A Brief Introduction to Matrix Algebra

Introducing Matrices

A *matrix* is a rectangular or square array of values arranged in rows and columns. An $m \times n$ matrix **A** has m rows and n columns, and has a general form of

$$\mathbf{A}_{m\times n}=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}_{m\times n}$$

The element a_{ij} denotes the element in the i^{th} row and the j^{th} column in matrix **A**.

Example:

$\mathbf{A} = \begin{bmatrix} 3 & 5 & 8 \\ 1 & 2 & 2 \end{bmatrix}$ is a 2×3 matrix; $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 3 & 6 \\ 7 & 2 \\ 2 & 9 \end{bmatrix}$ is a 4×2 matrix.

A *vector* is a matrix with only one column (a *column vector*) or only one row (a *row vector*). For example,

$\mathbf{x}=\begin{bmatrix} 3 \\ -2 \\ 1 \\ 5 \end{bmatrix}$ is a column vector,

and $\mathbf{x}'=\begin{bmatrix} 3 & -2 & 1 & 5 \end{bmatrix}$ and $\mathbf{y}'=\begin{bmatrix} 4 & 7 & -5 \end{bmatrix}$ are row vectors.

A single number such as 2.4 or −6 is called a *scalar*. The elements of a matrix are usually scalars, although a matrix can be expressed as a matrix of smaller matrices.

Basic Operations

Transpose

The *transpose* of a matrix **A** is the matrix whose columns are rows of **A** (and therefore whose rows are columns of **A**), with order retained, from first to last. The transpose of **A** is denoted by **A′**.

Example:

Let $\mathbf{A} = \begin{bmatrix} 3 & 2 & -6 \\ 7 & 1 & 2 \end{bmatrix}$, then $\mathbf{A}' = \begin{bmatrix} 3 & 7 \\ 2 & 1 \\ -6 & 2 \end{bmatrix}$.

You see that if **A** is 2×3, then **A′** is 3×2. In general if **A** is $m\times n$, then **A′** is $n\times m$, and $a'_{ij} = a_{ji}$.

The transpose of a row vector is a column vector.

Partitioned Matrices

The matrix **A** can be written as a matrix of matrices:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

This specification of **A** is called a *partitioning* of **A**, and the matrices $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21}$, $\mathbf{A}_{22}$ are said to be sub-matrices of **A**. **A** is called a *partitioned matrix*.

Example:

Let $\mathbf{A} = \begin{bmatrix} 1 & 6 & 9 & 4 & 5 & 8 \\ 5 & 4 & 7 & 2 & 0 & 2 \\ 9 & 2 & 8 & 1 & 7 & 1 \\ 9 & 1 & 7 & 6 & 2 & 3 \\ 2 & 5 & 4 & 8 & 1 & 7 \end{bmatrix}$; each of the arrays of numbers in the four sections of **A** engendered by the dashed lines is a matrix:

$$\mathbf{A}_{11} = \begin{bmatrix} 1 & 6 & 9 & 4 \\ 5 & 4 & 7 & 2 \\ 9 & 2 & 8 & 1 \end{bmatrix} \quad \mathbf{A}_{12} = \begin{bmatrix} 5 & 8 \\ 0 & 2 \\ 7 & 1 \end{bmatrix}$$

$$\mathbf{A}_{21} = \begin{bmatrix} 9 & 1 & 7 & 6 \\ 2 & 5 & 4 & 8 \end{bmatrix} \quad \mathbf{A}_{22} = \begin{bmatrix} 2 & 3 \\ 1 & 7 \end{bmatrix}$$

Trace

The sum of the diagonal elements of a square matrix is called the *trace* of the matrix, that is, tr(**A**) = $a_{11} + a_{22} + \dots + a_{nn} = \sum_{i=1}^{n} a_{ii}$.

Example:

Let $\mathbf{A} = \begin{bmatrix} 1 & 5 & 9 \\ -3 & 2 & 8 \\ 4 & 7 & 6 \end{bmatrix}$, then tr(**A**) = 1 + 2 + 6 = 9.

Addition and Subtraction

Matrices of the same size are added or subtracted by adding or subtracting corresponding elements.

Example:

Let $\mathbf{A} = \begin{bmatrix} 2 & 4 & 8 \\ 9 & -1 & 5 \end{bmatrix}$, and $\mathbf{B} = \begin{bmatrix} 1 & 5 & 3 \\ 2 & 3 & -7 \end{bmatrix}$, then

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 2+1 & 4+5 & 8+3 \\ 9+2 & (-1)+3 & 5+(-7) \end{bmatrix} = \begin{bmatrix} 3 & 9 & 11 \\ 11 & 2 & -2 \end{bmatrix} \text{ and}$$

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} 2-1 & 4-5 & 8-3 \\ 9-2 & -1-3 & 5-(-7) \end{bmatrix} = \begin{bmatrix} 1 & -1 & 5 \\ 7 & -4 & 12 \end{bmatrix}.$$

Multiplication

The *inner product* of two vectors $\mathbf{a}' = [a_1 \;\; a_2 \;\; \dots \;\; a_n]$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is defined as

$$\mathbf{a}'\mathbf{x} = a_1x_1 + a_2x_2 + \dots + a_nx_n = \sum_{i=1}^{n} a_i x_i .$$

Example:

Let $\mathbf{a}' = [3 \; 1 \; 10]$ and $\mathbf{x} = \begin{bmatrix} 5 \\ 10 \\ 3 \end{bmatrix}$, then the inner product $\mathbf{a}'\mathbf{x}$ = 3(5) + 1(10) + 10(3) = 55.

The product **AB** of two matrices **A** and **B** is defined and therefore exists only if the number of columns in **A** equals the number of rows in **B**. **AB** has the same number of rows as **A** and the same number of columns as **B**. The ij^{th} element of **AB** is the inner product of the i^{th} row of **A** and j^{th} column of **B**.

Example:

Let **A** = $\begin{bmatrix} 1 & 2 & 3 \\ -1 & 4 & 2 \end{bmatrix}$ and **B** = $\begin{bmatrix} 0 & 6 & 1 & 5 \\ 2 & 1 & -2 & 3 \\ 4 & 1 & 2 & 5 \end{bmatrix}$, then

AB = $\begin{bmatrix} 1(0)+2(2)+3(4) & 1(6)+2(1)+3(1) & 1(1)+2(-2)+3(2) & 1(5)+2(3)+3(5) \\ -1(0)+4(2)+2(4) & -1(6)+4(1)+2(1) & -1(1)+4(-2)+2(2) & -1(5)+4(3)+2(5) \end{bmatrix}$ =
$\begin{bmatrix} 16 & 11 & 3 & 26 \\ 16 & 0 & -5 & 17 \end{bmatrix}$

Direct or Kronecker Product ⊗

Suppose **A** is $m \times n$ and **B** is $p \times q$. Then the *direct* or *Kronecker product* **A**⊗**B** is of size $mp \times nq$ and is most easily described as the partitioned matrix:

$$\mathbf{A}_{m\times n} \otimes \mathbf{B}_{p\times q} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \dots & \dots & \dots & \dots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{bmatrix}$$

Example:

Let **A** = $\begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}$ and **B** = $\begin{bmatrix} 1 & -1 & 1 & 2 \\ 3 & 2 & 0 & 1 \\ -1 & 0 & 2 & 3 \end{bmatrix}$, then

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} 1 & -1 & 1 & 2 & 2 & -2 & 2 & 4 \\ 3 & 2 & 0 & 1 & 6 & 4 & 0 & 2 \\ -1 & 0 & 2 & 3 & -2 & 0 & 4 & 6 \\ -1 & 1 & -1 & -2 & 3 & -3 & 3 & 6 \\ -3 & -2 & 0 & -1 & 9 & 6 & 0 & 3 \\ 1 & 0 & -2 & -3 & -3 & 0 & 6 & 9 \end{bmatrix}$$

Some Special Matrices

A *square* matrix is a matrix whose number of columns equals the number of rows. The elements on the diagonal, a_{11}, a_{22}, …, a_{nn}, are referred to as the *diagonal elements* or *diagonal* of the matrix. Elements of a square matrix other than the diagonal elements are called *off-diagonal* or *non-diagonal* elements.

A *diagonal* matrix is a square matrix having zero for all its non-diagonal elements.

Example:

$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ is a diagonal matrix.

A *triangular* matrix is a square matrix with all elements above (or below) the diagonal being zero.

Example:

$\mathbf{A} = \begin{bmatrix} 1 & 4 & 6 \\ 0 & -2 & 9 \\ 0 & 0 & 7 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 3 & 0 & 0 \\ -7 & 2 & 0 \\ 1 & -2 & 5 \end{bmatrix}$ are triangular matrices.

A is an *upper triangular matrix* and **B** is a *lower triangular matrix*.

A diagonal matrix having all diagonal elements equal to unity is called an *identity* matrix, or sometimes a *unit* matrix. It is usually denoted by the letter **I**. For example,

$\mathbf{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ is an identity matrix of order 4.

For any matrix **A** of order $m \times n$, $\mathbf{I}_m\mathbf{A}_{m\times n} = \mathbf{A}_{m\times n}\mathbf{I}_n = \mathbf{A}_{m\times n}$.

A *symmetric* matrix is a square matrix when it equals its transpose.

Example:

Let $\mathbf{A} = \begin{bmatrix} 1 & -2 & 3 \\ -2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$, then $\mathbf{A}' = \begin{bmatrix} 1 & -2 & 3 \\ -2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix} = \mathbf{A}$.

The matrix **A** is symmetric. In symmetric matrices, the area above the diagonal is a mirror image of the area below the diagonal.

Vectors whose every element is unity are called *summing vectors* because they can be used to express a sum of numbers in matrix notation as an inner product.

Example:

$\mathbf{1}' = [1\ \ 1\ \ 1\ \ 1]$ is a summing vector of order 4. For $\mathbf{x}' = [2\ \ 4\ \ {-3}\ \ 8]$,

$$\mathbf{1}'\mathbf{x} = [1\ \ 1\ \ 1\ \ 1]\begin{bmatrix} 2 \\ 4 \\ -3 \\ 8 \end{bmatrix} = 2 + 4 - 3 + 8 = 11 = \mathbf{x}'\mathbf{1}.$$

It follows that $\mathbf{1}'_n\mathbf{1}_n = n$ and $\mathbf{1}_m\mathbf{1}'_n = \mathbf{J}_{m\times n}$, a matrix having all elements unity.

A square $\mathbf{J}$ matrix is denoted by $\mathbf{J}_n$. $\mathbf{J}_n = \mathbf{1}_n\mathbf{1}'_n$ and $\mathbf{J}_n^2 = n\mathbf{J}_n$. A useful variant of $\mathbf{J}_n$ is

$$\bar{\mathbf{J}}_n = \frac{1}{n}\mathbf{J}_n,\ \bar{\mathbf{J}}_n^2 = \bar{\mathbf{J}}_n.$$

Therefore, $\bar{\mathbf{J}}_n$ is an *idempotent* matrix which satisfies $\mathbf{K}^2 = \mathbf{K}$.

$\mathbf{C}_n = \mathbf{I} - \bar{\mathbf{J}}_n = \mathbf{I} - \frac{1}{n}\mathbf{J}_n$, is known as a *centring matrix*. It follows that

$\mathbf{C} = \mathbf{C}' = \mathbf{C}^2$, $\mathbf{C1} = 0$, and $\mathbf{CJ} = \mathbf{JC} = 0$.

An *orthogonal* matrix $\mathbf{A}$ is a matrix having the property $\mathbf{AA}' = \mathbf{I} = \mathbf{A}'\mathbf{A}$.

Example:

$$\text{Let } \mathbf{A} = \begin{bmatrix} 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \\ 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & -3/\sqrt{12} \end{bmatrix}$$

You can verify that $\mathbf{AA}' = \mathbf{I}$. Therefore, $\mathbf{A}$ is an orthogonal matrix. A matrix in the form shown above is called a *Helmert* matrix.

A *quadratic form* is the product of a row vector $\mathbf{x}'$, a matrix $\mathbf{A}$, and the column vector $\mathbf{x}$, that is, $\mathbf{x}'\mathbf{Ax}$. This is a quadratic function of the xs. Notice that to result in the same quadratic function of xs, you can use many different matrices. Each matrix has the same diagonal elements, and the sum of each pair of symmetrically placed off-diagonal elements a_{ij} and a_{ji} is the same.

For any particular quadratic form there is a unique *symmetric matrix* **A** for which the quadratic form can be expressed as **x′Ax**:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n} x_i^2 a_{ii} + 2\sum_{j=i+1}^{n}\sum_{i=1}^{n} x_i x_j a_{ij}$$

Example:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\mathbf{x}$$

$$= a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2(a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3) .$$

When $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \\ 3 & 1 & 4 \end{bmatrix}$, then

$$\mathbf{x}'\mathbf{A}\mathbf{x} = [x_1 \;\; x_2 \;\; x_3] \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \\ 3 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1^2 + 5x_2^2 + 4x_3^2 + 4x_1x_2 + 6x_2x_3 + 2x_2x_3$$

When **x′Ax** > 0 for all **x** other than **x** = 0, then **x′Ax** is a *positive definite* quadratic form, and **A** = **A′** is correspondingly a *positive definite (p.d.) matrix*.

Example:

Let $\mathbf{A} = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}$, then

$$\mathbf{x}'\mathbf{A}\mathbf{x} = [x_1 \;\; x_2 \;\; x_3] \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= 2x_1^2 + 5x_2^2 + 2x_3^2 + 4x_1x_2 + 2x_2x_3 + 2x_2x_3$$

$$= (x_1 + 2x_2)^2 + (x_1 + x_3)^2 + (x_2 + x_3)^2 > 0 \text{ other than } \mathbf{x} = 0.$$

The matrix **A** is a positive definite matrix.

When **x′Ax** ≥ 0 for all **x** and **x′Ax** = 0 for some **x** ≠ 0, then **x′Ax** is a *positive semi-definite* quadratic form, and **A** = **A′** is correspondingly a *positive semi-definite (p.s.d.) matrix*. The two classes of matrices taken together, positive definite and positive semi-definite, are called *nonnegative definite* (n.n.d.).

Example:

Let $\mathbf{A} = \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix}$, then

$$\mathbf{x}'\mathbf{A}\mathbf{x} = [x_1 \ \ x_2 \ \ x_3] \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= 37x_1^2 + 13x_2^2 + 17x_3^2 - 4x_1x_2 - 48x_2x_3 - 6x_2x_3$$

$$= (x_1 - 2x_2)^2 + (6x_1 - 4x_3)^2 + (3x_2 - x_3)^2$$

This is zero for $\mathbf{x}' = [2 \ \ 1 \ \ 3]$, and for any scalar multiple thereof, as well as for $\mathbf{x} = 0$. Then the matrix $\mathbf{A}$ is *positive semi-definite*.

Example:

$$\mathbf{x}'\mathbf{C}\mathbf{x} = \mathbf{x}'(\mathbf{I} - \bar{\mathbf{J}})\mathbf{x} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

is a positive semi-definite quadratic form because it is positive, except for being zero when all the x_is are equal. Its matrix, $\mathbf{I} - \bar{\mathbf{J}}$, which is idempotent, is also p.s.d., as are all symmetric idempotent matrices (except $\mathbf{I}$, which is the only p.d. idempotent matrix).

Determinant

The *determinant* of a square matrix of order n, (that is, $\mathbf{A} = \{a_{ij}\}$, $i, j = 1, 2, \ldots, n$) is the sum of all possible products of n elements of **A** such that

1. each product has one and only one element from every row and column of **A**,
2. the sign of a product being $(-1)^p$ for $p = \sum_{i=1}^{n} n_i$, where by writing
    a. the product with its i subscripts in natural order $a_{1j_1} a_{2j_2} \cdots a_{ij_i} \cdots a_{nj_n}$
    b. the j subscripts j_i, $i = 1, 2, \ldots, n$, being the first n integers in some order

n_i is defined as the number of js less than j_i that follow j_i in this order.

Therefore, the determinant of a matrix **A**, denoted by |**A**|, is a polynomial of the elements of a square matrix. It is a scalar. It is the sum of certain products of the elements of the matrix from which it is derived, each product being multiplied by +1 or –1 according to certain rules.

Example:

$$|\mathbf{A}| = \begin{vmatrix} 3 & 7 \\ 2 & 6 \end{vmatrix} = 3(6) - 7(2) = 4$$

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

$$= 1(50 - 48) - 2(40 - 42) + 3(32 - 35) = -3$$

The determinant that multiplies each element of the chosen row (in this case, the first row) is the determinant derived from |**A**| by crossing out the row and column containing the element concerned. For example, the first element, 1, is multiplied by the determinant $\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}$, which is obtained from |**A**| through crossing out the first row and column. Determinants obtained in this way are called *minors* of |**A**|, that is, $\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}$ is the minor of the element 1 in |**A**|, and $\begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix}$ is the minor of element 2.

The (+1) and (–1) factors in the expansion are decided on according to the following rule:

If **A** is written in the form $\mathbf{A} = \{a_{ij}\}$, the product of a_{ij} and its minor in the expansion of determinant |**A**| is multiplied by $(-1)^{i+j}$.

Therefore, because the element 1 in the example is the element a_{11}, its product with its minor is multiplied by $(-1)^{1+1} = +1$. For element 2, which is a_{12}, its product with its minor is multiplied by $(-1)^{1+2} = -1$.

Denote the minor of the element a_{ij} by $|\mathbf{M}_{ij}|$, where $\mathbf{M}_{ij}$ is a sub-matrix of **A** obtained by deleting the ith row and the jth column. The determinant of an n-order matrix is obtained by *expansion by the elements of a row* (*or column*), or *expansion by minors*:

$$|\mathbf{A}| = \sum_{j=1}^{n} a_{ij}(-1)^{i+j} \, |\mathbf{M}_{ij}| \text{ for any row } i\text{, or}$$

$$|\mathbf{A}| = \sum_{i=1}^{n} a_{ij}(-1)^{i+j} \, |\mathbf{M}_{ij}| \text{ for any column } j.$$

The signed minor $(-1)^{i+j} |\mathbf{M}_{ij}|$ is called a *cofactor*: $c_{ij} = (-1)^{i+j} |\mathbf{M}_{ij}|$.

This expansion is used recurrently when n is large, that is, each $|\mathbf{M}_{ij}|$ is expanded by the same procedure.

This method of evaluation requires lengthy computing for determinants of order exceeding 3 to 4. Fortunately, easier methods exist, but they are based on this expansion by minors method.

Inverse Matrices

The *inverse* of a square matrix **A** is a matrix whose product with **A** is the identity matrix **I**. The inverse matrix is denoted by $\mathbf{A}^{-1}$. The concept of "dividing" by **A** in matrix algebra is replaced by the concept of multiplying by the inverse matrix $\mathbf{A}^{-1}$.

An inverse matrix $\mathbf{A}^{-1}$ should have the following properties:

- $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$
- $\mathbf{A}^{-1}$ unique for given **A**.

An *adjugate* (or *adjoint*) of matrix **A**, denoted by adj **A**, is obtained by replacing the elements in **A** by their cofactors and then transposing it.

The inverse of matrix **A**, $\mathbf{A}^{-1}$, can be described as the adjugate of **A** multiplied by the scalar 1/|**A**|:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\text{adj }\mathbf{A}$$

where adj **A** is the adjugate (or adjoint) matrix of **A**. Therefore,

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{bmatrix} \mathbf{A}\text{ with every element} \\ \text{replaced by its cofactor} \end{bmatrix}^{\text{transposed}}$$

Example:

Let **A** = $\begin{bmatrix} 2 & 5 \\ 3 & 9 \end{bmatrix}$, the determinant of **A** is |**A**| = $\begin{vmatrix} 2 & 5 \\ 3 & 9 \end{vmatrix}$ = 18 – 15 = 3.

- The cofactor for $a_{11} = 2$ is $(-1)^{1+1}\,|9| = 9$.
- The cofactor for $a_{12} = 5$ is $(-1)^{1+2}\,|3| = -3$.
- The cofactor for $a_{21} = 3$ is $(-1)^{2+1}\,|5| = -5$.
- The cofactor for $a_{22} = 9$ is $(-1)^{2+2}\,|2| = 2$.

The adjugate matrix is $\begin{bmatrix} 9 & -5 \\ -3 & 2 \end{bmatrix}$, and so the inverse is

$$\mathbf{A}^{-1} = \frac{1}{3}\begin{bmatrix} 9 & -5 \\ -3 & 2 \end{bmatrix}$$

Conditions for existence of the inverse are

1. $\mathbf{A}^{-1}$ can exist only when **A** is square
2. $\mathbf{A}^{-1}$ does exist only if |**A**| is nonzero.

A square matrix is said to be *singular* when its determinant is zero and *non-singular* when its determinant is nonzero.

Several computing procedures for inverting matrices are based on solving linear equations by successive elimination and backward substitution. The basic idea is as follows:

To solve the equations $\mathbf{Ax} = \mathbf{b}$, write the following matrix $[\mathbf{A}\ \ \mathbf{I}]$, where $\mathbf{I}$ is the identity matrix of the same order as $\mathbf{A}$. Perform row operations to this matrix to make the left sub-matrix become $\mathbf{I}$, whereupon the right sub-matrix is $\mathbf{A}^{-1}$. Thus, starting with $[\mathbf{A}\ \ \mathbf{I}]$, you now have $[\mathbf{I}\ \ \mathbf{A}^{-1}]$, providing $\mathbf{A}^{-1}$ exists.

Linear Dependence and Rank

Define **X** as the matrix having columns $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$, and **a** as the vector of *a*s:

$\mathbf{X} = [\mathbf{x_1}\ \ \mathbf{x_2}\ \ \ldots\ \ \mathbf{x_n}]$ and $\mathbf{a}' = [a_1\ \ a_2\ \ \ldots\ \ a_n]$.

Then the *linear combination* of the set of *n* vectors is

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \sum_{i=1}^{n} a_i\mathbf{x}_i = \mathbf{Xa}$$

You see that

- **Xa** is a column vector, a linear combination of the columns of **X**.
- **b′X** is a row vector, a linear combination of the rows of **X**.
- **AB** is a matrix. Its rows are linear combinations of the rows of **B**, and its columns are linear combinations of the columns of **A**.

The vector **Xa** is sometimes called the *linear transformation* of the vector **a** to the vector **Xa**, with **X** being the matrix of the transformation.

Example:

Let $\mathbf{X} = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 4 & 1 \\ -1 & 3 & 5 \\ 6 & 7 & 5 \end{bmatrix}$ and $\mathbf{a}' = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$, then

$$\mathbf{Xa} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 4 & 1 \\ -1 & 3 & 5 \\ 6 & 7 & 5 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} a_1 - 2a_2 + 0a_3 \\ 0a_1 + 4a_2 + a_3 \\ -a_1 + 3a_2 + 5a_3 \\ 6a_1 + 7a_2 + 5a_3 \end{bmatrix}$$

is a vector for any scalar values a_1, a_2, and a_3.

If there exists a vector $\mathbf{a} \neq 0$, such that $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \mathbf{0}$, then provided none of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is null, those vectors are said to be *linearly dependent vectors*.

An alternative statement of the definition is the following

If $\mathbf{Xa} = 0$ for some non-null **a**, then the columns of **X** are *linearly dependent vectors*, provided none is null.

If $\mathbf{a} = 0$ is the only vector for which $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n = \mathbf{0}$, then provided none of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is null, those vectors (the columns of **X**) are said to be *linearly independent vectors*.

The *rank* of a matrix is the number of linearly independent rows (and columns) in the matrix. The rank of **A** is denoted by $r_{\mathbf{A}}$ or $r(\mathbf{A})$.

When $r(\mathbf{A}_{n \times n}) = n$, then **A** is *non-singular*, that is, $\mathbf{A}^{-1}$ exists.

When $r(\mathbf{A}_{n \times n}) < n$, then **A** is *singular* and $\mathbf{A}^{-1}$ does not exist.

When $r(\mathbf{A}_{p \times q}) = p < q$, then **A** has *full row rank,* that is, its rank equals its number of rows.

When $r(\mathbf{A}_{p \times q}) = q < p$, then **A** has *full column rank,* that is, its rank equals its number of columns.

When $r(\mathbf{A}_{n \times n}) = n$, **A** has *full rank*, that is, its rank equals its order, it is nonsingular, its inverse exists, and it is called *invertible*.

Because it is easier to work with rank than determinants, you often determine the existence of $\mathbf{A}^{-1}$ by ascertaining whether $r(\mathbf{A}) < n$ or $r(\mathbf{A}) = n$ rather than by ascertaining whether $|\mathbf{A}|$ is zero or not.

Example:

Consider the following vectors:

$$\mathbf{x}_1 = \begin{bmatrix} 3 \\ -6 \\ 9 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 5 \\ -5 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{x}_4 = \begin{bmatrix} -6 \\ 12 \\ -18 \end{bmatrix} \quad \mathbf{x}_5 = \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}$$

Because $2\mathbf{x}_1 + \mathbf{x}_4 = \begin{bmatrix} 6 \\ -12 \\ 18 \end{bmatrix} + \begin{bmatrix} -6 \\ 12 \\ -18 \end{bmatrix} = 0$, that is,

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_4 = \mathbf{0} \text{ for } \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \text{ which is non-null,}$$

$\mathbf{x}_1$ and $\mathbf{x}_4$ are *linearly dependent* vectors. So are $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ because

$$2\mathbf{x}_1 + 3\mathbf{x}_2 - 3\mathbf{x}_3 = \begin{bmatrix} 6 \\ -12 \\ 18 \end{bmatrix} + \begin{bmatrix} 0 \\ 15 \\ -15 \end{bmatrix} + \begin{bmatrix} -6 \\ -3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}$$

However, consider

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 = \begin{bmatrix} 3a_1 \\ -6a_1 \\ 9a_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 5a_2 \\ -5a_2 \end{bmatrix} = \begin{bmatrix} 3a_1 \\ -6a_1 + 5a_2 \\ 9a_1 - 5a_2 \end{bmatrix}$$

There are no values a_1 and a_2, which makes it a null vector other than $a_1 = 0 = a_2$. Therefore, $\mathbf{x}_1$ and $\mathbf{x}_2$ are *linearly independent* vectors.

The rank of the matrix $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \mathbf{x}_5 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 2 & -6 & 2 \\ -6 & 5 & 1 & 12 & -3 \\ 9 & -5 & 1 & -18 & 3 \end{bmatrix}$

is 3.

The calculation of ranks can be performed by row operations, which is discussed in many matrix algebra textbooks.

Note The linear dependence or linear independence of vectors is a characteristic pertaining to a set of vectors of the same order. It is not a characteristic of individual vectors.

Any non-null matrix **A** of rank r is equivalent to

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \mathbf{C}$$

where $\mathbf{I}_r$ is the identity matrix of order r, and the null sub-matrices are of appropriate order to make **C** the same order as **A**. For **A** of order $m \times n$, **P** and **Q** are non-singular matrices of order m and n, respectively, being products of elementary operators.

The matrix **C** is called the *equivalent canonical form* or the *canonical form under equivalence*. It always exists and can be used to determine the rank of **A**.

Generalized Inverse

The generalized inverse plays an important role in understanding the solutions to linear equations

$\mathbf{Ax} = \mathbf{y}$, when $\mathbf{A}$ has no inverse (but has generalized inverse).

A *generalized inverse* of a matrix $\mathbf{A}$ is any matrix $\mathbf{G}$ such that $\mathbf{AGA} = \mathbf{A}$. An alternative symbol for $\mathbf{G}$ is $\mathbf{A}^{-}$.

When $\mathbf{A}$ is not full rank, as occurs in many general linear models with $\mathbf{A} = \mathbf{X}'\mathbf{X}$, an infinite number of generalized inverses exist. Several ways of obtaining $\mathbf{G}$ exist. The approach to obtaining a generalized inverse used by SAS is to partition a singular matrix into several sets of matrices.

Example:

Consider an $\mathbf{X}'\mathbf{X}$ matrix of rank r that can be partitioned as

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

where $\mathbf{A}_{11}$ is $r \times r$ and of rank r. Then $\mathbf{A}_{11}^{-1}$ exists, and a generalized inverse of $\mathbf{X}'\mathbf{X}$ is

$$(\mathbf{X}'\mathbf{X})^{-} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Because the generalized inverse is not unique, the solutions to normal equations in general linear models, $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$, are not unique either. However, a class of linear functions ($\mathbf{Lb}$) called *estimable functions* exists, and $\mathbf{Lb}$ and its variance are invariant through all possible generalized inverses. In other words, the linear combination $\mathbf{Lb}$ is unique and is an unbiased estimate of $\mathbf{L}\boldsymbol{\beta}$.

Eigenvalues and Eigenvectors

Consider equations

$$\mathbf{Ax} = \lambda\mathbf{x} \quad \text{or} \quad (\mathbf{A}-\lambda\mathbf{I})\mathbf{x} = 0$$

for **x** (a vector) and λ (a scalar), and solve for **x**.

If the matrix $\mathbf{A}-\lambda\mathbf{I}$ is non-singular, the unique solution to these equations is $\mathbf{x} = 0$. You only get a nontrivial solution when $|\mathbf{A}-\lambda\mathbf{I}| = 0$. If you expand this determinant, the condition becomes a polynomial equation in λ of degree p. This is called the *characteristic equation* of **A**. Its p roots (which may be real or complex, simple or multiple) are called *eigenvalues* (or proper values, characteristic values, or latent roots) of **A**. If λ is an eigenvalue, a nonzero vector **x** satisfying $\mathbf{Ax} = \lambda\mathbf{x}$ is called an *eigenvector* (or proper vector, characteristic vector, or latent vector) corresponding to λ. It is often convenient to normalize each eigenvector to have a squared length of 1.

Example:

The matrix $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$ has a characteristic equation:

$\left| \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0$; that is, $\begin{vmatrix} 1-\lambda & 4 \\ 9 & 1-\lambda \end{vmatrix} = 0$.

$(1-\lambda)^2 - 36 = 0$, or $\lambda = -5$ or 7

It can be seen that

$$\begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ -3 \end{bmatrix} = -5\begin{bmatrix} 2 \\ -3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = 7\begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

therefore, $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$ is an eigenvector corresponding to the eigenvalue –5,

and $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ is an eigenvector corresponding to the eigenvalue 7.

Some of the Basic Properties of Eigenvalues and Eigenvectors

For $\mathbf{Ax} = \lambda\mathbf{x}$,

1. $\mathbf{A}^k\mathbf{x} = \lambda^k\mathbf{x}$ and $\mathbf{A}^{-1}\mathbf{x} = \lambda^{-1}\mathbf{x}$, when $\mathbf{A}$ is non-singular
2. $c\mathbf{Ax} = c\lambda\mathbf{x}$ for any scalar c
3. $f(\mathbf{A})\mathbf{x} = f(\lambda)\mathbf{x}$ for any polynomial function $f(\mathbf{A})$
4. $\sum_{i=1}^{n} \lambda_i = \mathrm{tr}(\mathbf{A})$ and $\prod_{i=1}^{n} \lambda_i = |\mathbf{A}|$, that is, the sum of eigenvalues of a matrix equals its trace, and their product equals its determinant
5. If $\mathbf{A}$ is symmetric, then
   - eigenvalues of matrix $\mathbf{A}$ are all real
   - $\mathbf{A}$ is diagonable
   - eigenvectors are orthogonal to each other
   - the rank of $\mathbf{A}$ equals the number of nonzero eigenvalues
   - positive definite matrices have eigenvalues all greater than zero and vice versa.

The matrix $\mathbf{A}$ is *diagonable* when a non-singular matrix $\mathbf{X}$ exists, and consists of the n eigenvectors, that is, $\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n]$, where all n eigenvectors are linearly independent, such that

$$\mathbf{X}^{-1}\mathbf{AX} = \mathbf{D} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$$

For symmetric matrix $\mathbf{A}$, because the eigenvectors are orthogonal to each other, the above equation becomes

$$\mathbf{X}'\mathbf{AX} = \mathbf{D} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$$

It follows that $\mathbf{A}$ can be written as $\mathbf{XDX}'$, or

$$\mathbf{A} = \lambda_1\mathbf{x}_1\mathbf{x}_1' + \lambda_2\mathbf{x}_2\mathbf{x}_2' + \ldots + \lambda_n\mathbf{x}_n\mathbf{x}_n'$$

This is called the *spectral decomposition* of the matrix $\mathbf{A}$.

When $\mathbf{A}$ is nonsingular, the spectral decomposition of $\mathbf{A}^{-1}$ is

$$\mathbf{A}^{-1} = \lambda_1^{-1}\mathbf{x}_1\mathbf{x}_1' + \lambda_2^{-1}\mathbf{x}_2\mathbf{x}_2' + \ldots + \lambda_n^{-1}\mathbf{x}_n\mathbf{x}_n'$$

When $\mathbf{A}$ is singular, a generalized inverse of $\mathbf{A}$ can be obtained from its spectral decomposition in exactly the same way, by omitting the terms for which $\lambda_i = 0$.

Cholesky Root

If **A** is positive definite of size $n \times n$, you can find an upper triangular matrix **U** such that $\mathbf{A} = \mathbf{U}'\mathbf{U}$, so that **U** is sort of a square root of **A**, also referred to as the *Cholesky root* of the matrix **A**.

The rule for the construction of **U** is simply

(column i of **U**) $\times$ (column j of **U**) $= a_{ij}$,

that is,

$$\begin{aligned} u_{11}^2 &= a_{11}, \\ u_{11}u_{12} &= a_{12}, \\ u_{11}u_{13} &= a_{13}, \\ &\cdots \\ u_{12}^2 + u_{22}^2 &= a_{22}, \\ u_{12}u_{13} + u_{22}u_{23} &= a_{23}, \\ &\cdots \\ u_{13}^2 + u_{23}^2 + u_{33}^2 &= a_{33}, \end{aligned}$$

and so on.

The procedure for forming **U** (called the *square-root* or *Cholesky* procedure) provides a basis for an excellent numerical method for solving simultaneous linear equations and inverting matrices. This procedure can also be applied when **A** is only positive semi-definite. If, when a zero u_{ii} occurs, you set all subsequent elements in the same row of **U** to zero, the relation $\mathbf{A} = \mathbf{U}'\mathbf{U}$ remains and the matrix $\mathbf{U}^{-}(\mathbf{U}')^{-}$ is a generalized inverse of **A**.