MATRIX ALGEBRA

Hervé Abdi[*]

James Joseph Sylvester developed the modern concept of matrices in the 19th century. For him a matrix was simply an array of numbers. He worked with systems of linear equations and so matrices provided a convenient way of working with the coefficients of these linear equations. In the process, matrix algebra was created to generalize number operations to set of numbers called matrices. Nowadays, matrix algebra is used in all branches of mathematics and the sciences and constitutes the basis of most statistical procedures.

Matrices: Definition

A matrix is a set of numbers arranged in a table. For example, Toto, Marius, and Olivette are looking at their possessions, and they are counting how many balls, cars, coins, and novels they each possess. Toto has 2 balls, 5 cars, 10 coins, and 20 novels. Marius has 1, 2, 3, and 4, and Olivette has 6, 1, 3, and 10. These data can be displayed in a table in which each row represents a person and each column a possession:

Person	*Balls*	*Cars*	*Coins*	*Novels*
Toto	2	5	10	20
Marius	1	2	3	4
Olivette	6	1	3	10

We can also say that these data are described by the matrix denoted **A** equal to

$$\mathbf{A} = \begin{bmatrix} 2 & 5 & 10 & 20 \\ 1 & 2 & 3 & 4 \\ 6 & 1 & 3 & 10 \end{bmatrix} \qquad (1)$$

Matrices are denoted by boldface uppercase letters.

To identify a specific element of a matrix, we use its row and column numbers. For example, the cell defined by Row 3 and Column 1 contains the value 6: We write that $a_{3,1} = 6$. With this notation, elements of a matrix are denoted with the same letter as the matrix but

---

[*] In Frey, B. (2022). *The SAGE Encyclopedia of Research Design*. Thousand Oaks: Sage.
Send correspondence to Hervé Abdi, Ph.D.
The University of Texas at Dallas, School of Behavioral and Brain Sciences
herve@utdallas.edu https://personal.utdallas.edu/~herve

written in lowercase italic. The first subscript always gives the row number of the element (i.e., 3), and second subscript always gives its column number (i.e., 1).

A generic element of a matrix is identified with indices such as i and j. So, $a_{i,j}$ is the element at the ith row and jth column of **A**. The *total* number of rows and columns is denoted with the same letters as the indices but in uppercase letters. The matrix **A** has I rows (here $I = 3$) and J columns (here $J = 4$), and it is made of $I \times J$ elements $a_{i,j}$ (here $3 \times 4 = 12$). The term "dimensions" is often used to refer to the number of rows and columns, so **A** has dimensions I by J.

As a shortcut, a matrix can be represented by its generic element written in brackets. So, the matrix **A** with I rows and J columns is denoted

$$\mathbf{A} = [a_{i,j}] = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,j} & \cdots & a_{1.J} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,j} & \cdots & a_{2.J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,1} & a_{i,2} & \cdots & a_{i,j} & \cdots & a_{i,\mathrm{J}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{I,1} & a_{I,1} & \cdots & a_{I,j} & \cdots & a_{I,J} \end{bmatrix} \quad (2)$$

For either convenience or clarity, the number of rows and columns can also be indicated as subscripts below the matrix name:

$$\mathbf{A} = \underset{I \times J}{\mathbf{A}} = [a_{i,j}]. \quad (3)$$

Vectors

A matrix with one column is called a *column vector* or simply a vector. Vectors are denoted with bold lowercase letters. For example, the first column of matrix **A** (of Equation 1) is a column vector that stores the number of balls of Toto, Marius, and Olivette. We can call it **b** (for balls), and so

$$\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix} \quad (4)$$

Vectors are the building blocks of matrices. For example, **A** (of Equation 1) is made of four column vectors, which represent the number of balls, cars, coins, and novels, respectively.

Norm of a Vector

We can associate to a vector a quantity, related to its variance and standard deviation, called the *norm* or *length*. The norm of a vector is the square root of the sum of squares of the elements. It is denoted by putting the name of the vector between a set of double bars (||). For example, for

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \qquad (5)$$

we find

$$||\mathbf{x}|| = \sqrt{2^2 + 1^2 + 2^2} = \sqrt{4 + 1 + 4} = \sqrt{9} = 3 \qquad (6)$$

Normalization of a Vector

A vector is normalized when its norm is equal to 1. To normalize a vector, we divide each of its elements by its norm. For example, vector **x** from Equation 5 is transformed into the normalized vector $\bar{\mathbf{x}}$ as

$$\bar{\mathbf{x}} = \frac{\mathbf{x}}{||\mathbf{x}||} \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{bmatrix} \qquad (7)$$

Operations for Matrices

Transposition

If we exchange the roles of the rows and the columns of a matrix, we *transpose* it. This operation is called the *transposition,* and the new matrix is called a *transposed* matrix. The **A** matrix transposed is denoted $\mathbf{A}^{\mathrm{T}}$. For example:

$$\mathbf{A} = \underset{3\times4}{\mathbf{A}} = \begin{bmatrix} 2 & 5 & 10 & 20 \\ 1 & 2 & 3 & 4 \\ 6 & 1 & 3 & 10 \end{bmatrix}, \text{then}$$

$$\mathbf{A}^{\mathrm{T}} = \underset{4\times3}{\mathbf{A}}^{T} = \begin{bmatrix} 2 & 1 & 6 \\ 5 & 2 & 1 \\ 10 & 3 & 3 \\ 20 & 4 & 10 \end{bmatrix}$$

Addition of Matrices

When two matrices have the same dimensions, we compute their sum by adding the corresponding elements. For example, with

$$\mathbf{A} = \begin{bmatrix} 2 & 5 & 10 & 20 \\ 1 & 2 & 3 & 4 \\ 6 & 1 & 3 & 10 \end{bmatrix} \text{and}$$

$$\mathbf{B} = \begin{bmatrix} 3 & 4 & 5 & 6 \\ 2 & 4 & 6 & 8 \\ 1 & 2 & 3 & 5 \end{bmatrix}, \qquad (9)$$

we find that

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 2+3 & 5+4 & 10+5 & 20+6 \\ 1+2 & 2+4 & 3+6 & 4+8 \\ 6+1 & 1+2 & 3+3 & 10+5 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 9 & 15 & 26 \\ 3 & 6 & 9 & 12 \\ 7 & 3 & 6 & 15 \end{bmatrix}. \qquad (10)$$

In general

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{1,1}+b_{1,1} & a_{1,2}+b_{1,2} & \cdots & a_{1,j}+b_{1,j} & \cdots & a_{1,J}+b_{1,J} \\ a_{2,1}+b_{2,1} & a_{2,2}+b_{2,2} & \cdots & a_{2,j}+b_{2,j} & \cdots & a_{2,J}+b_{2,J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,1}+b_{i,1} & a_{i,2}+b_{i,2} & \cdots & a_{i,j}+b_{i,j} & \cdots & a_{i,J}+b_{i,J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{I,1}+b_{I,1} & a_{I,2}+b_{I,2} & \cdots & a_{I,j}+b_{I,j} & \cdots & a_{I,J}+b_{I,J} \end{bmatrix}. \qquad (11)$$

Matrix addition behaves very much like usual addition. Specifically, matrix addition is commutative (i.e., **A** + **B** = **B** + **A**) and associative (i.e., A+ **[B** + **C]** = **[A** + **B]** + **C**).

Multiplication of a Matrix by a Scalar

To differentiate matrices from the usual numbers, we call usual numbers *scalar numbers* or simply *scalars*. To multiply a matrix by a scalar, we multiply each element of the matrix by this scalar. For example

$$10\times\mathbf{B} = 10\times\begin{bmatrix}3 & 4 & 5 & 6\\ 2 & 4 & 6 & 8\\ 1 & 2 & 3 & 5\end{bmatrix} \qquad (12)$$

$$= \begin{bmatrix}30 & 40 & 50 & 60\\ 20 & 40 & 60 & 80\\ 10 & 20 & 30 & 50\end{bmatrix}.$$

Multiplication: Product or Products?

There are *several* ways of generalizing the concept of *product* or *multiplication* to matrices. We will look at the most frequently used of these matrix products. Each of these products will behave like the product between scalars when the matrices have dimensions 1×1.

Hadamard Product

When generalizing product to matrices, the first approach is to multiply the corresponding elements of the two matrices that we want to multiply. This is called the *Hadamard product,* denoted by $\odot$. The Hadamard product exists only for matrices with the same dimensions. Formally, it is defined as shown below:

$$\boldsymbol{A}\odot\boldsymbol{B} = [a_{i,j}\times b_{i,j}]$$

$$= \begin{bmatrix} a_{1,1}\times b_{1,1} a_{1,2}\times b_{1,2}\cdots a_{1,j}\times b_{1,j}\cdots a_{1,J}\times b_{1,J}\\ a_{2,1}\times b_{2,1} a_{2,2}\times b_{2,2}\cdots a_{2,j}\times b_{2,j}\cdots a_{2,J}\times b_{2,J}\\ \vdots\vdots\ddots\vdots\ddots\vdots\\ a_{i,1}\times b_{i,1} a_{i,2}\times b_{i,2}\cdots a_{i,j}\times b_{1,j}\cdots a_{i,J}\times b_{i,J}\\ \vdots\vdots\ddots\vdots\ddots\vdots\\ a_{I,1}\times b_{I,1} a_{I,2}\times b_{I,2}\cdots a_{I,j}\times b_{I,j}\cdots a_{I,J}\times b_{I,J}\end{bmatrix} \quad (13)$$

For example, with

$$\mathbf{A} = \begin{bmatrix}2 & 5 & 10 & 20\\ 1 & 2 & 3 & 4\\ 6 & 1 & 3 & 10\end{bmatrix} \text{and}$$

$$\mathbf{B} = \begin{bmatrix}3 & 4 & 5 & 6\\ 2 & 4 & 6 & 8\\ 1 & 2 & 3 & 5\end{bmatrix}, \qquad (14)$$

we get

$$\mathbf{A}\odot\mathbf{B} = \begin{bmatrix}2\times 3 & 5\times 4 & 10\times 5 & 20\times 6\\ 1\times 2 & 2\times 2 & 3\times 6 & 4\times 8\\ 6\times 1 & 1\times 1 & 3\times 3 & 10\times 5\end{bmatrix} = \begin{bmatrix}6 & 20 & 50 & 120\\ 2 & 4 & 18 & 32\\ 6 & 1 & 9 & 50\end{bmatrix}. \qquad (15)$$

Standard or Cayley Product

The Hadamard product is straightforward, but it is *not* the matrix product that is used most often. The product most often used is called the *standard* or *Cayley* product, or simply *the* product (i.e., when the name of the product is not specified, it is the standard product). The definition of the Cayley product comes from the original use of matrices to solve equations. Its definition looks surprising at first because it is defined only when the number of columns of the first matrix is equal to the number of rows of the second matrix. When two matrices can be multiplied together, they are called *conformable.* This product will have the number of rows of the *first* matrix and the number of columns of the *second* matrix.

So, **A** with I rows and J columns can be multiplied by **B** with J rows and K columns to give **C** with I rows and K columns. A convenient way of checking that two matrices are conformable is to write the dimensions of the matrices as subscripts. For example

$$\underset{I\times J}{\mathbf{A}} \times \underset{J\times K}{\mathbf{B}} = \underset{I\times K}{\mathbf{C}}, \tag{16}$$

or even

$$\underset{I\times J}{\mathbf{A}}\,\underset{J\times K}{\mathbf{B}} = \underset{I\times K}{\mathbf{C}} \tag{17}$$

An element $c_{i,\,k}$ of matrix **C** is computed ass computed as

$$c_{i,k} = \sum_{j=1}^{J} a_{i,j} \times b_{j,k}. \tag{18}$$

So, $c_{i\,k}$ is the sum of J terms, each term being the product of the corresponding element of the ith row of **A** with the kth column of **B**.

For example, let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 5 \end{bmatrix} \text{and } \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}. \tag{19}$$

The product of these matrices is denoted $\mathbf{C} = \mathbf{A} \times \mathbf{B} = \mathbf{AB}$ (the × sign can be omitted when the context is clear). To compute $c_{2,1}$ we add three terms: (1) the product of the first element of the second row of **A** (i.e., 4) with the first element of the first column of **B** (i.e., 1); (2) the product of the second element of the second row of **A** (i.e., 5) with the second element of the first column of **B** (i.e., 3); and (3) the product of the third element of the second row of **A** (i.e., 5) with the third element of the first column of **B** (i.e., 5). Formally, the term $c_{2,1}$ is obtained as

$$
\begin{aligned}
c_{2,1} &= \sum_{j=1}^{J=3} a_{2,j} \times b_{j,1} \\
&= (a_{2,1}) \times (b_{1,1}) + (a_{2,2} \times b_{2,1}) + (a_{2,3} \times b_{3,1}) \\
&= (4 \times 1) + (5 \times 3) + (6 \times 5) \\
&= 49. \qquad (20)
\end{aligned}
$$

Matrix **C** is obtained as

$$
\begin{aligned}
\mathbf{AB} = \mathbf{C} &= [c_{i,k}] \\
&= \begin{bmatrix} 1 \times 1 + 2 \times 3 + 3 \times 5 & 1 \times 2 + 2 \times 4 + 3 \times 6 \\ 4 \times 1 + 5 \times 3 + 6 \times 5 & 4 \times 2 + 5 \times 4 + 6 \times 6 \end{bmatrix} \\
&= \begin{bmatrix} 22 & 28 \\ 49 & 64 \end{bmatrix} \qquad (21)
\end{aligned}
$$

Properties of the Product

Like the product between scalars, the product between matrices is *associative,* and *distributive* relative to addition. Specifically, for any set of three conformable matrices **A, B,** and **C**:

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) = \mathbf{ABC} \quad \text{(associativity)} \qquad (22)$$

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \quad \text{(distributivity).} \qquad (23)$$

The matrix products **AB** and **BA** do not always exist, but when they do, these products are *not,* in general, *commutative. so:*

$$\mathbf{AB} \neq \mathbf{BA}. \qquad (24)$$

For example, with

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 1 & -1 \\ -2 & 2 \end{bmatrix} (25)$$

we get:

$$\mathbf{AB} = \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \qquad (26)$$

But

$$\mathbf{BA} = \begin{bmatrix} 1 & -1 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ -8 & -4 \end{bmatrix}. \qquad (27)$$

Incidentally, we can combine transposition and product and get the following equation:

$$(\mathbf{AB})^{\mathsf{T}} = \mathbf{B}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}. \tag{28}$$

Exotic Product: Kronecker

Another product is the *Kronecker* product, also called the *direct, tensor,* or *Zehfuss* product. It is denoted $\otimes$ and is defined for all matrices. Specifically, with two matrices $\mathbf{A} = [a_{i,j}]$ (with dimensions I by J) and $\mathbf{B} = [b_{k,l}]$ (with dimensions K and L), the Kronecker product gives a matrix **C,** with dimensions $(I \times K)$ by $(J \times L)$, defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \cdots & a_{1,j}\mathbf{B} & \cdots & a_{1,J}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \cdots & a_{2,j}\mathbf{B} & \cdots & a_{2,J}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,1}\mathbf{B} & a_{i,2}\mathbf{B} & \cdots & a_{i,j}\mathbf{B} & \cdots & a_{i,J}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{I,1}\mathbf{B} & a_{I,2}\mathbf{B} & \cdots & a_{I,j}\mathbf{B} & \cdots & a_{I,J}\mathbf{B} \end{bmatrix}. \tag{29}$$

For example, with

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix} \tag{30}$$

we get

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} 1\times6 & 1\times7 & 2\times6 & 2\times7 & 3\times6 & 3\times7 \\ 1\times8 & 1\times9 & 2\times8 & 2\times9 & 3\times8 & 3\times9 \end{bmatrix}$$
$$= \begin{bmatrix} 6 & 7 & 12 & 14 & 18 & 21 \\ 8 & 9 & 16 & 18 & 24 & 27 \end{bmatrix}. \tag{31}$$

The Kronecker product is used to write design matrices. It is an essential tool for the derivation of expected values and sampling distributions.

Special Matrices

Certain special matrices have specific names.

Square and Rectangular Matrices

A matrix with the same number of rows and columns is a *square matrix.* By contrast, a matrix with different numbers of rows and columns is a *rectangular matrix.* So

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 5 \\ 7 & 8 & 0 \end{bmatrix} \qquad (32)$$

is a square matrix, but

$$\mathbf{B} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \qquad (33)$$

is a rectangular matrix.

Symmetric Matrix

A square matrix **A** with $a_{i,j} = a_{j,i}$ is *symmetric*. So

$$\mathbf{A} = \begin{bmatrix} 10 & 2 & 3 \\ 2 & 20 & 5 \\ 3 & 5 & 30 \end{bmatrix} \qquad (34)$$

is symmetric, but

$$\mathbf{A} = \begin{bmatrix} 12 & 2 & 3 \\ 4 & 20 & 5 \\ 7 & 8 & 30 \end{bmatrix} (35)$$

is not.

Note that for a symmetric matrix,

$$\mathbf{A} = \mathbf{A}^{\mathrm{T}}. \qquad (36)$$

A common mistake is to assume that the standard product of two symmetric matrices is commutative. But this is not true, as shown by the following example. With

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 1 \end{bmatrix} \text{and } \mathbf{B} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 3 \\ 2 & 3 & 1 \end{bmatrix} \qquad (37)$$

we get

$$\mathbf{AB} = \begin{bmatrix} 9 & 12 & 11 \\ 11 & 15 & 11 \\ 9 & 10 & 19 \end{bmatrix}, \text{but } \mathbf{BA} = \begin{bmatrix} 9 & 11 & 9 \\ 12 & 15 & 10 \\ 11 & 11 & 19 \end{bmatrix}. \qquad (38)$$

Note, however, that combining Equations 35 and 43 gives for symmetric matrices **A** and **B** the following equation:

$$\mathbf{AB} = (\mathbf{BA})^{\mathrm{T}}. \tag{39}$$

Diagonal Matrix

A square matrix is *diagonal* when all its elements are zero except the elements on the diagonal. Formally, a matrix is diagonal if $a_{i,j} = 0$ when $i \neq j$. Thus

$$\mathbf{A} = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 30 \end{bmatrix} \text{ is diagonal.} \tag{40}$$

Because only the diagonal elements matter for a diagonal matrix, we can specify only these diagonal elements. This is done with the following notation:

$$\begin{aligned} \mathbf{A} &= \operatorname{diag}\left\{\left[a_{1,1}, \ldots, a_{i,i}, \ldots, a_{I,I}\right]\right\} \\ &= \operatorname{diag}\left\{\left[a_{i,i}\right]\right\}. \end{aligned} \tag{41}$$

For example, the previous matrix can be rewritten as:

$$\mathbf{A} = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 30 \end{bmatrix} = \operatorname{diag}\{[10, 20, 30]\}. \tag{42}$$

The operator diag can also be used to isolate the diagonal of any square matrix. For example, with

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \tag{43}$$

we get

$$\operatorname{diag}\{\mathbf{A}\} = \operatorname{diag}\left\{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}\right\} = \begin{bmatrix} 1 \\ 5 \\ 9 \end{bmatrix}. \tag{44}$$

Note, incidentally, that

$$\operatorname{diag}\{\operatorname{diag}\{\mathbf{A}\}\} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{bmatrix}. \tag{45}$$

Multiplication by a Diagonal Matrix

Diagonal matrices are often used to multiply by a scalar all the elements of a given row or column. Specifically, when we premultiply a matrix by a diagonal matrix, the elements of the row of the second matrix are multiplied by the corresponding diagonal element. Likewise, when we postmultiply a matrix by a diagonal matrix, the elements of the column of the first matrix are multiplied by the corresponding diagonal element. For example, with:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{bmatrix}, \quad (46)$$

we get

$$\mathbf{BA} = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ 20 & 25 & 30 \end{bmatrix} \quad (47)$$

and

$$\mathbf{AC} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 8 & 18 \\ 8 & 20 & 36 \end{bmatrix} \quad (48)$$

and also

$$\mathbf{BAC} = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{bmatrix} = \begin{bmatrix} 4 & 16 & 36 \\ 40 & 100 & 180 \end{bmatrix}. \quad (49)$$

Identity Matrix

A diagonal matrix whose diagonal elements are all equal to 1 is called an *identity* matrix and is denoted **I**. If we need to specify its dimensions, we use subscripts such as, for example:

$$\underset{3\times3}{\mathbf{I}} = \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{(this is a } 3\times3 \text{ identity matrix).} \quad (50)$$

The identity matrix is the neutral element for the standard product. So

$$\mathbf{I} \times \mathbf{A} = \mathbf{A} \times \mathbf{I} = \mathbf{A} \quad (51)$$

for any matrix **A** conformable with **I**. For example:

$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\times\begin{bmatrix}1&2&3\\4&5&5\\7&8&0\end{bmatrix}=$$

$$\begin{bmatrix}1&2&3\\4&5&5\\7&8&0\end{bmatrix}\times\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}=\begin{bmatrix}1&2&3\\4&5&5\\7&8&0\end{bmatrix}. \qquad (52)$$

Matrix Full of Ones

A matrix whose elements are all equal to 1 is denoted by **1** or, when we need to specify its dimensions by writing, for example: $\underset{I\times J}{\mathbf{1}}$.

These matrices are neutral elements for the Hadamard product. So

$$\underset{2\times3}{\mathbf{A}}\odot\underset{2\times3}{\mathbf{1}}=\begin{bmatrix}1&2&3\\4&5&6\end{bmatrix}\odot\begin{bmatrix}1&1&1\\1&1&1\end{bmatrix} \qquad (53)$$

$$=\begin{bmatrix}1\times1&2\times1&3\times1\\4\times1&5\times1&6\times1\end{bmatrix}=\begin{bmatrix}1&2&3\\4&5&6\end{bmatrix}. \qquad (54)$$

The **1** matrices or vectors can also be used to compute sums of rows or columns:

$$\begin{bmatrix}1&2&3\end{bmatrix}\times\begin{bmatrix}1\\1\\1\end{bmatrix}=(1\times1)+(2\times1)+(3\times1)$$

$$=1+2+3=6, \qquad (55)$$

or also

$$\begin{bmatrix}1&1\end{bmatrix}\times\begin{bmatrix}1&2&3\\4&5&6\end{bmatrix}=\begin{bmatrix}5&7&9\end{bmatrix}. \qquad (56)$$

Matrix Full of Zeros

A matrix whose elements are all equal to 0 is the *null* or *zero* matrix. It is denoted by **0** or, when we need to specify its dimensions, by $\underset{I\times J}{\mathbf{0}}$. Null matrices are neutral elements for addition:

$$\begin{bmatrix}1&2\\3&4\end{bmatrix}+\underset{2\times2}{\mathbf{0}}=\begin{bmatrix}1+0&2+0\\3+0&4+0\end{bmatrix}=\begin{bmatrix}1&2\\3&4\end{bmatrix}. \qquad (57)$$

Null matrices are also null elements for the Hadamard product:

$$\begin{bmatrix}1&2\\3&4\end{bmatrix}\odot\underset{2\times2}{\mathbf{0}}=\begin{bmatrix}1\times0&2\times0\\3\times0&4\times0\end{bmatrix}=\begin{bmatrix}0&0\\0&0\end{bmatrix}=\underset{2\times2}{\mathbf{0}} \qquad (58)$$

As well as for the standard product:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \underset{2\times2}{\mathbf{0}} = \begin{bmatrix} 1\times0+2\times0 & 1\times0+2\times0 \\ 3\times0+4\times0 & 3\times0+4\times0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \underset{2\times2}{\mathbf{0}}. \qquad (59)$$

Triangular Matrix

A matrix is lower triangular when $a_{i,j} = 0$ for $i < j$. A matrix is upper triangular when $a_{i,j} = 0$ for $i > j$. For example,

$$\mathbf{A} = \begin{bmatrix} 10 & 0 & 0 \\ 2 & 20 & 0 \\ 3 & 5 & 30 \end{bmatrix} \text{ is lower triangular,} \qquad (60)$$

and

$$\mathbf{B} = \begin{bmatrix} 12 & 2 & 3 \\ 0 & 20 & 5 \\ 0 & 0 & 30 \end{bmatrix} \text{ is upper triangular.} \qquad (61)$$

Cross-Product Matrix

A cross-product matrix is obtained by multiplication of a matrix by its transpose. Therefore, a cross-product matrix is square and symmetric. For example, the matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 4 \end{bmatrix} \qquad (62)$$

premultiplied by its transpose

$$\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 4 \end{bmatrix} \qquad (63)$$

gives the following cross-product matrix:

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \begin{bmatrix} 1\times1+2\times2+3\times3 & 1\times1+2\times4+3\times4 \\ 1\times1+4\times2+4\times3 & 1\times1+4\times4+4\times4 \end{bmatrix} = \begin{bmatrix} 14 & 21 \\ 14 & 33 \end{bmatrix}. \qquad (64)$$

A Particular Case of Cross-Product Matrix: Variance–Covariance Matrix

A particular case of cross-product matrices is correlation or covariance matrices. A variance–covariance matrix is obtained from a data matrix by three steps: (1) subtract the mean of each column from each element of this column (this is *centering*), (2) compute the cross-product matrix from the centered matrix, and (3) divide each element of the cross-product matrix by the number of rows of the data matrix. For example, if we take the $I = 3$ by $J = 2$ matrix **A**,

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 5 & 10 \\ 8 & 10 \end{bmatrix}, \qquad (65)$$

we obtain the means of each column as

$$\begin{aligned} \mathbf{m} &= \frac{1}{I} \times \underset{1\times I}{\mathbf{1}} \times \underset{I\times J}{\mathbf{A}} \\ &= \frac{1}{3} \times \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 2 & 1 \\ 5 & 10 \\ 8 & 10 \end{bmatrix} = \begin{bmatrix} 5 & 7 \end{bmatrix}. \end{aligned} \qquad (66)$$

To center the matrix, we subtract the mean of each column from all its elements. This centered matrix gives the deviations of each element from the mean of its column. Centering is performed as

$$\mathbf{D} = \mathbf{A} - \underset{J\times 1}{\mathbf{1}} \times \mathbf{m} = \begin{bmatrix} 2 & 1 \\ 5 & 10 \\ 8 & 10 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 5 & 7 \end{bmatrix} \qquad (67)$$

$$= \begin{bmatrix} 2 & 1 \\ 5 & 10 \\ 8 & 10 \end{bmatrix} - \begin{bmatrix} 5 & 7 \\ 5 & 7 \\ 5 & 7 \end{bmatrix} = \begin{bmatrix} -3 & -6 \\ 0 & 3 \\ 3 & 3 \end{bmatrix}. \qquad (68)$$

We denote by **S** the variance–covariance matrix derived from **A**. It is computed as

$$\begin{aligned} \mathbf{S} &= \frac{1}{I}\mathbf{D}^{\mathsf{T}}\mathbf{D} = \frac{1}{3}\begin{bmatrix} -3 & 0 & 3 \\ -6 & 3 & 3 \end{bmatrix} \times \begin{bmatrix} -3 & -6 \\ 0 & 3 \\ 3 & 3 \end{bmatrix} \\ &= \frac{1}{3} \times \begin{bmatrix} 18 & 27 \\ 27 & 54 \end{bmatrix} = \begin{bmatrix} 6 & 9 \\ 9 & 18 \end{bmatrix}. \end{aligned} \qquad (69)$$

(Variances are on the diagonal; covariances are off-diagonal.)

The Inverse of a Square Matrix

An operation similar to division exists, but only for (some) square matrices. This operation uses the notion of inverse operation and defines the inverse of a matrix. The inverse is defined by analogy with the scalar number case, for which division actually corresponds to multiplication by the inverse, namely,

$$\frac{a}{b} = a \times b^{-1} \text{ with } b \times b^{-1} = 1. \qquad (70)$$

The inverse of a square matrix **A** is denoted $\mathbf{A}^{-1}$. It has the following property:

$$\mathbf{A} \times \mathbf{A}^{-1} = \mathbf{A}^{-1} \times \mathbf{A} = \mathbf{I}. \qquad (71)$$

The definition of the inverse of a matrix is simple, but its computation is complicated and is best left to computers.

As an example, for

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad (72)$$

the inverse is:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & -2 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (73)$$

All square matrices do not necessarily have an inverse. The inverse of a matrix does not exist if the rows (and the columns) of this matrix are linearly dependent. For example, this matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 4 & 2 \\ 1 & 0 & 2 \\ 2 & 1 & 3 \end{bmatrix} \qquad (74)$$

does not have an inverse because the second column is a linear combination of the two other columns:

$$\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} = 2 \times \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}. \qquad (75)$$

A matrix without an inverse is called *singular*. When $\mathbf{A}^{-1}$ exists, it is unique.

Inverse matrices are used for solving linear equations and least square problems in, for example, multiple regression analysis, analysis of variance, and, of course, multivariate analysis.

Inverse of a Diagonal Matrix

The inverse of a diagonal matrix is easy to compute: The inverse of

$$\mathbf{A} = \operatorname{diag}\{a_{i,i}\} \qquad (76)$$

is the diagonal matrix

$$\mathbf{A}^{-1} = \operatorname{diag}\{a_{i,i}^{-1}\} = \operatorname{diag}\{1/a_{i,i}\}. \qquad (77)$$

For example,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & 4 \end{bmatrix} \text{and} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & .25 \end{bmatrix} \qquad (78)$$

are inverse of each other.

The Big Tool: The Eigen-decomposition

So far, matrix operations are very similar to operations with numbers. The next notion is specific to matrices. This is the idea of decomposing a matrix into simpler matrices. A lot of the power of matrices follows from this. A first decomposition is called the *eigen-decomposition,* and it applies only to square matrices. The generalization of the eigen-decomposition to rectangular matrices is called the *singular value* decomposition.

Eigenvectors and *eigenvalues* are numbers and vectors associated with square matrices. Together they constitute the eigen-decomposition. Even though the eigen-decomposition does not exist for all square matrices, it has a particularly simple expression for a class of matrices often used in multivariate analysis such as correlation, covariance, or cross-product matrices. The eigen-decomposition of these matrices is important in statistics because it is used to find the maximum (or minimum) of functions involving these matrices. For example, principal components analysis is obtained from the eigen-decomposition of a covariance or correlation matrix and gives the least square estimate of the original data matrix.

Notations and Definition

An eigenvector of matrix **A** is a vector **u** that satisfies the following equation:

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u}, \qquad (79)$$

where λ is a scalar called the *eigenvalue* associated to the eigenvector. When rewritten, Equation 79 becomes

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = 0. \qquad (80)$$

Therefore, **u** is eigenvector of **A** if the multiplication of **u** by **A** changes the length of **u** but not its orientation. For example,

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \qquad (81)$$

has for eigenvectors

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \text{with eigenvalue} \quad \lambda_1 = 4 \qquad (82)$$

and

$$\mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \text{ with eigenvalue } \lambda_2 = -1 \qquad (83)$$

When $\mathbf{u}_1$ and $\mathbf{u}_2$ are multiplied by **A**, only their length changes. That is,

$$\mathbf{A}\mathbf{u}_1 = \lambda_1\mathbf{u}_1 = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 12 \\ 8 \end{bmatrix} = 4\begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad (84)$$

and

$$\mathbf{A}\mathbf{u}_2 = \lambda_2\mathbf{u}_2 = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1\begin{bmatrix} -1 \\ 1 \end{bmatrix}. \qquad (85)$$

This is illustrated in Figure 1.

For convenience, eigenvectors are generally normalized such that

$$\mathbf{u}^{\mathrm{T}}\mathbf{u} = 1. \qquad (86)$$

For the previous example, normalizing the eigenvectors gives

$$\mathbf{u}_1 = \begin{bmatrix} .8321 \\ .5547 \end{bmatrix} \text{and } \mathbf{u}_2 = \begin{bmatrix} -.7071 \\ .7071 \end{bmatrix}. \qquad (87)$$

We can check that

$$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} .8321 \\ .5547 \end{bmatrix} = \begin{bmatrix} 3.3284 \\ 2.2188 \end{bmatrix} = 4\begin{bmatrix} .8321 \\ .5547 \end{bmatrix} \qquad (88)$$

and

$$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} -.7071 \\ .7071 \end{bmatrix} = \begin{bmatrix} .7071 \\ -.7071 \end{bmatrix} = -1\begin{bmatrix} -.7071 \\ .7071 \end{bmatrix}. \qquad (89)$$

Eigenvector and Eigenvalue Matrices

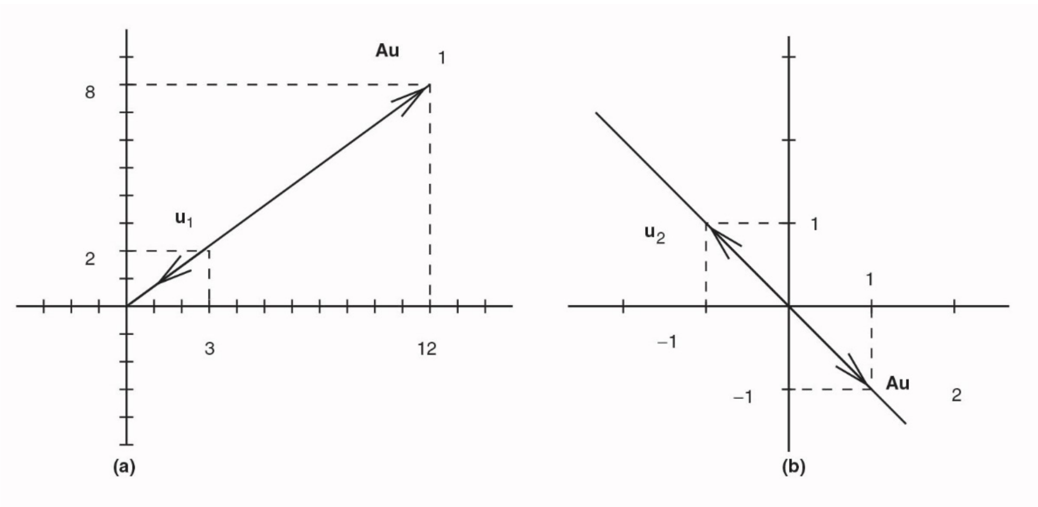
Traditionally, we store the eigenvectors of matrix **A** as the columns of a matrix denoted **U**. Eigenvalues are stored in a diagonal matrix (denoted **Λ,** which is the upper case Greek letter Lambda). Therefore, Equation 79 becomes

$$\mathbf{AU} = \mathbf{U\Lambda}. \qquad (90)$$

For example, with **A** (from Equation 81), we have

$$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix}. \qquad (91)$$

Figure 1 Two Eigenvectors of a Matrix



Reconstitution of a Matrix

The eigen-decomposition can also be used to build back a matrix from its eigenvectors and eigenvalues. This is shown by rewriting Equation 90 as

$$\mathbf{A} = \mathbf{U\Lambda U}^{-1}. \qquad (92)$$

For example, because

$$\mathbf{U}^{-1} = \begin{bmatrix} .2 & .2 \\ -.4 & .6 \end{bmatrix},$$

we obtain

$$\mathbf{A} = \mathbf{U\Lambda U}^{-1} = \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} .2 & .2 \\ -.4 & .6 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}. \qquad (93)$$

Digression: An Infinity of Eigenvectors for One Eigenvalue

It is only through a slight abuse of language that we talk about *the* eigenvector associated with *one* eigenvalue. Any scalar multiple of an eigenvector is an eigenvector, so for each eigenvalue, there is an infinite number of eigenvectors, all proportional to each other. For example,

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{94}$$

is an eigenvector of **A**:

$$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}. \tag{95}$$

Therefore,

$$2 \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \end{bmatrix} \tag{96}$$

is also an eigenvector of **A:**

$$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} = -1 \times 2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{97}$$

Positive (Semi)Definite Matrices

Some matrices, such as $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, do not have eigenvalues. Fortunately, the matrices used often in statistics belong to a category called *positive semidefinite.* The eigen-decomposition of these matrices always exists and has a particularly convenient form. A matrix is positive semidefinite when it can be obtained as the product of a matrix by its transpose. This implies that a positive semidefinite matrix is always symmetric. So, formally, the matrix **A** is positive semidefinite if it can be obtained as

$$\mathbf{A} = \mathbf{X}\mathbf{X}^{\mathrm{T}} \tag{98}$$

for a certain matrix **X**. Positive semidefinite matrices include correlation, covariance, and cross-product matrices.

The eigenvalues of a positive semidefinite matrix are always positive or null, and the eigenvectors of a positive semidefinite matrix are composed of real values and are pairwise orthogonal when their eigenvalues are different. This implies the following equality:

$$\mathbf{U}^{-1} = \mathbf{U}^{\mathrm{T}}.(99)$$

We can, therefore, express the positive semidefinite matrix **A** as

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^{\mathrm{T}} \tag{100}$$

where $\mathbf{U}$ (with $\mathbf{U}^T\mathbf{U} = \mathbf{I}$) are the normalized eigenvectors. For example

$$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \tag{101}$$

can be decomposed as

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T = \begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \tag{102}$$

with

$$\begin{bmatrix} \sqrt{(\frac{1}{2})} & \sqrt{(\frac{1}{2})} \\ \sqrt{(\frac{1}{2})} & -\sqrt{(\frac{1}{2})} \end{bmatrix} \times \begin{bmatrix} \sqrt{(\frac{1}{2})} & \sqrt{(\frac{1}{2})} \\ \sqrt{(\frac{1}{2})} & -\sqrt{(\frac{1}{2})} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{103}$$

Diagonalization

When a matrix is positive semidefinite, we can rewrite Equation 100 as

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \Leftrightarrow \Lambda = \mathbf{U}^T\mathbf{A}\mathbf{U}. \tag{104}$$

This shows that we can transform $\mathbf{A}$ into a *diagonal* matrix, and so the eigen-decomposition of a positive semidefinite matrix is often called its *diagonalization.*

Another Definition for Positive Semidefinite Matrices

A matrix $\mathbf{A}$ is positive semidefinite if for any nonzero vector $\mathbf{x}$, we have

$$\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0 \qquad \forall \mathbf{x}\ . \tag{105}$$

When *all* the eigenvalues of a matrix are positive, the matrix is *positive definite.* In that case, Equation 105 becomes

$$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0 \qquad \forall \mathbf{x}\ . \tag{106}$$

Trace, Determinant, and Rank

The eigenvalues of a matrix are closely related to three important numbers associated to a square matrix: the *trace,* the *determinant,* and the *rank.*

Trace

The trace of $\mathbf{A}$, denoted $\text{trace}\{\mathbf{A}\}$, is the sum of its diagonal elements. For example, with

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad (107)$$

we obtain

$$\text{trace}\{\mathbf{A}\} = 1 + 5 + 9 = 15. \quad (108)$$

The trace of a matrix is also equal to the sum of its eigenvalues:

$$\text{trace}\{\mathbf{A}\} = \sum_{\ell} \lambda_{\ell} = \text{trace}\{\mathbf{\Lambda}\} \quad (109)$$

with $\mathbf{\Lambda}$ being the matrix of the eigenvalues of $\mathbf{A}$. For the previous example, we have

$$\mathbf{\Lambda} = \text{diag}\{16.1168, -1.1168, 0\}. \quad (110)$$

We can verify that

$$\text{trace}\{\mathbf{A}\} = \sum_{\ell} \lambda_{\ell} = 16.1168 + (-1.1168) = 15. \quad (111)$$

Determinant

The *determinant* is important for finding the solution of systems of linear equations (i.e., the determinant *determines* the existence of a solution). The determinant of a matrix is equal to the product of its eigenvalues. If $\det\{\mathbf{A}\}$ is the determinant of $\mathbf{A}$,

$$\det\{\mathbf{A}\} = \prod_{\ell} \lambda_{\ell} \text{ with } \lambda_{\ell} \text{ being the } \ell\text{th eigenvalue of } \mathbf{A}. \quad (112)$$

For example, the determinant of $\mathbf{A}$ from Equation 107 is equal to

$$\det\{\mathbf{A}\} = 16.1168 \times -1.1168 \times 0 = 0. \quad (113)$$

Rank

Finally, the *rank* of a matrix is the number of nonzero eigenvalues of the matrix. For our example,

$$\text{rank}\{\mathbf{A}\} = 2. \quad (114)$$

The rank of a matrix gives the dimensionality of the Euclidean space that can be used to represent this matrix. Matrices whose rank is equal to their dimensions are *full rank*, and they are invertible. When the rank of a matrix is smaller than its dimensions, the matrix is not invertible and is called *rank-deficient, singular*, or *multicolinear*. For example, matrix $\mathbf{A}$ from Equation 107 is a 3×3 square matrix, its rank is equal to 2, and therefore it is rank-deficient and does not have an inverse.

Statistical Properties of the Eigen-decomposition

The eigen-decomposition is essential in optimization and in statistics. For example, principal components analysis is a technique used to analyze an $I \times J$ matrix $\mathbf{X}$ in which the rows are observations and the columns are variables. Principal components analysis finds orthogonal row *factor scores* that ‘‘explain’’ as much of the variance of $\mathbf{X}$ as possible. They are obtained as

$$\mathbf{F} = \mathbf{XQ} \tag{115}$$

where $\mathbf{F}$ is the matrix of factor scores and $\mathbf{Q}$ is the matrix of loadings of the variables. These loadings give the coefficients of the linear combination used to compute the factor scores from the variables. In addition to Equation 115, we impose the constraints that

$$\mathbf{F}^T\mathbf{F} = \mathbf{Q}^T\mathbf{X}^T\mathbf{XQ} \tag{116}$$

is a diagonal matrix (i.e., $\mathbf{F}$ is an orthogonal matrix) and that

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I} \tag{117}$$

(i.e., $\mathbf{Q}$ is an orthonormal matrix). The solution is obtained by using Lagrange multipliers in which the constraint from Equation 117 is expressed as the multiplication with a diagonal matrix of Lagrange multipliers denoted $\mathbf{\Lambda}$; in order to give the following expression:

$$\mathbf{\Lambda}\left(\mathbf{Q}^T\mathbf{Q} - \mathbf{I}\right) \tag{118}$$

This amounts to defining the following equation:

$$\mathcal{L} = \mathbf{F}^T\mathbf{F} - \mathbf{\Lambda}\left(\mathbf{Q}^T\mathbf{Q} - \mathbf{I}\right). \tag{119}$$

The values of $\mathbf{Q}$ that give the maximum values of $\mathcal{L}$ are found by first computing the derivative of $\mathcal{L}$ relative to $\mathbf{Q}$,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{Q}} = 2\mathbf{X}^T\mathbf{XQ} - 2\mathbf{\Lambda Q}, \tag{120}$$

and setting this derivative to zero:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{Q}} = \mathbf{X}^T\mathbf{XQ} - \mathbf{\Lambda Q} = \mathbf{0} \iff \mathbf{X}^T\mathbf{XQ} = \mathbf{\Lambda Q}. \tag{121}$$

Because $\mathbf{\Lambda}$ is diagonal, this is an eigen-decomposition problem, $\mathbf{\Lambda}$ is the matrix of eigenvalues of the positive semidefinite matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ ordered from the largest to the smallest, and $\mathbf{Q}$ is the matrix of eigenvectors of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. Finally, the factor matrix is

$$\mathbf{F} = \mathbf{XQ}. \tag{122}$$

The variance of the factor scores is equal to the eigenvalues:

$$\mathbf{F}^T\mathbf{F} = \mathbf{Q}^T\mathbf{X}^T\mathbf{XQ} = \mathbf{Q}^T\mathbf{Q\Lambda Q}^T\mathbf{Q} = \mathbf{\Lambda}. \tag{123}$$

Because the sum of the eigenvalues is equal to the trace of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, the set of the first factor scores "extracts" as much of the variance of the original data as possible, the second set of factor scores extracts as much of the variance left unexplained by the first factor as possible, and so on for the remaining factors. The diagonal elements of the matrix $\mathbf{\Lambda}^{\frac{1}{2}}$, which are the standard deviations of the factor scores, are called the *singular values* of **X**.

A Tool for Rectangular Matrices: The Singular Value Decomposition

The singular value decomposition (SVD) generalizes the eigen-decomposition to rectangular matrices. The eigen-decomposition decomposes a matrix into *two* simple matrices, and the SVD decomposes a rectangular matrix into *three* simple matrices: two orthogonal matrices and one diagonal matrix. The SVD uses the eigen-decomposition of a positive semidefinite matrix to derive a similar decomposition for rectangular matrices.

Definitions and Notations

The SVD decomposes matrix **A** as

$$\mathbf{A} = \mathbf{P}\mathbf{\Delta}\mathbf{Q}^{\mathrm{T}}, \qquad (124)$$

where **P** is the matrix storing the (normalized) eigenvectors of the matrix $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ (i.e., $\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I}$). The columns of **P** are called the left singular vectors of **A**. Matrix **Q** stores the (normalized) eigenvectors of the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ (i.e., $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}$). The columns of **Q** are called the right singular vectors of **A**. The matrix $\mathbf{\Delta}$ is the diagonal matrix of the *singular values*, $\mathbf{\Delta} = \mathbf{\Lambda}^{\frac{1}{2}}$, with $\mathbf{\Lambda}$ being the diagonal matrix of the eigenvalues of $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$.

The SVD is derived from the eigen-decomposition of a positive semidefinite matrix. This is shown by considering the eigen-decomposition of the two positive semidefinite matrices obtained from **A**, namely, $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$. If we express these matrices in terms of the SVD of **A**, we find

$$\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{P}\mathbf{\Delta}\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\mathbf{\Delta}\mathbf{P}^{\mathrm{T}} = \mathbf{P}\mathbf{\Delta}^{2}\mathbf{P}^{\mathrm{T}} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{\mathrm{T}} \qquad (125)$$

and

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{Q}\mathbf{\Delta}\mathbf{P}^{\mathrm{T}}\mathbf{P}\mathbf{\Delta}\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}\mathbf{\Delta}^{2}\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathrm{T}}. \qquad (126)$$

This equation shows that Δ is the square root of $\mathbf{\Lambda}$, that $\mathbf{P}$ is the matrix of the eigenvectors of $\mathbf{AA}^{\mathrm{T}}$, and that $\mathbf{Q}$ is the matrix of the eigenvectors of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$. For example, the matrix

$$\mathbf{A} = \begin{bmatrix} 1.1547 & -1.1547 \\ -1.0774 & 0.0774 \\ -0.0774 & 1.0774 \end{bmatrix} \qquad (127)$$

can be expressed as

$$\mathbf{A} = \mathbf{PQ}^{\mathrm{T}} = \begin{bmatrix} 0.8165 & 0 \\ -.4082 & -0.7071 \\ -0.4082 & -0.7071 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & -0.7071 \end{bmatrix} = \begin{bmatrix} 1.1547 & -1.1547 \\ -0.0774 & 0.0774 \\ -0.0774 & 1.0774 \end{bmatrix}. \qquad (128)$$

We can check that

$$\mathbf{AA}^{\mathrm{T}} = \begin{bmatrix} 0.8165 & 0 \\ -0.4082 & -0.7071 \\ -0.4082 & -0.7071 \end{bmatrix} \begin{bmatrix} 2^2 & 0 \\ 0 & 1^2 \end{bmatrix} \begin{bmatrix} 0.8165 & -0.4082 & -0.4082 \\ 0 & -0.7071 & 0.7071 \end{bmatrix} = \begin{bmatrix} 2.6667 & 1.3333 & 1.3333 \\ 1.3333 & 1.1667 & 0.1667 \\ 1.3333 & 0.1667 & 1.1667 \end{bmatrix}$$

and that

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \begin{bmatrix} 2^2 & 0 \\ 0 & 1^2 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & 0.7071 \end{bmatrix} = \begin{bmatrix} 2.5 & -1.5 \\ -1.5 & 2.5 \end{bmatrix}. \qquad (130)$$

Generalized or Pseudoinverse

The inverse of a matrix is defined only for full rank square matrices. The generalization of the inverse for other matrices is called *generalized inverse, pseudoinverse,* or *Moore–Penrose inverse* and is denoted by $\mathbf{X}^{+}$. The pseudoinverse of $\mathbf{A}$ is the unique matrix that satisfies the following four properties:

$$\begin{aligned} &\mathbf{AA}^{+}\mathbf{A} = \mathbf{A} && (i) \\ &\mathbf{A}^{+}\mathbf{AA}^{+} = \mathbf{A}^{+} && (ii) \\ &(\mathbf{AA}^{+})^{\mathrm{T}} = \mathbf{AA}^{+} \text{(symmetry 1)} && (iii) \\ &(\mathbf{A}^{+}\mathbf{A})^{\mathrm{T}} = \mathbf{A}^{+}\mathbf{A} \quad \text{(symmetry 2)} && \text{(iv).} \quad (131) \end{aligned}$$

For example, with

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \qquad (132)$$

we find that the pseudoinverse $\mathbf{A}^{+}$ is equal to

$$\mathbf{A}^{+} = \begin{bmatrix} .25 & -.25 & .5 \\ -.25 & .25 & .5 \end{bmatrix}. \qquad (133)$$

This example shows that the product of a matrix and its pseudoinverse does not always give the identity matrix:

$$\mathbf{A}\mathbf{A}^{+} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} .25 & -.25 & .5 \\ -.25 & .25 & .5 \end{bmatrix} = \begin{bmatrix} 0.3750 & 0.1250 \\ 0.1250 & 0.3750 \end{bmatrix} \qquad (134)$$

Pseudoinverse and SVD

The SVD is the building block for the Moore–Penrose pseudoinverse because any matrix **A** with SVD equal to $\mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{\mathsf{T}}$ has for pseudoinverse

$$\mathbf{A}^{+} = \mathbf{Q}\boldsymbol{\Delta}^{-1}\mathbf{P}^{\mathsf{T}}. \qquad (135)$$

For the preceding example, we obtain

$$\mathbf{A}^{+} = \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \begin{bmatrix} 2^{-1} & 0 \\ 0 & 1^{-1} \end{bmatrix} \begin{bmatrix} 0.8165 & -0.4082 & -0.4082 \\ 0 & -0.7071 & 0.7071 \end{bmatrix} = \begin{bmatrix} 0.2887 & -0.6443 & 0.3557 \\ -0.2887 & -0.3557 & 0.6443 \end{bmatrix}. \qquad (136)$$

Hervé Abdi

See also Analysis of Covariance (ANCOVA); Analysis of Variance (ANOVA); Canonical Correlation Analysis; Confirmatory Factor Analysis; Correspondence Analysis; General Linear Model; Latent Variable; Mauchly Test; Multiple Regression; Principal Components Analysis; Sphericity; Structural Equation Modeling

Further Readings

Abdi, H., & Williams, L.J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 433–459.

Abdi H., & Beaton, D. (2022). *Principal component and correspondence analyses using R.* New York: Springer.

Guillemot, V., Beaton, D., Gloaguen, A., Lofstedt, T., Levine, B., Raymond, N., Tenenhaus, A. & Abdi, H. (2019). A constrained singular value decomposition method that integrates sparsity and orthogonality. *PLoSOne*, **14(1)**, 1–39. https://doi.org/10.1371/journal.pone.0211463

Deisenroth, M.P., Alod Faisal, A., & Soon Ong, C., (2020). *Mathematics for machine learning*. Cambridge: Cambridge University Press.

Rao, C.R., Mitra, S.K. (1971). *Generalized inverse of matrices and its applications.* New York: Wiley.

Searle, S. R. & Khuri, A.I. (2017). *Matrices algebra useful for statistics.* New York: Wiley.