x]\mathbf{U}^{-1} = \mathbf{U}[\lambda_i^x]\mathbf{U}'$  (property of the inverse of an orthonormal matrix, Section 2.8). This equality is true only if  $\mathbf{U}$  has been normalized.

**Second property.** — It was shown in Section 2.7 that, when the rank ( $r$ ) of matrix  $\mathbf{A}_{nn}$  is smaller than its order ( $r < n$ ), the determinant  $|\mathbf{A}|$  is 0. It was also shown that, when it is necessary to know the rank of a matrix, as for instance in dimensional analysis (Section 3.3),  $|\mathbf{A}| = 0$  indicates that one must check the rank of  $\mathbf{A}$ . Such a test naturally follows from the calculation of eigenvalues. Indeed, for a square symmetric matrix  $\mathbf{A}$ , the determinant is equal to the product of its eigenvalues:

$$|\mathbf{A}| = \prod_{i=1}^n \lambda_i \quad (2.30)$$

so that  $|\mathbf{A}| = 0$  if one or several of the eigenvalues is 0. When the rank of a matrix is smaller than its order ( $r < n$ ), this matrix has  $(n - r)$  null eigenvalues. Thus, *eigenvalues can be used to determine the rank of a square symmetric matrix*: the rank is equal to *the number of nonzero eigenvalues*. In the case of a covariance or cross-product matrix among variables, the number of nonzero eigenvalues (i.e. the rank of  $\mathbf{A}$ ) is the number of linearly independent dimensions required to account for all the variance (Chapter 9).

**Third property.** — It was implicitly assumed, up to this point, that the eigenvalues were *all different* from one another. It may happen, however, that some (say,  $m$ ) eigenvalues are equal. These are known as *multiple eigenvalues*. In such a case, the question is whether or not matrix  $\mathbf{A}_{nn}$  has  $n$  *distinct* eigenvectors. In other words, are there  $m$  *linearly independent* eigenvectors which correspond to the *same* eigenvalue? In principal component analysis (Section 9.1), a solution corresponding to that situation is called circular.

Values  $\lambda_i$  are chosen in such a way that the determinant  $|\mathbf{A} - \lambda_i\mathbf{I}|$  is null (eq. 2.23):

$$|\mathbf{A} - \lambda_i\mathbf{I}| = 0$$

Multiple  
eigenvalues

which means that the rank of  $(\mathbf{A} - \lambda_i \mathbf{I})$  is smaller than  $n$ . In the case of multiple eigenvalues, if there are  $m$  *distinct* eigenvectors corresponding to the  $m$  identical eigenvalues  $\lambda_i$ , the determinant of  $(\mathbf{A} - \lambda_i \mathbf{I})$  must be null for each of these eigenvalues, but in a different way each time. When  $m = 1$ , the condition for  $|\mathbf{A} - \lambda_i \mathbf{I}| = 0$  is for its rank to be  $r = n - 1$ . Similarly, in a case of *multiplicity*, the condition for  $|\mathbf{A} - \lambda_i \mathbf{I}|$  to be null  $m$  times, but distinctly, is for its rank to be  $r = n - m$ . Consequently, for  $n$  distinct eigenvectors to exist, the rank of  $(\mathbf{A} - \lambda_i \mathbf{I})$  must be  $r = n - m$ , and this for any eigenvalue  $\lambda_i$  of multiplicity  $m$ .

**Numerical example.** Here is an example of a full-rank asymmetric matrix  $\mathbf{A}$  that has two equal eigenvalues corresponding to distinct eigenvectors. The full-rank condition is shown by the fact that  $\det(\mathbf{A}) = -1$ , which differs from 0. The eigenvalues are  $\lambda_1 = \lambda_2 = 1$  and  $\lambda_3 = -1$ :

$$\mathbf{A} = \begin{bmatrix} -1 & -2 & -2 \\ 1 & 2 & 1 \\ -1 & -1 & 0 \end{bmatrix} \quad \text{so that, for } \lambda_1 = \lambda_2 = 1, \quad (\mathbf{A} - 1\mathbf{I}) = \begin{bmatrix} -2 & -2 & -2 \\ 1 & 1 & 1 \\ -1 & -1 & -1 \end{bmatrix}$$

The multiplicity, or number of multiple eigenvalues, is  $m = 2$ . The rank of  $(\mathbf{A} - \lambda_i \mathbf{I})$  is  $r = 1$  because all three columns of this matrix are identical. Thus, for  $\lambda_1 = \lambda_2 = 1$ ,  $n - m = 3 - 2 = 1$ , which shows that  $r = n - m$  in this example. It follows that there exist two distinct eigenvectors  $\mathbf{u}_1$  and  $\mathbf{u}_2$ . They can indeed be calculated:

$$\text{for } \lambda_1 = 1, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \text{for } \lambda_2 = 1, \mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \text{whereas for } \lambda_3 = -1, \mathbf{u}_3 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

Eigenvectors  $\mathbf{u}_1$  and  $\mathbf{u}_2$  both correspond to the multiple eigenvalue  $\lambda = 1$ . Any linear combination of such eigenvectors is also an eigenvector of  $\mathbf{A}$  corresponding to  $\lambda$ . For example:

$$\mathbf{u}_1 - \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \quad \mathbf{u}_1 + 2\mathbf{u}_2 = \begin{bmatrix} 3 \\ -2 \\ -1 \end{bmatrix}$$

It can easily be verified that the above two eigenvectors, or any other linear combination of  $\mathbf{u}_1$  and  $\mathbf{u}_2$ , are eigenvectors of  $\mathbf{A}$  corresponding to  $\lambda = 1$ . Of course, the new eigenvectors are not linearly independent of  $\mathbf{u}_1$  and  $\mathbf{u}_2$ , so that there are still only two distinct eigenvectors corresponding to the multiple eigenvalue  $\lambda = 1$ .

**Numerical example.** Here is an example of a full-rank asymmetric matrix  $\mathbf{A}$  that has two indistinguishable eigenvectors. The full-rank condition is shown by the fact that  $\det(\mathbf{A}) = 3$ , which differs from 0. The eigenvalues are  $\lambda_1 = 3$  and  $\lambda_2 = \lambda_3 = 1$ :

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 3 & -2 \\ 4 & 1 & 0 \end{bmatrix} \quad \text{so that, for } \lambda_2 = \lambda_3 = 1, \quad (\mathbf{A} - 1\mathbf{I}) = \begin{bmatrix} 1 & -1 & 1 \\ 3 & 2 & -2 \\ 4 & 1 & -1 \end{bmatrix}$$

**Table 2.2** Types of *symmetric* matrices and corresponding characteristics of their eigenvalues.

<i>Symmetric</i> matrix	Eigenvalues
All elements of matrix <b>A</b> are <i>real</i> (i.e. non-imaginary)	All eigenvalues are <i>real</i> (i.e. non-imaginary)
Matrix <b>A</b> is <i>positive definite</i>	All eigenvalues are <i>positive</i>
Matrix <b>A<sub>nn</sub></b> is <i>positive semidefinite</i> and of rank <i>r</i>	There are <i>r</i> positive and ( <i>n</i> − <i>r</i> ) null eigenvalues
Matrix <b>A<sub>nn</sub></b> is <i>negative semidefinite</i> and of rank <i>r</i>	There are <i>r</i> negative and ( <i>n</i> − <i>r</i> ) null eigenvalues
Matrix <b>A<sub>nn</sub></b> is <i>indefinite</i> and of rank <i>r</i>	There are <i>r</i> non-null (positive and negative) and ( <i>n</i> − <i>r</i> ) null eigenvalues
Matrix <b>A</b> is <i>diagonal</i>	The <i>diagonal elements</i> are the eigenvalues

The multiplicity, or number of multiple eigenvalues, is  $m = 2$ . The rank of  $(\mathbf{A} - \lambda_i \mathbf{I})$  is  $r = 2$  because any two of the three rows (or columns) of this matrix are independent of each other. Thus, for  $\lambda_2 = \lambda_3 = 1$ ,  $n - m = 3 - 2 = 1$ , which shows that  $r \neq n - m$  in this example. The conclusion is that there do not exist two independent eigenvectors associated with the eigenvalue of multiplicity  $m = 2$ . The eigenvectors are the following:

$$\text{for } \lambda_1 = 3, \quad \mathbf{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \quad \text{whereas for } \lambda_2 = \lambda_3 = 1, \quad \mathbf{u}_1 = \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

In the case of a square *symmetric* matrix, it is always possible to calculate  $m$  *orthogonal* eigenvectors corresponding to multiple eigenvalues, when present. This is not necessarily true for *non-symmetric* matrices, where the number of eigenvectors may be smaller than  $m$ . Therefore, whatever their multiplicity, eigenvalues of most matrices of interest to ecologists, including association matrices (Section 2.2), have distinct eigenvectors associated with them. In any case, it is unlikely that eigenvalues of matrices computed from field data be exactly equal (i.e. multiple).

**Fourth property.** — A property of square *symmetric* matrices may be used to predict the nature of their eigenvalues (Table 2.2). A symmetric matrix **A** may be combined with any vector  $\mathbf{t} \neq \mathbf{0}$ , in a matrix expression of the form  $\mathbf{t}'\mathbf{A}\mathbf{t}$  which is

Quadratic form known as a *quadratic form*. This expression results in a scalar whose value leads to the following definitions:

- if  $\mathbf{t}'\mathbf{A}\mathbf{t}$  is always positive, matrix  $\mathbf{A}$  is *positive definite*;
- if  $\mathbf{t}'\mathbf{A}\mathbf{t}$  can be either positive or null, matrix  $\mathbf{A}$  is *positive semidefinite*;
- if  $\mathbf{t}'\mathbf{A}\mathbf{t}$  can be either negative or null, matrix  $\mathbf{A}$  is *negative semidefinite*;
- if  $\mathbf{t}'\mathbf{A}\mathbf{t}$  can be either negative, null or positive, matrix  $\mathbf{A}$  is *indefinite*.

## 2.11 Singular value decomposition

Another useful method of matrix decomposition is *singular value decomposition* (SVD). The *approximation theorem* of Schmidt (1907), later rediscovered by Eckart & Young (1936), showed that any *rectangular* matrix  $\mathbf{Y}$  can be decomposed as follows:

$$\mathbf{Y}(n \times p) = \mathbf{V}(n \times k) \mathbf{W}(\text{diagonal}, k \times k) \mathbf{U}'(k \times p) \quad (2.31)$$

Singular value

where both  $\mathbf{U}$  and  $\mathbf{V}$  are orthonormal matrices (i.e. matrices containing column vectors that are normalized and orthogonal to one another; Section 2.8).  $\mathbf{W}$  is a diagonal matrix  $\mathbf{D}(w_i)$ , of order  $k = \min(n, p)$ , containing the *singular values*; the illustration hereunder assumes that  $n > p$  so that  $k = p$ . The notation  $\mathbf{D}(w_i)$  for the diagonal matrix of singular values will be used in the remainder of this section. The early history of singular value decomposition has been recounted by Stewart (1993). The following illustration shows the shapes of these matrices:

$$\begin{bmatrix} \mathbf{Y}_{(n \times p)} \end{bmatrix} = \begin{bmatrix} \mathbf{V}_{(n \times p)} \end{bmatrix} \begin{bmatrix} w_1 & 0 & 0 & \dots & 0 \\ 0 & w_2 & 0 & \dots & 0 \\ 0 & 0 & w_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & w_p \end{bmatrix} \begin{bmatrix} \mathbf{U}'_{(p \times p)} \end{bmatrix}$$

Demonstrating eq. 2.31 is beyond the scope of this book. The diagonal values  $w_i$  in  $\mathbf{D}(w_i)$  are non-negative; they are the singular values of  $\mathbf{Y}$ . SVD functions are found in advanced statistical languages such as R, S-PLUS® and MATLAB®. The notation used in different manuals and computer software may, however, differ from the one used here. That is the case of the R language, where function *svd()* is said to decomposes  $\mathbf{Y}$  into  $\mathbf{U}\mathbf{D}(w_i)\mathbf{V}'$ , instead of the notation  $\mathbf{V}\mathbf{D}(w_i)\mathbf{U}'$  used here to insure consistency between the results of eigenanalysis and SVD in Subsection 9.1.9.

**Application 1: Rank of a rectangular matrix.** — The rank of a rectangular matrix is equal to the number of singular values larger than 0. As an illustration, consider the matrix in Numerical example 2 of Section 2.7:

$$\mathbf{Y} = \begin{bmatrix} 2 & 1 & 3 & 4 \\ -1 & 6 & -3 & 0 \\ 1 & 20 & -3 & 8 \end{bmatrix}$$

In this example,  $(n = 3) < (p = 4)$ , hence  $k = n = 3$ , and the dimensions of matrices in eq. 2.31 are  $\mathbf{V}(3 \times 3)$ ,  $\mathbf{W}(3 \times 3)$  and  $\mathbf{U}'(3 \times 4)$ . Singular value decomposition of that matrix produces two singular values larger than zero and one null singular value. SVD of the transposed matrix produces the same singular values.  $\mathbf{Y}$  is thus of rank 2. After elimination of the third (null) singular value and the corresponding vector in both  $\mathbf{V}$  and  $\mathbf{U}'$ , the singular value decomposition of  $\mathbf{Y}$  gives:

$$\mathbf{Y} = \begin{bmatrix} -0.08682 & 0.84068 \\ -0.26247 & -0.53689 \\ -0.96103 & 0.07069 \end{bmatrix} \begin{bmatrix} 22.650 & 0 \\ 0 & 6.081 \end{bmatrix} \begin{bmatrix} -0.03851 & -0.92195 & 0.15055 & -0.35477 \\ 0.37642 & -0.15902 & 0.64476 & 0.64600 \end{bmatrix}$$

**Application 2: Decomposition of a cross-product matrix.** — A covariance matrix is a type of cross-product matrix (Chapter 4). Consider the covariance matrix  $\mathbf{S}$  of the data used to illustrate principal component analysis in Section 9.1. It is decomposed as follows by SVD:

$$\mathbf{S} = \mathbf{V} \mathbf{D}(w_i) \mathbf{U}'$$

$$\begin{bmatrix} 8.2 & 1.6 \\ 1.6 & 5.8 \end{bmatrix} = \begin{bmatrix} -0.8944 & -0.4472 \\ -0.4472 & 0.8944 \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} -0.8944 & -0.4472 \\ -0.4472 & 0.8944 \end{bmatrix}$$

The singular values of  $\mathbf{S}$ , found on the diagonal of  $\mathbf{D}(w_i)$ , are equal to the eigenvalues. This is true for any square symmetric matrix. Matrices  $\mathbf{V}$  and  $\mathbf{U}$  contain vectors identical to the eigenvectors obtained by eigenanalysis; eigenvectors may vary in their signs depending on the program or the computer platform. Negative eigenvalues, which may be found in principal coordinate analysis of symmetric distance matrices (PCoA, Section 9.3), will come out as singular values with positive signs. Example:

$$\text{matrix} \quad \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

has the singular values [9.6235, 6.2348, 0.0000] and the following set of eigenvalues: [9.6235, 0.0000, -6.2348]. The singular value 6.2348 with a positive sign corresponds to the negative eigenvalue -6.2348.

**Application 3: Generalized matrix inversion.** — SVD offers a way of inverting matrices that are singular (Section 2.8) or numerically very close to being singular. SVD may either give users a clear diagnostic of the problem, or solve it. Singularity may be encountered in regression for example: if the matrix of explanatory variables  $\mathbf{X}$  is not of full rank, the cross-product matrix  $\mathbf{A} = [\mathbf{X}'\mathbf{X}]$  is singular and it cannot be inverted with the methods described in Section 2.8, although inversion is necessary to solve eq. 2.19.

Inversion of  $\mathbf{A} = [\mathbf{X}'\mathbf{X}]$  by SVD involves the following steps. First,  $\mathbf{A}$  is decomposed using eq. 2.31:

$$\mathbf{A} = \mathbf{V} \mathbf{D}(w_i) \mathbf{U}'$$

Since  $\mathbf{A}$  is symmetric,  $\mathbf{V}$ ,  $\mathbf{D}(w_i)$ , and  $\mathbf{U}$  are all square matrices of the same size as  $\mathbf{A}$ . Using property 5 of matrix inverses (above), the inverse of  $\mathbf{A}$  is easy to compute:

$$\mathbf{A}^{-1} = [\mathbf{V} \mathbf{D}(w_i) \mathbf{U}']^{-1} = [\mathbf{U}']^{-1} [\mathbf{D}(w_i)]^{-1} [\mathbf{V}]^{-1}$$

Because  $\mathbf{U}$  and  $\mathbf{V}$  are orthonormal, their inverses are equal to their transposes (property 7 of inverses), whereas the inverse of a diagonal matrix is a diagonal matrix whose elements are the reciprocals of the original elements (property 8). Hence:

$$\mathbf{A}^{-1} = \mathbf{U} \mathbf{D}(1/w_i) \mathbf{V}' \quad (2.32)$$

Singular  
matrix

Ill-  
conditioned  
matrix

It may happen that one or more of the  $w_i$ 's are zero, so that their reciprocals are infinite;  $\mathbf{A}$  is then a *singular matrix*. This is what happens in the regression case when  $\mathbf{X}$  is not of full rank. It may also happen that one or more of the  $w_i$ 's are numerically so small that their values cannot be properly computed because of the machine's precision in floating-point calculation; in that case,  $\mathbf{A}$  is said to be *ill-conditioned*. When  $\mathbf{A}$  is singular, the columns of  $\mathbf{U}$  corresponding to the zero elements in  $\mathbf{D}(w_i)$  form an orthonormal basis\* for the space where the system of equations has no solution, whereas the columns of  $\mathbf{V}$  corresponding to the non-zero elements in  $\mathbf{D}(w_i)$  are an orthonormal basis for the space where the system has a solution. When  $\mathbf{A}$  is singular or ill-conditioned, it is still possible to find its inverse, either exactly or approximately, and use it to compute a regression model. Here is an example:

$$\mathbf{y} = \begin{bmatrix} 1.25 \\ 1.13 \\ 1.60 \\ 2.08 \\ 2.10 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 5 \\ 1 & 2 & 2 & 2 \\ 1 & 3 & 3 & 4 \\ 1 & 4 & 4 & 3 \\ 1 & 5 & 5 & 1 \end{bmatrix}$$

\* A set of  $k$  linearly independent vectors form a basis for a  $k$ -dimensional vector space. Any vector in that space can be uniquely written as a linear combination of the base vectors.

The first column of  $\mathbf{X}$  contains 1's to estimate the intercept. Columns 2 and 3 are identical, so  $\mathbf{X}$  is not of full rank. Equation 2.31 produces a decomposition of  $\mathbf{A} = [\mathbf{X}'\mathbf{X}]$  that has 3 (not 4) singular values larger than 0. A generalized inverse is obtained by computing eq. 2.32 after removing the last column from  $\mathbf{U}$  and  $\mathbf{V}$  and the last row and column from  $\mathbf{D}(w_i)$ :

$$\begin{aligned} \mathbf{A}^{-1} &= \mathbf{U}\mathbf{D}(1/w_i)\mathbf{V}' \\ &= \begin{bmatrix} -0.17891 & 0.07546 & 0.98097 \\ -0.59259 & -0.37762 & -0.07903 \\ -0.59259 & -0.37762 & -0.07903 \\ -0.51544 & 0.84209 & -0.15878 \end{bmatrix} \begin{bmatrix} 0.00678 & 0 & 0 \\ 0 & 0.04492 & 0 \\ 0 & 0 & 6.44243 \end{bmatrix} \begin{bmatrix} -0.17891 & -0.59259 & -0.59259 & -0.51544 \\ 0.07546 & -0.37762 & -0.37762 & 0.84209 \\ 0.98097 & -0.07903 & -0.07903 & -0.15878 \end{bmatrix} \\ &= \begin{bmatrix} 6.20000 & -0.50000 & -0.50000 & -1.00000 \\ -0.50000 & 0.04902 & 0.04902 & 0.06863 \\ -0.50000 & 0.04902 & 0.04902 & 0.06863 \\ -1.00000 & 0.06863 & 0.06863 & 0.19608 \end{bmatrix} \end{aligned}$$

Using the generalized inverse  $\mathbf{A}^{-1}$ , the regression coefficients can now be computed (eq. 2.19):

$$\mathbf{b} = [\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y} = \mathbf{A}^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0.21200 \\ 0.17539 \\ 0.17539 \\ 0.12255 \end{bmatrix}$$

The first value in vector  $\mathbf{b}$  is the intercept. Now remove column 2 from  $\mathbf{X}$  and compute a multiple linear regression equation. The regression coefficients are:

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.21200 \\ 0.35078 \\ 0.12255 \end{bmatrix}$$

The regression coefficient for the second column of  $\mathbf{X}$ , 0.35078, has been split in two equal coefficients of 0.17539 in the SVD solution when the two identical variables were kept in the analysis.

Similar problems may be encountered when solving sets of simultaneous linear equations represented by matrix equation  $\mathbf{A}\mathbf{b} = \mathbf{c}$  (Section 2.8). In this book, SVD will also be used in algorithms for principal component analysis (Subsection 9.1.9) and correspondence analysis (Subsection 9.2.1).



---

## 2.12 Software

Functions for all matrix operations described in this chapter are available in the R language. Standard matrix operations are available in the BASE package while more specialized operations are found in the MATRIX package.

Among the functions found in BASE are *det()* to compute a determinant, *solve()* for matrix inversion or solving a system of linear equations, *eigen()* for eigenvalue decomposition, and *svd()* for singular value decomposition. Other useful decompositions used in later chapters but not discussed in Chapter 2 are the QR decomposition (function *qr()* of BASE) and Cholesky factorization (functions *chol()* of BASE and MATRIX). Package MASS offers function *ginv()* for general inversion.

Functions implementing matrix algebra are also available in S-PLUS<sup>®</sup>, MATLAB<sup>®</sup> and SAS<sup>®</sup>.