MATRIX ALGEBRA

By gathering the elements of the equation $B = A^{-1}$ under (16) and (17), we derive the following expression for the inverse of a 2×2 matrix:

$$(18) \qquad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Here the quantity $a_{11}a_{22} - a_{12}a_{21}$ in the denominator of the scalar factor which multiplies the matrix on the RHS is the so-called determinant of the original matrix A denoted by $\det(A)$ or $|A|$. The inverse of the matrix A exist if and only if the determinant has a nonzero value.

The formula above can be generalised to accommodate square matrices of an arbitrary order n. However, the general formula rapidly becomes intractable as the value of n increases. In the case of a matrix of order 3×3, the inverse is given by

$$(19) \qquad A^{-1} = \frac{1}{|A|} \begin{bmatrix} c_{11} & -c_{21} & c_{31} \\ -c_{12} & c_{22} & -c_{32} \\ c_{13} & -c_{23} & c_{33} \end{bmatrix},$$

where c_{ij} is the determinant of a submatrix of A formed by omitting the ith row and the jth column, and where $|A|$ is the determinant of A which is given by

$$(20) \qquad \begin{aligned} |A| &= a_{11}c_{11} - a_{12}c_{12} + a_{13}c_{13} \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}\big(a_{22}a_{33} - a_{23}a_{32}\big) - a_{12}\big(a_{21}a_{33} - a_{31}a_{23}\big) + a_{13}\big(a_{21}a_{32} - a_{31}a_{22}\big). \end{aligned}$$

Already this is becoming unpleasantly complicated. The rules for the general case can be found in textbooks of matrix algebra.

In fact, the general formula leads to a method of deriving the inverse matrix which is inefficient from a computational point of view. It is both laborious and prone to numerical rounding errors. The method of finding the inverse which is used by the computer is usually that of Gaussian reduction which depends upon the so-called elementary matrix transformations which are described below.

The basic rules affecting matrix inversion are as follows:

(21)

(i)   The inverse of A^{-1} is A itself, that is $(A^{-1})^{-1} = A$,

(ii)   The inverse of the transpose is the transpose of the inverse, that is, $(A')^{-1} = (A^{-1})'$,

(iii)   If $C = AB$, then $C^{-1} = B^{-1}A^{-1}$.

The first of these comes directly from the definition of the inverse matrix whereby $AA^{-1} = A^{-1}A = I$. The second comes from comparing the equation $(AA^{-1})' = (A^{-1})'A' = I$, which is the consequence of the reversal rule of matrix transposition, with the equation $A'(A')^{-1} = (A')^{-1}A' = I$ which is from the definition of the

inverse of A'. The third rule, which is the reversal rule of matrix inversion, is understood by considering the two equations which define the inverse of the product AB:

(22)
$$\text{(i)} \quad (AB)^{-1}AB = B^{-1}\{A^{-1}A\}B = B^{-1}B = I,$$
$$\text{(ii)} \quad AB(AB)^{-1} = A\{BB^{-1}\}A^{-1} = AA^{-1} = I.$$

One application of matrix inversion is to the problem of finding the solution of a system of linear equations such as the system under (1) which is expressed in summary matrix notation under (3). Here we are looking for the value of the vector x of unknowns given the vector y of constants and the matrix A of coefficients. The solution is indicated by the fact that

(23)   If $y = Ax$ and if A^{-1} exists, then $A^{-1}y = A^{-1}Ax = x$.

Example. Consider a pair of simultaneous equations written in matrix form:

$$\begin{bmatrix} 1 & 5 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}.$$

The solution using the formula for the inverse of 2×2 matrix found under (18) is

(24)
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{-1}{7} \begin{bmatrix} 3 & -5 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

We have not used the formula under (19) to find the solution of the system of three equations under (4) because there is an easier way which depends upon the method of Gaussian elimination.

Elementary Operations

There are three elementary row operations which serve, in a process described as Gaussian elimination, to reduce an arbitrary matrix to one which has units and zeros on its principal diagonal and zeros everywhere else. We shall illustrate these elementary operations in the context of a 2×2 matrix; but there should be no difficulty in seeing how they may be applied to square matrices of any order.

The first operation is the multiplication of a row of the matrix by a scalar factor:

(25)
$$\begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \lambda a_{11} & \lambda a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

The second operation is the addition of one row to another, or the subtraction of one row from another:

(26)
$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} - a_{11} & a_{22} - a_{12} \end{bmatrix}.$$

The third operation is the interchange of one row with another:

(27)
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{bmatrix}.$$

If the matrix A possesses an inverse, then, by the application of a succession of such operations, one may reduce it to the identity matrix. Thus, if the elementary operations are denoted by $E_j; j = 1, \ldots, q$, then we have

(28)
$$\{E_q \cdots E_2 E_1\} A = I \quad \text{or, equivalently,} \quad BA = I,$$
$$\text{where} \quad B = E_q \cdots E_2 E_1 = A^{-1}.$$

Example. A 2×2 matrix which possesses an inverse may be reduced to the identity matrix by applying four elementary operations:

(29)
$$\begin{bmatrix} 1/\alpha_{11} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} 1 & \alpha_{12}/\alpha_{11} \\ \alpha_{21} & \alpha_{22} \end{bmatrix},$$

(30)
$$\begin{bmatrix} 1 & 0 \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} 1 & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} = \begin{bmatrix} 1 & \beta_{12} \\ 0 & \beta_{22} - \beta_{21}\beta_{12} \end{bmatrix},$$

(31)
$$\begin{bmatrix} 1 & 0 \\ 0 & 1/\gamma_{22} \end{bmatrix} \begin{bmatrix} 1 & \gamma_{12} \\ 0 & \gamma_{22} \end{bmatrix} = \begin{bmatrix} 1 & \gamma_{12} \\ 0 & 1 \end{bmatrix},$$

(32)
$$\begin{bmatrix} 1 & -\gamma_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \gamma_{12} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Here, we have expressed the results of the first two opertations in new symbols in the interests of notational simplicity. Were we to retain the original notation throughout, then we could obtain the expression for the inverse matrix by forming the product of the four elementary matrices.

It is notable that, in this example, we have used only the first two of the three elementary operations. The third would be called for if we were confronted by a matrix in the form of

(33)
$$\begin{bmatrix} \alpha & \beta \\ \gamma & 0 \end{bmatrix};$$

for then we should wish to reorder the rows before embarking on the process of elimination.

Example. Consider the equations under (4) which can be written in matrix format as

(34)
$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 5 & -6 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \\ 8 \end{bmatrix}.$$

Taking four times the first equation from the second equation and three times the first equation from the third gives

(35)
$$
\begin{bmatrix} 1 & 3 & 2 \\ 0 & -7 & -14 \\ 0 & -7 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ -28 \\ -19 \end{bmatrix}.
$$

Taking the second equation of this transformed system from the third gives

(36)
$$
\begin{bmatrix} 1 & 3 & 2 \\ 0 & -7 & -14 \\ 0 & 0 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ -28 \\ 9 \end{bmatrix}.
$$

An equivalent set of equations is obtained by dividing the second equation by -7 and the third equation by 9:

(37)
$$
\begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 9 \\ 4 \\ 1 \end{bmatrix}.
$$

Now we can solve the system by backsubstitution. That is to say

(38)
$$
\begin{aligned}
x_3 &= 1, \\
x_2 &= 4 - 2x_3 = 2, \\
x_3 &= 9 - 3x_2 - 2x_3 = 1.
\end{aligned}
$$

Exercise. As an exercise, you are asked to find the matrix which transforms the equations under (34) to the equations of (35) and the matrix which transforms the latter into the final equation under (37).

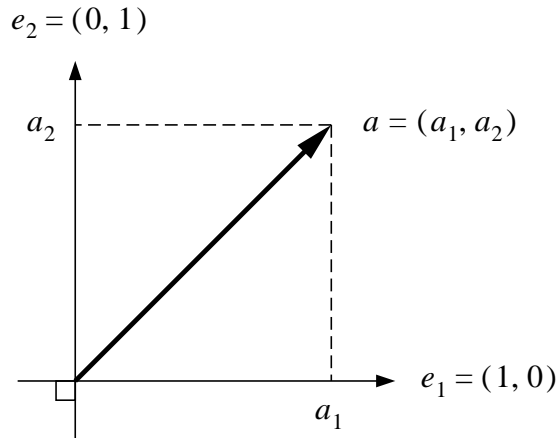### Geometric Vectors in the Plane

A vector of order n, which is defined as an ordered set of n real numbers, can be regarded as point in a space of n dimensions. It is difficult to draw pictures in three dimensions and it is impossible directly to envisage spaces of more than thee dimensions. Therefore, in discussing the geometric aspects of vectors and matrices, we shall confine ourselves to the geometry of the two-dimensional plane.

A point in the plane is primarily a geometric object; but, if we introduce a coordinate system, then it may be described in terms of an ordered pair of numbers.

In constructing a coordinate system, it is usually convenient to introduce two perpendicular axes and to use the same scale of measurement on both axes. The point of intersection of these axes is called the origin and it is denoted by 0. The point on the first axis at a unit distance from the origin 0 is denoted by e_1 and the point on the second axis at a unit distance from 0 is denoted by e_2.

An arbitrary point a in the plane can be represented by its coordinates a_1 and a_2 relative to these axes. The coordinates are obtained by the perpendicular projections of the point onto the axes. If we are prepared to identify the point with

its coordinates, then we may write $a = (a_1, a_2)$. According to this convention, we may also write $e_1 = (1,0)$ and $e_2 = (0,1)$.


Figure 1. The coordinates of a vector a relative to two perpendicular axes

The directed line segment running from the origin 0 to the point a is described as a geometric vector which is bound to the origin. The ordered pair $(a_1, a_2) = a$ may be described as an algebraic vector. In fact, it serves little purpose to make a distinction between these two entities—the algebraic vector and the geometric vector—which may be regarded hereafter as alternative representations of the same object a. The unit vectors $e_1 = (1,0)$ and $e_2 = (0,1)$, which serve, in fact, to define the coordinate system, are described as the basis vectors.

The sum of two vectors $a = (a_1, a_2)$ and $b = (b_1, b_2)$ is defined by

(39)
$$a + b = (a_1, a_2) + (b_1, b_2)$$
$$= (a_1 + b_1, a_2 + b_2).$$

The geometric representation of vector addition corresponds to a parallelogram of forces. Forces, which have both magnitude and direction, may be represented by directed line segments whose lengths correspond to the magnitudes. Hence forces may be described as vectors; and, as such, they obey the law of addition given above.
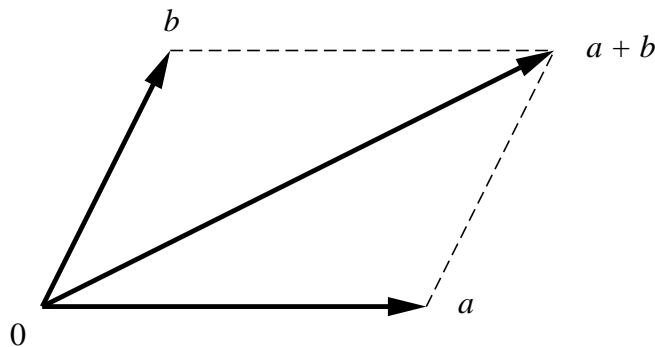
Figure 2. The parallelogram law of vector addition

9

If $a = (a_1, a_2)$ is a vector and λ is a real number, which is also described as a scalar, then the product of a and λ is defined by

(40)
$$\lambda a = \lambda(a_1, a_2)$$
$$= (\lambda a_1, \lambda a_2).$$

The geometric counterpart of multiplication by a scalar is a stretching or a contraction of the vector which affects its length but not its direction.

The axes of the coordinate system are provided by the lines $\mathcal{E}_1 = \{\lambda e_1\}$ and $\mathcal{E}_2 = \{\lambda e_2\}$ which are defined by letting λ take every possible value. In terms of the basis vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$, the point $a = (a_1, a_2)$ can be represented by

(41)
$$a = (a_1, a_2)$$
$$= a_1 e_1 + a_2 e_2.$$

Norms and Inner Products.

The length or norm of a vector $a = (a_1, a_2)$ is
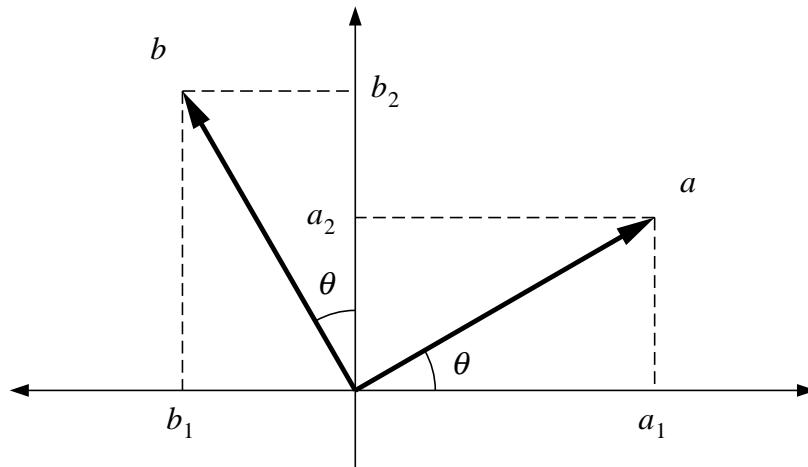
(42)
$$\|a\| = \sqrt{a_1^2 + a_2^2};$$

and this may be regarded either as an algebraic definition or as a consequence of the geometric theorem of Pythagoras.

The inner product of the vectors $a = (a_1, a_2)$ and $b = (b_1, b_2)$ is the scalar quantity

(43)
$$a'b = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = a_1 b_1 + a_2 b_2.$$

This is nothing but the product of the matrix a' of order 1×2 and the matrix b of order 2×1.

Figure 3. Two vectors $a = (a_1, a_2)$ and $b = (b_1, b_2)$ which are at right angles have a zero-valued inner product.

The vectors a, b are said to be mutually orthogonal if $a'b = 0$. It can be shown by simple trigonometry that a, b fulfil the condition of orthogonality if and only if the line segments are at right angles. Indeed, this condition is indicated by the etymology of the word orthogonal—Gk. *orthos* right, *gonia* angle.

In the diagram, we have a vector a of length $p = \sqrt{a_1^2 + a_2^2}$ and a vector b of length $q = \sqrt{b_1^2 + b_2^2}$. The vectors are at right angles. By trigonometry, we find that

(44)
$$
\begin{aligned}
a_1 &= p \cos \theta, \\
a_2 &= p \sin \theta, \\
b_1 &= -q \sin \theta, \\
b_2 &= q \cos \theta.
\end{aligned}
$$

Therefore

(45)
$$
a'b = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = a_1 b_1 + a_2 b_2 = 0.
$$

Simultaneous Equations

Consider the equations

(46)
$$
\begin{aligned}
ax + by &= e, \\
cx + dy &= f,
\end{aligned}
$$

which describe two lines in the plane. The coordinates (x, y) of the point of intersection of the lines is the algebraic solution of the simultaneous equations.

The equations may be written in matrix form as

(47)
$$
\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix}.
$$

The necessary and sufficient condition for the existence of a unique solution is that

(48)
$$
\text{Det} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc \neq 0.
$$

Then the solution is given by

(49)
$$
\begin{aligned}
\begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \begin{bmatrix} e \\ f \end{bmatrix} \\
&= \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix}.
\end{aligned}
$$

We may prove that

(50)
$$\text{Det} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = 0 \quad \text{if and only if}$$

$$a = \lambda c \quad \text{and} \quad b = \lambda d \quad \text{for some scalar} \quad \lambda.$$

Proof. From $a = \lambda c$ and $b = \lambda d$ we derive, by cross multiplication, the identity $\lambda a d = \lambda b c$, whence $ad - bc = 0$ and the determinant is zero. Conversely, if $ad - bc = 0$, then we can deduce that $a = (b/d)c$ and $b = (a/c)d$ together with the identity $(b/d) = (a/c) = \lambda$, which implies that $a = \lambda c$ and $b = \lambda d$.

When the determinant is zero-valued one of two possibilities ensues. The first in when $e = \lambda f$. Then the two equations describe the same line and there is infinite number of solutions, with each solution corresponding to a point on the line. The second possibility is when $e \neq \lambda f$. Then the equations describe parallel lines and there are no solutions. Therefore, we say that the equations are inconsistent.

It is appropriate to end this section by giving a geometric interpretation of

(51)
$$\text{Det} \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} = a_1 b_2 - a_2 b_1.$$

This is simply the area enclosed by the parallelogram of forces which is formed by adding the vectors $a = (a_1, a_2)$ and $b = (b_1, b_2)$. The result can be established by subtracting triangles from the rectangle in the accompanying figure to show that the area of the shaded region is $\frac{1}{2}(a_1 b_2 - a_2 b_1)$. The shaded region comprises half of the area which is enclosed by the parallelogram of forces.
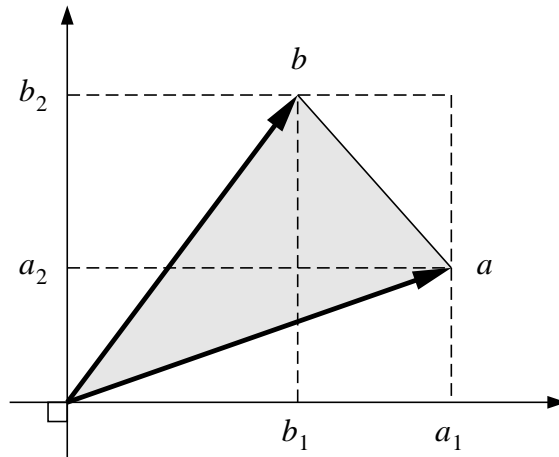
Figure 4. The determinant corresponds to the
area enclosed by the parallelogram of forces.