Matrix and Matrix Algebra

Wang Xiao

Shanghai University of Finance and Economics

March 9, 2026

Matrix as a notation

A linear equation in the n variables $x_1, ..., x_n$ is an equation of the form

$$a_1x_1 + a_2x_2 + ... + a_nx_n = b,$$

where the *coefficients* a_i are fixed real numbers and b is a fixed real numer. If we let $\mathbf{a} = (a_1, ..., a_n)^\top$ and $\mathbf{x} = (x_1, ..., x_n)^\top$, then we can write this equation in vector notation as

$$\mathbf{a} \cdot \mathbf{x} = b.$$

This is the equation of a hyperplane in $\mathbb{R}^n$, and a vector $\mathbf{x}$ solves the equation precisely when the point $\mathbf{x}$ lies on that hyperplane.

A system of m linear equations in n variables consists of m such equations:

$$\begin{aligned} a_{11}x_1 + ... + a_{1n}x_n &= b_1 \\ a_{21}x_1 + ... + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + ... + a_{mn}x_n &= b_m. \end{aligned}$$

Matrix, as a notation making describing and solving a general system of linear equations. The linear system can be written in vector notation:

$$\begin{aligned} \mathbf{A}_1 \cdot \mathbf{x} &= b_1 \\ \mathbf{A}_2 \cdot \mathbf{x} &= b_2 \\ &\vdots \\ \mathbf{A}_m \cdot \mathbf{x} &= b_m \end{aligned}$$

where $\mathbf{A}_i = (a_{i1}, ..., a_{in}) \in \mathbb{R}^n$.

To simplify the notation further, we use the $m \times n$ matrix

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

and the *column vectors*

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \quad \text{and } \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m,$$

and write the equations as

$$A\mathbf{x} = \mathbf{b},$$

where the multiplication on the left-hand side is defined to be

$$A\mathbf{x} := \begin{bmatrix} \mathbf{A}_1 \cdot \mathbf{x} \\ \vdots \\ \mathbf{A}_m \cdot \mathbf{x} \end{bmatrix}$$

Matrix as a mapping, a function, a transformation

We have learnt notations and concepts of functions and mappings, we are investigating linear mappings between vectors spaces.

Definition

A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear transformation* or *linear map* if it satisfies

i. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$;

ii. $T(c\mathbf{v}) = cT(\mathbf{v})$ for all $\mathbf{v} \in \mathbb{R}^n$ and scalars c.

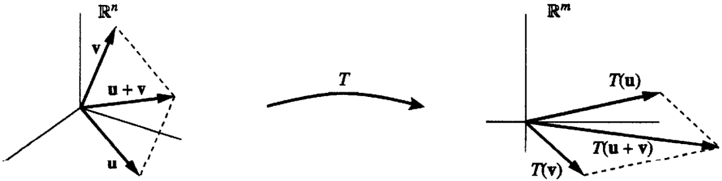
Consider linear combination of vectors, every vector in $\mathbb{R}^n$ can be written as a linear combination of the vectors

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

The vectors $\mathbf{e}_1, ..., \mathbf{e}_n$ are often called the *standard basis* vectors for $\mathbb{R}^n$. Obviously, given the vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{we have } \mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + ... + x_n\mathbf{x}_n.$$

Back to linear transformation, if we think visually of T as mapping $\mathbb{R}^n$ to $\mathbb{R}^m$, then we have a diagram like following figure:



The main point of the linearity properties is that the values of T on the standard basis vectors $\mathbf{e}_1, ..., \mathbf{e}_n$ completely determine the function T: for suppose $\mathbf{x} = x_1\mathbf{e}_1 + ... + x_n\mathbf{e}_n \in \mathbb{R}^n$; then

$$T(\mathbf{x}) = T(x_1\mathbf{e}_1 + ... + x_n\mathbf{e}_n) = x_1T(\mathbf{e}_1) + ... + x_nT(\mathbf{e}_n).$$

In particular, let

$$T(\mathbf{e}_j) = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \in \mathbb{R}^m;$$

then to T we can naturally associate the $m \times n$ array

$$A = \begin{bmatrix} a_1 1 & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix},$$

which we call the *standard matrix* for T, and we will often denote this by $[T]$. To emphasize: the jth column of A is the vector in $\mathbb{R}^m$ obtained by applying T to the jth standard basis vector $\mathbf{e}_j$.

Example

The most basic example of a linear map is the following. Fix $\mathbf{a} \in \mathbb{R}^n$, and define $T : \mathbb{R}^n \to \mathbb{R}$ by $T(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$. By linearity of dot product, we have

$$T(\mathbf{u}+\mathbf{v}) = a \cdot (\mathbf{u}+\mathbf{v}) = (\mathbf{a}\cdot\mathbf{u}) + (\mathbf{a}\cdot\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$$

and

$$T(c\mathbf{v}) = \mathbf{a} \cdot (c\mathbf{v}) = c(\mathbf{a}\cdot\mathbf{v}) = cT(\mathbf{v}),$$

moreover, it is easy to see that

$$\text{if } \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad \text{then } [T] = [a_1 \ a_2 \ \ldots \ a_n]$$

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map, and let A be its standard matrix. We want to define the product of the $m \times n$ matrix A with the vector $\mathbf{x} \in \mathbb{R}^n$ is such a way that the vector $T(\mathbf{x}) \in \mathbb{R}^m$ is equal to $A\mathbf{x}$. In accordance with the formula of $T(\mathbf{x})$ in standard basis, we have

$$A\mathbf{x} = T(\mathbf{x}) = \sum_{i=1}^{n} x_i T(\mathbf{e}_i) = \sum_{i=1}^{n} x_i \mathbf{a}_i,$$

where

$$\mathbf{a}_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix}, \ \mathbf{a}_2 = \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix}, \ldots, \ \mathbf{a}_n = \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} \in \mathbb{R}^m$$

are the column vectors of the matrix A. That is, $A\mathbf{x}$ is the linear combination of the vectors $\mathbf{a}_1, ..., \mathbf{a}_n$, weighted according to the coordinates of the vector $\mathbf{x}$.

Algebra of Matrices (Linear Mappings)

Denote by $\mathcal{M}_{m\times n}$ the set of all $m \times n$ matrices. In an obvious way this set can be identified with $\mathbb{R}^{mn}$ (how?). We can add $m \times n$ matrices and multiply them by scalars, just as we did with vectors. For future references (no formal definition needed), we call a matrix *square* if $m = n$. We refer to the entries $a_{ii}, i = 1, ..., n$, as *diagonal* entries. We call the square matrix a *diagonal matrix* if $a_{ij} = 0$ whenever $i \neq j$.

A square matrix all of whose entries below the diagonal are 0 is called *upper triangular*; one all of whose entries above the diagonal are 0 is called *lower triangular*.

If $S, T : \mathbb{R}^n \to \mathbb{R}^m$ are linear maps and $c \in \mathbb{R}$, then we obviously form the linear maps $cT : \mathbb{R}^n \to \mathbb{R}^m$ and $S + T : \mathbb{R}^n \to \mathbb{R}^m$, defined, respectively, by

$$(cT)(\mathbf{x}) \stackrel{\text{def}}{=} c(T(\mathbf{x}))$$

$$(S + T)(\mathbf{x}) \stackrel{\text{def}}{=} S(\mathbf{x}) + T(\mathbf{x}).$$

The corresponding algebraic manipulations with matrices are clear: if $A = [a_{ij}]$, then cA is the matrix whose entries are ca_{ij}:

$$cA = c \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} ca_{11} & \dots & ca_{1n} \\ ca_{21} & \dots & ca_{2n} \\ \dots & \ddots & \dots \\ ca_{m1} & \dots & ca_{mn} \end{bmatrix}$$

Given two matrices A and $B \in \mathcal{M}_{m\times n}$, we define their *sum* entry by entry.

$$A + B = \begin{bmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & \dots & a_{2n} + b_{2n} \\ \dots & \ddots & \dots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

Denote by O the zero matrix, any matrix added by O equals itself.

By the definition of matrix addition and scalar product (same for linear maps), we can check that scalar multiplication and addition satisfy the same properties as scalar multiplication and addition of vectors. We list them here for reference.

Proposition

Let $A, B, C \in \mathcal{M}_{m\times n}$ *and let* $c, d \in \mathbb{R}$.

1. $A + B = B + A$;
2. $(A + B) + C = A + (B + C)$;
3. $O + A = A$;
4. *There is a matrix* $-A$ *so that* $A + (-A) = O$;
5. $c(dA) = (cd)A$;
6. $c(A + B) = cA + cB$;
7. $(c + d)A = cA + dA$;
8. $1A = A$.

Recall that when $g(x)$ is the domain of f, we define $(f \circ g)(x) = f(g(x))$. So suppsoe we have linear maps $S : \mathbb{R}^{|} \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^m$. Then we can define $T \circ S : \mathbb{R}^p \to \mathbb{R}^m$ by $(T \circ S)(\mathbf{x}) = T(S(\mathbf{x}))$. Keep in mind that composition of functions is not commutative, so is linear maps. But composition is associative. Matrix multiplication can be defined to correspond to the composition of linear maps. Let A be the $m \times n$ matrix representing T and let B be the $n \times p$ matrix representing S. We expect that the $m \times p$ matrix C representing $T \circ S$ can be expressed in terms of A and B. The jth column of C is the vector $(T \circ S)(\mathbf{e}_j) \in \mathbb{R}^m$. Now let

$$T(S(\mathbf{e}_j)) = T \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix} = b_{1j}\mathbf{a}_1 + b_{2j}\mathbf{a}_2 + ... + b_{nj}\mathbf{a}_n,$$

where $\mathbf{a}_1, ..., \mathbf{a}_n$ are the column vectors of A.

Above formulation can be made into the definition:

Definition

Let A be an $m \times n$ matrix and B an $n \times p$ matrix. Their product AB is the $m \times p$ matrix whose jth column is the product of A with the jth column of B. That is, its ij-entry is

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj},$$

i.e., the dot product of the ith row vector of A and the jth column vector of A, both of which are vectors in $\mathbb{R}^n$. Graphically, this calculation is illustrated as follows

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & & \vdots & \\ \color{red}{a_{i1}} & \color{red}{a_{i1}} & \color{red}{\cdots} & \color{red}{a_{in}} \\ & & \vdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & & \color{red}{b_{1j}} & & b_{1p} \\ b_{21} & & \color{red}{b_{2j}} & & b_{2p} \\ \vdots & \cdots & \color{red}{\vdots} & \cdots & \vdots \\ b_{n1} & & \color{red}{b_{nj}} & & b_{np} \end{bmatrix}$$

$$= \begin{bmatrix} \cdots & \cdots & \cdots \\ & \vdots & \\ \cdots & (AB)_{ij} & \cdots \\ & \vdots & \\ \cdots & \cdots & \cdots \end{bmatrix}$$

The jth column of AB is the product of A with the jth column vector of B.

Proposition

Let A *and* A' *be* $m \times n$ *matrices; let* B *and* B' *be* $n \times p$ *matrices; let* C *be a* $p \times q$ *matrix, and let* c *be a scalar. Then*

1. $AI_n = A = I_m A$. *For this reason,* I_n *is called the* $n \times n$ *identity matrix.*
2. $(A + A')B = AB + A'B$ *and* $A(B + B') = AB + AB'$. *This is the distribution property of matrix multiplication over matrix addition.*
3. $(cA)B = c(AB) = A(cB)$.
4. $(AB)C = A(BC)$. *This is the associative property of matrix multiplication.*

Inverse and Transpose

Definition

Given an $m \times n$ matrix A, an $n \times m$ matrix B is called a right inverse of A if $AB = I_m$. Similarly, an $n \times m$ matrix C is called a left inverse of A if $CA = I_n$.

Example

Let $A = \begin{bmatrix} 2 & -1 & 0 \\ 1 & -2 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & -1 \\ -1 & -2 \\ -1 & -1 \end{bmatrix}$, then $AB = I_2$.

Suppose A is a square, $n \times n$, matrix with right inverse B and left inverse C, so that

$$AB = I_n = CA.$$

Then associativity of matrix multiplication implies

$$C = CI_n = C(AB) = (CA)B = I_nB = B.$$

That is, if A is both a left inverse and a right inverse, they must be equal.

Definition

Let A be an $n \times n$ matrix. We say A is invertible if there is an $n \times n$ matrix B so that

$$AB = BA = I_n.$$

We call B the inverse of the matrix A and denote this by $B = A^{-1}$.

Proposition

Suppose A and B are invertible $n \times n$ matrices. Then their product AB is invertible, and

$$(AB)^{-1} = B^{-1}A^{-1}.$$

The final matrix operation we will discuss is the *transpose*. When A is an $m \times n$ matrix with entries a_{ij}, the matrix $A^\top$ is the $n \times m$ matrix whose ij-entry is a_{ji}. We say a square matrix A is *symmetric* if $A^\top = A$ and *skew-symmetric* if $A^\top = -A$. These two important cases will be discussed in much detail in the future.

Proposition

Let A and A' be $m \times n$ matrices, let B be an $n \times p$ matrix, and c be a scalar. Then

1. $(A^\top)^\top = A$;
2. $(cA)^\top = cA^\top$;
3. $(A + A')^\top = A^\top + A'^\top$;
4. $(AB)^\top = B^\top A^\top$.

Determinants of 2 and 3×3 matrices

Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in $\mathbb{R}^2$ and consider the parallelogram P they span. The area of P is nonzero if and only if $\mathbf{x}$ and $\mathbf{y}$ are not collinear. The area of P is bh where $b = \|\mathbf{x}\|$ and $h = \|\mathbf{y}\| \sin\theta$, and we can calculate $\sin\theta$ from the formula

$$\cos\theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \, \|\mathbf{y}\|}$$

Recall the ex on plane geometry:

Example

If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, set $\rho(\mathbf{x}) = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$

(a) Check that $\rho(\mathbf{x})$ is orthogonal to $\mathbf{x}$. $\rho(\mathbf{x})$ is obtained by rotating an angle $\frac{\pi}{2}$ counterclockwise.

(b) Give $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, prove that $\mathbf{x} \cdot \rho(\mathbf{y}) = -\rho(\mathbf{x}) \cdot \mathbf{y}$

If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1, y_2 \end{bmatrix}$, then we have from above exercise

$$\text{area}(P) = \rho(\mathbf{x}) \cdot \mathbf{y} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = x_1 y_2 - x_2 y_1.$$

The *signed area* of the parallelogram P to be the area of P when one turns counterclockwise from $\mathbf{x}$ to $\mathbf{y}$ and to be *negative* the area of P when one turns clockwise from $\mathbf{x}$ to $\mathbf{y}$. Then we have

$$\text{signed are}(P) = x_1 y_2 - x_2 y_1.$$

We consider the function

$$\mathcal{D}(\mathbf{x}, \mathbf{y}) = x_1 y_2 - x_2 y_1,$$

which is the function associates to each ordered pair of vectors $\mathbf{x}, \mathbf{y}$ the signed area of the parallelogram they span.

Some properties of $\mathcal{D}$.

Property 1

If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, then $\mathcal{D}(\mathbf{y}, \mathbf{x}) = -\mathcal{D}(\mathbf{x}, \mathbf{y})$.
Algebraically, this is $y_1 x_2 - y_2 x_1 = -(x_1 y_2 - x_2 y_1)$.

Property 2

$\mathcal{D}(c\mathbf{x}, \mathbf{y}) = c\mathcal{D}(\mathbf{x}, \mathbf{y}) = \mathcal{D}(\mathbf{x}, c\mathbf{y})$.
Check $(cx_1)y_2 - (cx_2)y_1 = c(x_1 y_2 - x_2 y_1)$.

Property 3

$\mathcal{D}(\mathbf{x} + \mathbf{y}, \mathbf{z}) = \mathcal{D}(\mathbf{x}, \mathbf{z}) + \mathcal{D}(\mathbf{y}, \mathbf{z})$ and $\mathcal{D}(\mathbf{x}, \mathbf{y} + \mathbf{z}) = \mathcal{D}(\mathbf{x}, \mathbf{y}\mathcal{D}(\mathbf{x}, \mathbf{z})$.
Check that

$$(x_1 + y_1)z_2 - (x_2 + y_2)z_1 = (x_1 z_2 - x_2 z_1) + (y_1 z_2 - y_2 z_1)$$

Property 4

For the standard basis, $\mathcal{D}(\mathbf{e}_1, \mathbf{e}_2) = 1$.

The expression $\mathcal{D}$ is a 2×2 *determinant*, written $|\mathbf{x} \quad \mathbf{y}|$. Indeed, given a 2×2 matrix A with column vectors $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^2$, we define

$$\det A = \mathcal{D}(\mathbf{a}_1, \mathbf{a}_2) = |\mathbf{a}_1 \quad \mathbf{a}_2|.$$

Given three vectors,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \quad \text{and } \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix},$$

we define

$$\mathcal{D}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \begin{vmatrix} \vert & \vert & \vert \\ \mathbf{x} & \mathbf{y} & \mathbf{z} \\ \vert & \vert & \vert \end{vmatrix} = x_1 \begin{vmatrix} y_2 & z_2 \\ y_3 & z_3 \end{vmatrix} - x_3 \begin{vmatrix} y_1 & z_1 \\ y_3 & z_3 \end{vmatrix} + x_3 \begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix}$$

Comparing to function $\mathcal{D}$ of two vectors, $\mathcal{D}$ of three vectors has similar properties. Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3$ and c is a scalar, then

$$\mathcal{D}(\mathbf{x}, \mathbf{z}, \mathbf{y}) = -\mathcal{D}(\mathbf{x}, \mathbf{y}, \mathbf{z}),$$

$$\mathcal{D}(\mathbf{x}, c\mathbf{y}, \mathbf{z}) = c\mathcal{D}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathcal{D}(\mathbf{x}, \mathbf{y}, c\mathbf{z}),$$

$$\mathcal{D}(\mathbf{x}, \mathbf{y} + \mathbf{w}, \mathbf{z}) = \mathcal{D}(\mathbf{x}, \mathbf{y}, \mathbf{z}) + \mathcal{D}(\mathbf{x}, \mathbf{w}, \mathbf{z}).$$

To verify that $\mathcal{D}(\mathbf{z}, \mathbf{x}, \mathbf{z}) = -\mathcal{D}(\mathbf{x}, \mathbf{y}, \mathbf{z})$:

$$\begin{aligned}
\mathcal{D}(\mathbf{y}, \mathbf{x}, \mathbf{z}) &= y_1 \begin{vmatrix} x_2 & z_2 \\ x_3 & z_3 \end{vmatrix} - y_2 \begin{vmatrix} x_1 & z_1 \\ x_3 & z_3 \end{vmatrix} + y_3 \begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} \\
&= y_1(x_2 z_3 - x_3 z_2) + y_2(x_3 z_1 - x_1 z_3) + y_3(x_1 z_2 - x_2 z_1) \\
&= -x_1(y_2 z_3 - y_3 z_2) + x_2(y_1 z_3 - y_3 z_1) - x_3(y_1 z_2 - y_2 z_1) \\
&= -x_1 \begin{vmatrix} y_2 & z_2 \\ y_3 & z_3 \end{vmatrix} + x_2 \begin{vmatrix} y_1 & z_1 \\ y_3 & z_3 \end{vmatrix} - x_3 \begin{vmatrix} y_1 & z_2 \\ y_2 & z_2 \end{vmatrix} \\
&= -\mathcal{D}(\mathbf{x}, \mathbf{y}, \mathbf{z}).
\end{aligned}$$

Given two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$, define a vector, called their *cross product*, by

$$\begin{aligned}\mathbf{x} \times \mathbf{y} &= (x_2y_3 - x_3y_2)\mathbf{e}_1 + (x_3y_1 - x_1y_3)\mathbf{e}_2 + (x_1y_2 - x_2y_1)\mathbf{e}_3 \\ &= \begin{vmatrix} \mathbf{e}_1 & x_1 & y_1 \\ \mathbf{e}_2 & x_2 & y_2 \\ \mathbf{e}_3 & x_3 & y_3 \end{vmatrix},\end{aligned}$$

where the latter is to be interpreted "formally". The geometric meaning of the cross product is the following proposition.

Proposition

The cross product $\mathbf{x} \times \mathbf{y}$ *of two vectors is orthogonal to both* $\mathbf{x}$ *and* $\mathbf{y}$ *and* $\|\mathbf{x} \times \mathbf{y}\|$ *is the area of the parallelogram* P *spanned by* $\mathbf{x}$ *and* $\mathbf{y}$*. Moreover, when* $\mathbf{x}$ *and* $\mathbf{y}$ *are nonparallel, the vectors* $\mathbf{x}, \mathbf{y}, \mathbf{x} \times \mathbf{y}$ *determine a parallelepiped of positive signed volume.*

Proof. Formula for the cross product gives

$$\mathbf{z} \cdot (\mathbf{x} \times \mathbf{y}) = \mathcal{D}(\mathbf{z}, \mathbf{x}, \mathbf{y}).$$

So $\mathbf{x} \cdot (\mathbf{x} \times \mathbf{y}) = 0$. Since $\mathcal{D}(\mathbf{x}, \mathbf{y}, \mathbf{x} \times \mathbf{y})$ is the signed volume of the parallelepiped spanned by $\mathbf{x}, \mathbf{y}$, and $\mathbf{x} \times \mathbf{y}$, $\mathbf{x} \times \mathbf{y}$ is orthogonal to the plane spanned by $\mathbf{x}$ and $\mathbf{y}$, that volume is the product of the area of P and $\|\mathbf{x} \times \mathbf{y}\|$. On the other hand,

$$\mathcal{D}(\mathbf{x}, \mathbf{y}, \mathbf{x} \times \mathbf{y}) = \mathcal{D}(\mathbf{x} \times \mathbf{y}, \mathbf{x}, \mathbf{y}) = (\mathbf{x} \times \mathbf{y}) \cdot (\mathbf{x} \times \mathbf{y}) = \|\mathbf{x} \times \mathbf{y}\|^2.$$

We can infer that

$$\|\mathbf{x} \times \mathbf{y}\| = \text{area}(P).$$