In Neil Salkind (Ed.), ***Encyclopedia of Research Design.***
Thousand Oaks, CA: Sage. 2010

Matrix Algebra

Hervé Abdi · Lynne J. Williams

1 Introduction

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

2 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,

Hervé Abdi
The University of Texas at Dallas

Lynne J. Williams
The University of Toronto Scarborough

Address correspondence to:
Hervé Abdi
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75083–0688, USA
E-mail: herve@utdallas.edu http://www.utd.edu/~herve

and 4 and Olivette has 6, 1, 3 and 10. These data can be displayed in a table where each row represents a person and each column a possession:

	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 $\mathbf{A}$ equal to:

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

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 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 the i-th row and j-th column of $\mathbf{A}$. The *total* number of rows and columns is denoted with the same letters as the indices but in uppercase letters. The matrix $\mathbf{A}$ 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 $3 \times 4 = 12$). We often use the term *dimensions* to refer to the number of rows and columns, so $\mathbf{A}$ has dimensions I by J.

As a shortcut, a matrix can be represented by its generic element written in brackets. So, $\mathbf{A}$ with I rows and J columns is denoted:

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

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

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

2.1 Vectors

A matrix with one column is called a *column vector* or simply a vector. Vectors are denoted with bold lower case letters. For example, the first column of matrix $\mathbf{A}$ (of Equation 1) is a column vector which stores the number of balls of Toto, Marius, and Olivette. We can call it $\mathbf{b}$ (for balls), and so:

$$\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix} . \tag{4}$$

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

2.2 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

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} , \tag{5}$$

we find

$$\| \mathbf{x} \| = \sqrt{2^2 + 1^2 + 2^2} = \sqrt{4 + 1 + 4} = \sqrt{9} = 3 \ . \tag{6}$$

2.3 Normalization of a vector

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

$$\overline{\mathbf{x}} = \frac{\mathbf{x}}{\| \mathbf{x} \|} = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ \frac{2}{3} \end{bmatrix} . \tag{7}$$

3 Operations for matrices

3.1 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 $\mathbf{A}$ transposed is denoted $\mathbf{A}^T$. For example:

$$\text{if } \mathbf{A} = \underset{3 \times 4}{\mathbf{A}} = \begin{bmatrix} 2 & 5 & 10 & 20 \\ 1 & 2 & 3 & 4 \\ 6 & 1 & 3 & 10 \end{bmatrix} \text{ then } \mathbf{A}^T = \underset{4 \times 3}{\mathbf{A}^T} = \begin{bmatrix} 2 & 1 & 6 \\ 5 & 2 & 1 \\ 10 & 3 & 3 \\ 20 & 4 & 10 \end{bmatrix} . \tag{8}$$

3.2 Addition (sum) of matrices

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

$$\mathbf{A} = \begin{bmatrix} 2 & 5 & 10 & 20 \\ 1 & 2 & 3 & 4 \\ 6 & 1 & 3 & 10 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 3 & 4 & 5 & 6 \\ 2 & 4 & 6 & 8 \\ 1 & 2 & 3 & 5 \end{bmatrix}, \tag{9}$$

we find

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 2+3 & 5+4 & 10+5 & 20+6 \\ 1+2 & 2+4 & 3+6 & 4+8 \\ 6+1 & 1+2 & 3+3 & 10+5 \end{bmatrix} = \begin{bmatrix} 5 & 9 & 15 & 26 \\ 3 & 6 & 9 & 12 \\ 7 & 3 & 6 & 15 \end{bmatrix}. \tag{10}$$

In general

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

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

3.3 Multiplication of a matrix by a scalar

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

$$10 \times \mathbf{B} = 10 \times \begin{bmatrix} 3 & 4 & 5 & 6 \\ 2 & 4 & 6 & 8 \\ 1 & 2 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 30 & 40 & 50 & 60 \\ 20 & 40 & 60 & 80 \\ 10 & 20 & 30 & 50 \end{bmatrix}. \tag{12}$$

3.4 Multiplication: Product or products?

There are *several* ways of generalizing the concept of product 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.

3.5 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:

$$\mathbf{A} \odot \mathbf{B} = [a_{i,j} \times b_{i,j}]$$
$$= \begin{bmatrix} a_{1,1} \times b_{1,1} & a_{1,2} \times b_{1,2} & \cdots & a_{1,j} \times b_{1,j} & \cdots & a_{1,J} \times b_{1,J} \\ a_{2,1} \times b_{2,1} & a_{2,2} \times b_{2,2} & \cdots & a_{2,j} \times b_{2,j} & \cdots & a_{2,J} \times b_{2,J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,1} \times b_{i,1} & a_{i,2} \times b_{i,2} & \cdots & a_{i,j} \times b_{i,j} & \cdots & a_{i,J} \times b_{i,J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{I,1} \times b_{I,1} & a_{I,2} \times b_{I,2} & \cdots & a_{I,j} \times b_{I,j} & \cdots & a_{I,J} \times b_{I,J} \end{bmatrix} . \tag{13}$$

For example, with

$$\mathbf{A} = \begin{bmatrix} 2 & 5 & 10 & 20 \\ 1 & 2 & 3 & 4 \\ 6 & 1 & 3 & 10 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 3 & 4 & 5 & 6 \\ 2 & 4 & 6 & 8 \\ 1 & 2 & 3 & 5 \end{bmatrix} , \tag{14}$$

we get:

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

3.6 Standard (a.k.a.) Cayley product

The Hadamard product is straightforward, but, unfortunately, it is *not* the matrix product most often used. This product is called the *standard* or *Cayley* product, or simply *the* product (*i.e.,* when the name of the product is not specified, this is the standard product). Its definition 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, $\mathbf{A}$ with I rows and J columns can be multiplied by $\mathbf{B}$ with J rows and K columns to give $\mathbf{C}$ 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:

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

An element $c_{i,k}$ of the matrix $\mathbf{C}$ is 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 i-th row of $\mathbf{A}$ with the k-th column of $\mathbf{B}$.

For example, let:

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

The product of these matrices is denoted $\mathbf{C} = \mathbf{A} \times \mathbf{B} = \mathbf{AB}$ (the $\times$ sign can be omitted when the context is clear). To compute $c_{2,1}$ we add 3 terms: (1) the product of the first element of the second row of $\mathbf{A}$ (*i.e.,* 4) with the first element of the first column of $\mathbf{B}$ (*i.e.,* 1); (2) the product of the second element of the second row of $\mathbf{A}$ (*i.e.,* 5) with the second element of the first column of $\mathbf{B}$ (*i.e.,* 3); and (3) the product of the third element of the second row of $\mathbf{A}$ (*i.e.,* 6) with the third element of the first column of $\mathbf{B}$ (*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 $\mathbf{C}$ is obtained as:

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

3.6.1 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 $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$:

$$
\begin{aligned}
(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) = \mathbf{ABC} \quad & \text{associativity} \qquad (22) \\
\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \quad & \text{distributivity.} \qquad (23)
\end{aligned}
$$

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

$$\mathbf{AB} \neq \mathbf{BA} \ . \tag{24}$$

For example, with

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

we get:

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

But

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

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

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

3.7 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 $\mathbf{B}$ (with dimensions K and L), the Kronecker product gives a matrix $\mathbf{C}$ (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:

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

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

4 Special matrices

Certain special matrices have specific names.

4.1 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:

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

is a square matrix, but

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

is a rectangular matrix.

4.2 Symmetric matrix

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

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

is symmetric, but

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

is not.

Note that for a symmetric matrix:

$$\mathbf{A} = \mathbf{A}^T \ . \tag{36}$$

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:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 1 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 3 \\ 2 & 3 & 1 \end{bmatrix} \ . \tag{37}$$

We get

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

Note, however, that combining Equations 28 and 36, gives for symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, the following equation:

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

4.3 Diagonal matrix

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

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

Because only the diagonal elements matter for a diagonal matrix, we just need to specify them. This is done with the following notation:

$$\mathbf{A} = \mathsf{diag}\left\{[a_{1,1}, \ldots, a_{i,i}, \ldots, a_{I,I}]\right\} = \mathsf{diag}\left\{[a_{i,i}]\right\} . \tag{41}$$

For example, the previous matrix can be rewritten as:

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

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:

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

Note, incidently, that:

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

4.4 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 pre-multiply 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 post-multiply 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:

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

we get

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

and

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

and also

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

4.5 Identity matrix

A diagonal matrix whose diagonal elements are all equal to 1 is called an *identity* matrix and is denoted $\mathbf{I}$. If we need to specify its dimensions, we use subscripts such as

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

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

$$\mathbf{I} \times \mathbf{A} = \mathbf{A} \times \mathbf{I} = \mathbf{A} \tag{51}$$

for any matrix $\mathbf{A}$ conformable with $\mathbf{I}$. For example:

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

4.6 Matrix full of ones

A matrix whose elements are all equal to 1, is denoted $\mathbf{1}$ or, when we need to specify its dimensions, by $\underset{I \times J}{\mathbf{1}}$. These matrices are neutral elements for the Hadamard product. So:

$$\underset{2 \times 3}{\mathbf{A}} \odot \underset{2 \times 3}{\mathbf{1}} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \odot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{53}$$

$$= \begin{bmatrix} 1 \times 1 & 2 \times 1 & 3 \times 1 \\ 4 \times 1 & 5 \times 1 & 6 \times 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} . \tag{54}$$

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

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = (1 \times 1) + (2 \times 1) + (3 \times 1) = 1 + 2 + 3 = 6 , \tag{55}$$

or also

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

4.7 Matrix full of zeros

A matrix whose elements are all equal to 0, is the *null* or *zero* matrix. It is denoted by $\mathbf{0}$ or, when we need to specify its dimensions, by

$\underset{I \times J}{\mathbf{0}}$. Null matrices are neutral elements for addition

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \underset{2 \times 2}{\mathbf{0}} = \begin{bmatrix} 1+0 & 2+0 \\ 3+0 & 4+0 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} . \tag{57}$$

They are also null elements for the Hadamard product.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \odot \underset{2 \times 2}{\mathbf{0}} = \begin{bmatrix} 1 \times 0 & 2 \times 0 \\ 3 \times 0 & 4 \times 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \underset{2 \times 2}{\mathbf{0}} \tag{58}$$

and for the standard product:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \underset{2 \times 2}{\mathbf{0}} = \begin{bmatrix} 1 \times 0 + 2 \times 0 & 1 \times 0 + 2 \times 0 \\ 3 \times 0 + 4 \times 0 & 3 \times 0 + 4 \times 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \underset{2 \times 2}{\mathbf{0}} . \tag{59}$$

4.8 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:

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

and

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

4.9 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:

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

pre-multiplied by its transpose

$$\mathbf{A}^T = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 4 \end{bmatrix} \tag{63}$$

gives the cross-product matrix:

$$\begin{aligned}\mathbf{A}^T\mathbf{A} &= \begin{bmatrix} 1 \times 1 + 2 \times 2 + 3 \times 3 & 1 \times 1 + 2 \times 4 + 3 \times 4 \\ 1 \times 1 + 4 \times 2 + 4 \times 3 & 1 \times 1 + 4 \times 4 + 4 \times 4 \end{bmatrix} \\ &= \begin{bmatrix} 14 & 21 \\ 21 & 33 \end{bmatrix} .\end{aligned} \tag{64}$$

4.9.1 A particular case of cross-product matrix: Variance/Covariance

A particular case of cross-product matrices are correlation or covariance matrices. A variance/covariance matrix is obtained from a data matrix with 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 $\mathbf{A}$:

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

we obtain the means of each column as:

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

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

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

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

We note $\mathbf{S}$ the variance/covariance matrix derived from $\mathbf{A}$, it is computed as:

$$\mathbf{S} = \frac{1}{I}\mathbf{D}^T\mathbf{D} = \frac{1}{3}\begin{bmatrix} -3 & 0 & 3 \\ -6 & 3 & 3 \end{bmatrix} \times \begin{bmatrix} -3 & -6 \\ 0 & 3 \\ 3 & 3 \end{bmatrix}$$

$$= \frac{1}{3} \times \begin{bmatrix} 18 & 27 \\ 27 & 54 \end{bmatrix} = \begin{bmatrix} 6 & 9 \\ 9 & 18 \end{bmatrix} . \tag{69}$$

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

5 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:

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

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

$$\mathbf{A} \times \mathbf{A}^{-1} = \mathbf{A}^{-1} \times \mathbf{A} = \mathbf{I} . \tag{71}$$

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

For example, for:

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

the inverse is:

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

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,

$$\mathbf{A} = \begin{bmatrix} 3 & 4 & 2 \\ 1 & 0 & 2 \\ 2 & 1 & 3 \end{bmatrix} , \tag{74}$$

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

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

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

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

5.1 Inverse of a diagonal matrix

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

$$\mathbf{A} = \mathsf{diag}\left\{a_{i,i}\right\} \tag{76}$$

is the diagonal matrix

$$\mathbf{A}^{-1} = \mathsf{diag}\left\{a_{i,i}^{-1}\right\} = \mathsf{diag}\left\{1/a_{i,i}\right\} \tag{77}$$

For example,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & 4 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & .25 \end{bmatrix}, \tag{78}$$

are the inverse of each other.

6 The Big tool: eigendecomposition

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 *eigendecomposition* and it applies only to square matrices, the generalization of the eigendecomposition to rectangular matrices is called the *singular value* decomposition.

Eigenvectors and *eigenvalues* are numbers and vectors associated with square matrices, together they constitute the *eigendecomposition*. Even though the eigendecomposition 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 eigendecomposition 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 component analysis is obtained from the eigendecomposition of a covariance or correlation matrix and gives the least square estimate of the original data matrix.

6.1 Notations and definition

An eigenvector of matrix $\mathbf{A}$ is a vector $\mathbf{u}$ that satisfies the following equation:

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u} \ , \tag{79}$$

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

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

Therefore $\mathbf{u}$ is eigenvector of $\mathbf{A}$ if the multiplication of $\mathbf{u}$ by $\mathbf{A}$ changes the length of $\mathbf{u}$ but not its orientation. For example,

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \tag{81}$$

has for eigenvectors:

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

and

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

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

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

and

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

This is illustrated in Figure 1.

For convenience, eigenvectors are generally normalized such that:

$$\mathbf{u}^{\mathsf{T}}\mathbf{u} = 1 \ . \tag{86}$$

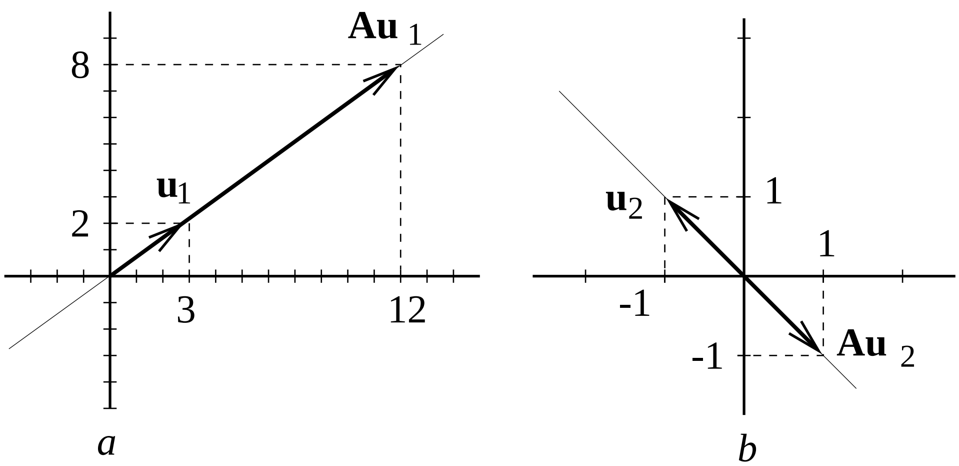

Figure 1: Two eigenvectors of a matrix.

For the previous example, normalizing the eigenvectors gives:

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

We can check that:

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

and

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

6.2 Eigenvector and eigenvalue matrices

Traditionally, we store the eigenvectors of $\mathbf{A}$ as the columns a matrix denoted $\mathbf{U}$. Eigenvalues are stored in a diagonal matrix (denoted $\mathbf{\Lambda}$). Therefore, Equation 79 becomes:

$$\mathbf{AU} = \mathbf{U\Lambda} . \tag{90}$$

For example, with $\mathbf{A}$ (from Equation 81), we have

$$\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix} \tag{91}$$

6.3 Reconstitution of a matrix

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

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1} \ . \tag{92}$$

For example, because

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

we obtain:

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

6.4 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 $\mathbf{A}$:

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

Therefore:

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

is also an eigenvector of $\mathbf{A}$:

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

6.5 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 semi-definite*. The eigendecomposition of these matrices always exists and has a particularly convenient form. A matrix is positive semi-definite when it can be obtained as the product of a matrix by its transpose. This implies that a positive semi-definite matrix is always symmetric. So, formally, the matrix $\mathbf{A}$ is positive semi-definite if it can be obtained as:

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

for a certain matrix $\mathbf{X}$. Positive semi-definite matrices include correlation, covariance, and cross-product matrices.

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

$$\mathbf{U}^{-1} = \mathbf{U}^{\mathsf{T}} . \tag{99}$$

We can, therefore, express the positive semi-definite matrix $\mathbf{A}$ as:

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

where $\mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{I}$ are the normalized eigenvectors.

For example,

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

can be decomposed as:

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

with

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

6.5.1 Diagonalization

When a matrix is positive semi-definite we can rewrite Equation 100 as

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathsf{T}} \Longleftrightarrow \boldsymbol{\Lambda} = \mathbf{U}^{\mathsf{T}}\mathbf{A}\mathbf{U} . \tag{104}$$

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

6.5.2 Another definition for positive semi-definite matrices

A matrix $\mathbf{A}$ is positive semi-definite if for any non-zero vector $\mathbf{x}$ we have:

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

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

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

6.6 Trace, Determinant, etc.

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

6.6.1 Trace

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

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

we obtain:

$$\mathsf{trace}\{\mathbf{A}\} = 1 + 5 + 9 = 15 \ . \tag{108}$$

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

$$\mathsf{trace}\{\mathbf{A}\} = \sum_{\ell} \lambda_\ell = \mathsf{trace}\{\mathbf{\Lambda}\} \tag{109}$$

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

$$\mathbf{\Lambda} = \mathsf{diag}\{16.1168, -1.1168, 0\} \ . \tag{110}$$

We can verify that:

$$\mathsf{trace}\{\mathbf{A}\} = \sum_{\ell} \lambda_\ell = 16.1168 + (-1.1168) = 15 \tag{111}$$

6.6.2 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 $\mathsf{det}\{\mathbf{A}\}$ is the determinant of $\mathbf{A}$:

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

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

$$\mathsf{det}\{\mathbf{A}\} = 16.1168 \times -1.1168 \times 0 = 0 \ . \tag{113}$$

6.6.3 Rank

Finally, the *rank* of a matrix is the number of non-zero eigenvalues of the matrix. For our example:

$$\mathsf{rank}\{\mathbf{A}\} = 2 \ . \tag{114}$$

The rank of a matrix gives the dimensionality of the Euclidean space which 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.

6.7 Statistical properties of the eigen-decomposition

The eigen-decomposition is essential in optimization. For example, principal component analysis (PCA) is a technique used to analyze a $I \times J$ matrix $\mathbf{X}$ where the rows are observations and the columns are variables. PCA finds orthogonal row *factor scores* which "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

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

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

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

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

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

This amounts to defining the following equation

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

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

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

and setting this derivative to zero:

$$\mathbf{X}^{\mathsf{T}}\mathbf{X}\mathbf{Q} - \mathbf{\Lambda}\mathbf{Q} = \mathbf{0} \Longleftrightarrow \mathbf{X}^{\mathsf{T}}\mathbf{X}\mathbf{Q} = \mathbf{\Lambda}\mathbf{Q} . \tag{121}$$

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

$$\mathbf{F} = \mathbf{X}\mathbf{Q} . \tag{122}$$

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

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

Because the sum of the eigenvalues is equal to the trace of $\mathbf{X}^\mathsf{T}\mathbf{X}$, the first factor scores "extract" as much of the variances of the original data as possible, and the second factor scores extract as much of the variance left unexplained by the first factor, 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 $\mathbf{X}$.

7 A tool for rectangular matrices: The singular value decomposition

The singular value decomposition (SVD) generalizes the eigendecomposition to rectangular matrices. The eigendecomposition, 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 eigendecomposition of a positive semi-definite matrix to derive a similar decomposition for rectangular matrices.

7.1 Definitions and notations

The SVD decomposes matrix $\mathbf{A}$ as:

$$\mathbf{A} = \mathbf{P}\mathbf{\Delta}\mathbf{Q}^\mathsf{T} . \tag{124}$$

where $\mathbf{P}$ is the (normalized) eigenvectors of the matrix $\mathbf{A}\mathbf{A}^\mathsf{T}$ (*i.e.,* $\mathbf{P}^\mathsf{T}\mathbf{P} = \mathbf{I}$). The columns of $\mathbf{P}$ are called the *left singular vectors* of $\mathbf{A}$. $\mathbf{Q}$ is the (normalized) eigenvectors of the matrix $\mathbf{A}^\mathsf{T}\mathbf{A}$ (*i.e.,* $\mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{I}$). The columns of $\mathbf{Q}$ are called the *right singular vectors* of $\mathbf{A}$. $\mathbf{\Delta}$ 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 $\mathbf{A}\mathbf{A}^\mathsf{T}$ and $\mathbf{A}^\mathsf{T}\mathbf{A}$.

The SVD is derived from the eigendecomposition of a positive semi-definite matrix. This is shown by considering the eigendecomposition of the two positive semi-definite matrices obtained from $\mathbf{A}$:

namely $\mathbf{A}\mathbf{A}^{\mathsf{T}}$ and $\mathbf{A}^{\mathsf{T}}\mathbf{A}$. If we express these matrices in terms of the SVD of $\mathbf{A}$, we find:

$$\mathbf{A}\mathbf{A}^{\mathsf{T}} = \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{\mathsf{T}}\mathbf{Q}\boldsymbol{\Delta}\mathbf{P}^{\mathsf{T}} = \mathbf{P}\boldsymbol{\Delta}^2\mathbf{P}^{\mathsf{T}} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{\mathsf{T}} \ , \tag{125}$$

and

$$\mathbf{A}^{\mathsf{T}}\mathbf{A} = \mathbf{Q}\boldsymbol{\Delta}\mathbf{P}^{\mathsf{T}}\mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{\mathsf{T}} = \mathbf{Q}\boldsymbol{\Delta}^2\mathbf{Q}^{\mathsf{T}} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathsf{T}} \ . \tag{126}$$

This shows that $\boldsymbol{\Delta}$ is the square root of $\boldsymbol{\Lambda}$, that $\mathbf{P}$ are eigenvectors of $\mathbf{A}\mathbf{A}^{\mathsf{T}}$, and that $\mathbf{Q}$ are eigenvectors of $\mathbf{A}^{\mathsf{T}}\mathbf{A}$.

For example, the matrix:

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

can be expressed as:

$$\begin{aligned} \mathbf{A} &= \mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{\mathsf{T}} \\ &= \begin{bmatrix} 0.8165 & 0 \\ -0.4082 & -0.7071 \\ -0.4082 & 0.7071 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \\ &= \begin{bmatrix} 1.1547 & -1.1547 \\ -1.0774 & 0.0774 \\ -0.0774 & 1.0774 \end{bmatrix} \ . \end{aligned} \tag{128}$$

We can check that:

$$\mathbf{A}\mathbf{A}^{\mathsf{T}} = \begin{bmatrix} 0.8165 & 0 \\ -0.4082 & -0.7071 \\ -0.4082 & 0.7071 \end{bmatrix} \begin{bmatrix} 2^2 & 0 \\ 0 & 1^2 \end{bmatrix} \begin{bmatrix} 0.8165 & -0.4082 & -0.4082 \\ 0 & -0.7071 & 0.7071 \end{bmatrix}$$

$$= \begin{bmatrix} 2.6667 & -1.3333 & -1.3333 \\ -1.3333 & 1.1667 & 0.1667 \\ -1.3333 & 0.1667 & 1.1667 \end{bmatrix} \tag{129}$$

and that:

$$\mathbf{A}^{\mathsf{T}}\mathbf{A} = \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \begin{bmatrix} 2^2 & 0 \\ 0 & 1^2 \end{bmatrix} \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix}$$

$$= \begin{bmatrix} 2.5 & -1.5 \\ -1.5 & 2.5 \end{bmatrix} . \tag{130}$$

7.2 Generalized or pseudo-inverse

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*, *pseudo-inverse* or *Moore-Penrose inverse* and is denoted by $\mathbf{X}^+$. The pseudo-inverse of $\mathbf{A}$ is the unique matrix that satisfies the following four constraints:

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

For example, with

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \tag{132}$$

we find that the pseudo-inverse is equal to

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

This example shows that the product of a matrix and its pseudo-inverse does not always gives the identity matrix:

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

7.3 Pseudo-inverse and singular value decomposition

The SVD is the building block for the Moore-Penrose pseudo-inverse. Because any matrix $\mathbf{A}$ with SVD equal to $\mathbf{P}\boldsymbol{\Delta}\mathbf{Q}^{\mathsf{T}}$ has for pseudo-inverse:

$$\mathbf{A}^{+} = \mathbf{Q}\boldsymbol{\Delta}^{-1}\mathbf{P}^{T} . \tag{135}$$

For the preceding example we obtain:

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

Pseudo-inverse matrices are used to solve multiple regression and analysis of variance problems.

Related entries

Analysis of variance and covariance, canonical correlation, correspondence analysis, confirmatory factor analysis, discriminant analysis, general linear, latent variable, Mauchly test, multiple regression, principal component analysis, sphericity, structural equation modelling

Further readings

1. Abdi, H. (2007a). Eigendecomposition: eigenvalues and eigenvecteurs. In N.J. Salkind (Ed.): *Encyclopedia of Measurement and Statistics*. Thousand Oaks (CA): Sage. pp. 304–308.
2. Abdi, H. (2007b). Singular Value Decomposition (SVD) and Generalized Singular Value Decomposition (GSVD). In N.J. Salkind (Ed.): *Encyclopedia of Measurement and Statistics*. Thousand Oaks (CA): Sage. pp. 907–912.
3. Basilevsky, A. (1983). *Applied Matrix Algebra in the Statistical Sciences*. New York: North-Holland.
4. Graybill, F.A. (1969). *Matrices with Applications in Statistics*. New York: Wadworth.
5. Healy, M.J.R. (1986). *Matrices for Statistics*. Oxford: Oxford University Press.
6. Searle, S.R. (1982). *Matrices Algebra Useful for Statistics*. New York: Wiley.