Vector Algebra

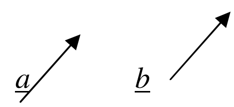
1.   Review of Basic Ideas

In engineering and science, physical quantities which are completely specified by their magnitude (size) are known as scalars. Examples are: mass, temperature, volume, resistance, charge, voltage, current, etc.

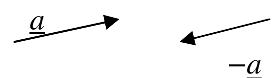
Other quantities possess both magnitude and direction and may be represented geometrically by directed line segments known as vectors. The length of the line is known as the magnitude of the vector and its direction is the direction of the vector. Examples of vector quantities are: velocity, acceleration, force, electric field, magnetic field etc and will be denoted by $\underline{v}$, $\underline{a}$, $\underline{F}$, $\underline{E}$, $\underline{B}$, etc.

1. Two vectors $\underline{a}$ and $\underline{b}$ are equal if they have the same magnitude and direction irrespective of their initial points. We write

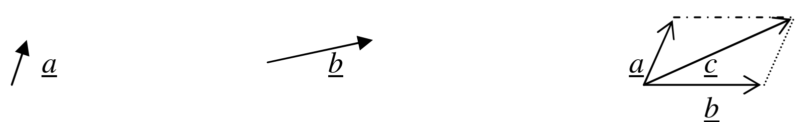
$$\underline{a} = \underline{b}$$



2. A vector having the same magnitude as $\underline{a}$ but the opposite direction is denoted by $-\underline{a}$.

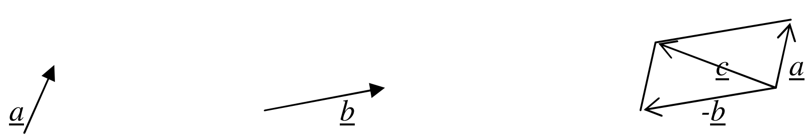


3. Geometrically the sum of two vectors is given by the parallelogram law



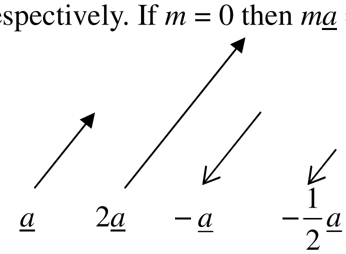
$$\underline{c} = \underline{a} + \underline{b}$$

4. The difference of two vectors $\underline{a}$ and $\underline{b}$, represented by $\underline{c} = \underline{a} - \underline{b}$ is defined as $\underline{c} = \underline{a} + (-\underline{b})$



5. If $\underline{a} = \underline{b}$ then $\underline{a} - \underline{b}$ is the zero vector denoted by $\underline{0}$. This has magnitude 0 but no direction.

6. Multiplication of $\underline{a}$ by a scalar, m, produces a vector $m\underline{a}$ with magnitude m times that of $\underline{a}$ and direction the same as or opposite to that of $\underline{a}$ according to whether m is positive or negative respectively. If $m = 0$ then $m\underline{a} = \underline{0}$.



7. Unit vectors are vectors with magnitude 1. If $\underline{a}$ is any vector then we usually denote its magnitude by $|\underline{a}|$. A unit vector with the same direction as $\underline{a}$ will be $\frac{\underline{a}}{|\underline{a}|}$.

2. Components of a Vector

In a rectangular coordinate system in $3-\mathrm{D}$ Euclidean space R^3, orthogonal (perpendicular) unit vectors in the directions of the positive x, y and z axis are denoted by $i=\begin{pmatrix}1\\0\\0\end{pmatrix}$, $j=\begin{pmatrix}0\\1\\0\end{pmatrix}$ and $\underline{k}=\begin{pmatrix}0\\0\\1\end{pmatrix}$ respectively. The vector from the origin O to a point P is known as the position vector of P. If P has Cartesian coordinates $\begin{pmatrix}x\\y\\z\end{pmatrix}$, then the position vector of P, $\underline{r}$, may be written as,

$$\overrightarrow{OP}=\underline{r}=\begin{pmatrix}x\\y\\z\end{pmatrix}=x\begin{pmatrix}1\\0\\0\end{pmatrix}+y\begin{pmatrix}0\\1\\0\end{pmatrix}+z\begin{pmatrix}0\\0\\1\end{pmatrix}=x\underline{i}+y\underline{j}+z\underline{k}$$

$$\underline{r}=\begin{pmatrix}x\\y\\z\end{pmatrix}=x\begin{pmatrix}1\\0\\0\end{pmatrix}+y\begin{pmatrix}0\\1\\0\end{pmatrix}+z\begin{pmatrix}0\\0\\1\end{pmatrix}=xi+yj+zk\,.$$

x, y, z are know as the components or coordinates of $\underline{r}$ with respect to the vectors $\underline{i}$, $\underline{j}$, and $\underline{k}$.

If $P:\begin{pmatrix}x_1\\y_1\\z_1\end{pmatrix}$ and $Q:\begin{pmatrix}x_2\\y_2\\z_2\end{pmatrix}$ are two points, the vector from P to Q, $\overrightarrow{PQ}$ will be

$$\overrightarrow{PQ}=\underline{a}=a_1\underline{i}+a_2\underline{j}+a_3\underline{k}=\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix},$$ where the components of $\underline{a}$ are

$a_1=x_2-x_1$, $a_2=y_2-y_1$ and $a_3=z_2-z_1$.

Note that the ordered triple of components of a vector is unique with respect to a given coordinate system.

If $\underline{a}=\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix}$ and $\underline{b}=\begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix}$, in terms of components we have:

Equality :

$\underline{a} = \underline{b} \quad iff \quad a_1 = b_1, a_2 = b_2, a_3 = b_3$

Addition:

$$\underline{a} + \underline{b} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{pmatrix} = (a_1 + b_1)\underline{i} + (a_2 + b_2)\underline{j} + (a_3 + b_3)\underline{k}$$

Scalar Multiplication:

$$m\underline{a} = \begin{pmatrix} ma_1 \\ ma_2 \\ ma_3 \end{pmatrix} = ma_1\underline{i} + ma_2\underline{j} + ma_3\underline{k}$$

Zero Vector :

$$\underline{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Magnitude :

$|\underline{a}| = \sqrt{a_1^{\,2} + a_2^{\,2} + a_3^{\,3}}$ (Pythagoras)

Unit Vector:

$$\frac{\underline{a}}{|\underline{a}|} = \frac{1}{\sqrt{a_1^{\,2} + a_2^{\,2} + a_3^{\,2}}} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \frac{a_1\underline{i} + a_2\underline{j} + a_3\underline{k}}{\sqrt{a_1^{\,2} + a_2^{\,2} + a_3^{\,2}}}$$

Notice that the above are also applicable to the *n*-component vectors $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in R^n,\ n \geq 1$**.**

Example

The vector $\underline{a}$ from $P : \begin{pmatrix} 3 \\ -2 \\ 1 \end{pmatrix}$ to $Q : \begin{pmatrix} 1 \\ 2 \\ -4 \end{pmatrix}$ has components

$a_1 = 1 - 3 = -2,\ a_2 = 2 - (-2) = 4,\ a_3 = -4 - 1 = -5$. Hence

$$\underline{a} = \begin{pmatrix} -2 \\ 4 \\ -5 \end{pmatrix} = -2\underline{i} + 4\underline{j} - 5\underline{k}, \ |\underline{a}| = \sqrt{(-2)^2 + 4^2 + (-5)^2} = \sqrt{45}$$

And a unit vector in the direction of $\underline{a}$ is $\dfrac{1}{\sqrt{45}}\begin{pmatrix} -2 \\ 4 \\ -5 \end{pmatrix}$

If $\underline{a}$ has the initial point $R:\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, then its terminal point is $S:\begin{pmatrix} -1 \\ 6 \\ -2 \end{pmatrix}$.

<u>Example</u>

The vector $\underline{a}$ from $P:\begin{pmatrix} 1 \\ 2 \\ -2 \\ 1 \end{pmatrix}$ to $Q:\begin{pmatrix} 1 \\ 1 \\ -4 \\ 2 \end{pmatrix}$ has components

$a_1 = 1 - 1 = 0, \ a_2 = 1 - 2 = -1, \ a_3 = -4 - (-2) = -2, \ a_3 = 2 - 1 = 1$. Hence

$$\underline{a} = \begin{pmatrix} 0 \\ -1 \\ -2 \\ 1 \end{pmatrix}, \ |\underline{a}| = \sqrt{0^2 + (-1)^2 + (-2)^2 + 1^2} = \sqrt{6}$$

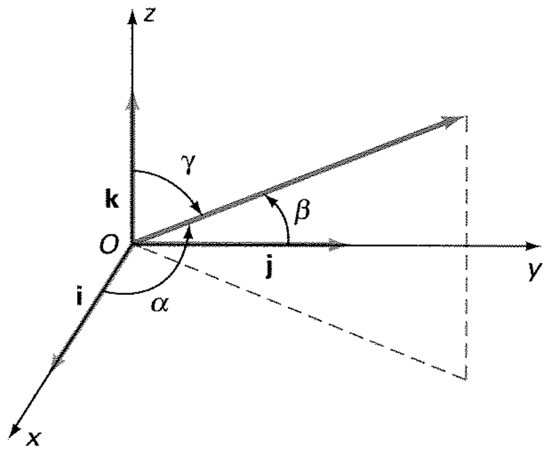
And a unit vector in the direction of $\underline{a}$ is $\dfrac{1}{\sqrt{6}}\begin{pmatrix} 0 \\ -1 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}$

If $\underline{a}$ has the terminal point $R:\begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix}$, then its initial point S is:

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix} - S = \begin{pmatrix} 0 \\ -1 \\ -2 \\ 1 \end{pmatrix} \Rightarrow S = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ -1 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 5 \\ 0 \end{pmatrix}.$$

Direction Cosines

If $\underline{r} = x\underline{i} + y\underline{j} + z\underline{k}$, the direction of $\underline{r}$ may be specified by the cosines of the angles made by $\underline{r}$ with the 3 coordinate axes.



$$l = \cos\alpha = \frac{x}{|\underline{r}|}$$

$$m = \cos\beta = \frac{y}{|\underline{r}|}$$

$$n = \cos\gamma = \frac{z}{|\underline{r}|}$$

l, m, and n are known as the direction cosines of $\underline{r}$, $l\underline{i} + m\underline{j} + n\underline{k}$ is a unit vector along $\underline{r}$ and

$\underline{r} = |\underline{r}|(l\underline{i} + m\underline{j} + n\underline{k})$

Example

Let $\overrightarrow{OP}$ be $\begin{pmatrix} 3 \\ 2 \\ 6 \end{pmatrix} = 3\underline{i} + 2\underline{j} + 6\underline{k}$, then $|\underline{r}| = 7,\ l = 3/7,\ m = 2/7,\ n = 6/7$ and

$\alpha = \cos^{-1}(3/7),\ \beta = \cos^{-1}(2/7),\ \gamma = \cos^{-1}(6/7).$

If $\underline{a}, \underline{b}, \underline{c} \in R^n$ are vectors and m, n are scalars (real numbers), then we have

1. $\underline{a} + \underline{b} = \underline{b} + \underline{a}$ Commutative law of vector addition
2. $\underline{a} + (\underline{b} + \underline{c}) = (\underline{a} + \underline{b}) + \underline{c}$ Associative law of vector addition
3. $\underline{a} + \underline{0} = \underline{a}$ Existence of $\underline{0}$ as an additive vector identity
4. $\underline{a} + (-\underline{a}) = \underline{0}$ Existence of additive inverses
5. $m(\underline{a} + \underline{b}) = m\underline{a} + m\underline{b}$ Scalar distribution over vector addition
6. $(m + n)\underline{a} = m\underline{a} + n\underline{a}$ Vector distribution over scalar addition
7. $(mn)\underline{a} = m(n\underline{a})$ Associative law for scalar multiplication
8. $1\underline{a} = \underline{a}$ Multiplicative scalar identity

3. Vector Products

Let $\underline{a}$ and $\underline{b}$ be two 3-component vectors, their dot product or scalar product, written $\underline{a} \bullet \underline{b}$, is defined as,

$$\underline{a} \bullet \underline{b} = \begin{cases} |\underline{a}||\underline{b}|\cos\theta & \text{if } \underline{a} \neq 0,\ \underline{b} \neq 0 \\ 0 & \text{otherwise} \end{cases}$$, where θ is the angle between $\underline{a}$ and $\underline{b}$

Example

Given non-zero position vectors $\underline{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ and $\underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$, show that $\underline{a} \bullet \underline{b} = a_1 b_1 + a_2 b_2 + a_3 b_3$

Proof:

According to Cosine Law, we have $|\underline{b} - \underline{a}|^2 = |\underline{a}|^2 + |\underline{b}|^2 - 2|\underline{a}||\underline{b}|\cos\theta$.

$$|\underline{b} - \underline{a}|^2 = (b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2,\quad |\underline{a}|^2 = a_1^2 + a_2^2 + a_3^2,\quad |\underline{b}|^2 = b_1^2 + b_2^2 + b_3^2$$

Then we have

$$|\underline{b} - \underline{a}|^2 = |\underline{a}|^2 + |\underline{b}|^2 - 2|\underline{a}||\underline{b}|\cos\theta$$
$$\Rightarrow (b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2 = a_1^2 + a_2^2 + a_3^2 + b_1^2 + b_2^2 + b_3^2 - 2|\underline{a}||\underline{b}|\cos\theta$$
$$\Rightarrow 2a_1 b_1 + 2a_2 b_2 + 2a_3 b_3 = 2|\underline{a}||\underline{b}|\cos\theta$$
$$\Rightarrow \underline{a} \bullet \underline{b} = |\underline{a}||\underline{b}|\cos\theta = a_1 b_1 + a_2 b_2 + a_3 b_3$$

Accordingly, we can define the dot product of two n-component vectors $\underline{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in R^n$ as

$$\underline{a} \bullet \underline{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

Properties of the dot product $\underline{a} \bullet \underline{b}$, $\underline{a}, \underline{b} \in R^n$

(i) The result is a scalar

(ii) $\underline{a} \bullet \underline{b}$ is zero if $\underline{a} = \underline{0}$ or $\underline{b} = \underline{0}$ or $\underline{a}$ and $\underline{b}$ are perpendicular (orthogonal)

(iii) $|\underline{a}| = \sqrt{\underline{a} \bullet \underline{a}}$

(iv) $\underline{a} \bullet \underline{b} = \underline{b} \bullet \underline{a}$ (symmetry)

(v) $(m\underline{a} + n\underline{b}) \bullet \underline{c} = m(\underline{a}, \underline{c}) + n(\underline{b}, \underline{c})$ $\forall \underline{a}, \underline{b} \in R^3$ and $m, n \in R$ (Linearity)

(vi) $\underline{a} \bullet \underline{a} \geq 0$ and $\underline{a} \bullet \underline{a} = 0$ iff $\underline{a} = \underline{0}$ (Positive definiteness)

(vii) $|\underline{a} \bullet \underline{b}| \leq |\underline{a}||\underline{b}|$ (Schwartz inequality)

(viii) $|\underline{a}+\underline{b}| \le |\underline{a}|+|\underline{b}|$ (Triangle inequality)

(ix) $|\underline{a}+\underline{b}|^2 + |\underline{a}-\underline{b}|^2 = 2\left(|\underline{a}|^2+|\underline{b}|^2\right)$

We observe that $\underline{i}\bullet\underline{i}=\underline{j}\bullet\underline{j}=\underline{k}\bullet\underline{k}=1$ and $\underline{i}\bullet\underline{j}=\underline{j}\bullet\underline{k}=\underline{k}\bullet\underline{i}=0$

Example

Let $\underline{a} = 5\underline{i} + 4\underline{j} + 2\underline{k}$ and $\underline{b} = 4\underline{i} - 5\underline{j} + 3\underline{k}$, find $\underline{a}\bullet\underline{b}$ and the angle between the vectors.

Solution:

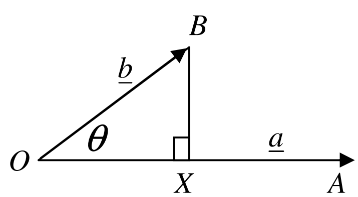
$\underline{a}\bullet\underline{b} = (5\times4)+(4\times(-5))+(2\times3)=6$. But $|\underline{a}|=\sqrt{45},\ |\underline{b}|=\sqrt{50}$, hence

$$\cos\theta = \frac{\underline{a}\bullet\underline{b}}{|\underline{a}||\underline{b}|} = \frac{6}{\sqrt{45}\times\sqrt{50}} = \frac{2}{5\sqrt{10}} \Rightarrow \theta=\arccos\left(\frac{2}{5\sqrt{10}}\right)$$

Example

Given two vectors $\underline{a}=\overrightarrow{OA},\ \underline{b}=\overrightarrow{OB}$, find the projection $\text{pro}_{\underline{a}}\underline{b}$ of $\underline{b}$ in the direction of $\underline{a}$ and the coefficient of $\text{pro}_{\underline{a}}\underline{b}$ (i.e. the coefficient of the projection of $\underline{b}$ in the direction of $\underline{a}$).

Solution:



$$\begin{aligned}\underline{a}\cdot\underline{b} &= |\underline{a}||\underline{b}|\cos\theta \\ &= OA\times OB\cos\theta \\ &= OA\times OX\end{aligned}$$

Then

$$\frac{\underline{a}\bullet\underline{b}}{|\underline{a}|} = OB\cos\theta = OX$$

The projection $\text{pro}_{\underline{a}}\underline{b}$ of $\underline{b}$ in the direction of $\underline{a}$ is:

$$(OB\cos\theta)\frac{\underline{a}}{|\underline{a}|} = \frac{\underline{a}\bullet\underline{b}}{|\underline{a}|}\frac{\underline{a}}{|\underline{a}|} = \frac{\underline{a}\bullet\underline{b}}{|\underline{a}|^2}\underline{a} = \frac{\underline{a}\bullet\underline{b}}{\underline{a}\bullet\underline{a}}\underline{a}.$$

the coefficient of $\text{pro}_{\underline{a}}\underline{b}$ (i.e. the coefficient of the projection of $\underline{b}$ in the direction of $\underline{a}$) is:

$OX = |\underline{b}|\cos\theta = \dfrac{1}{|\underline{a}|}\underline{a}\bullet\underline{b} = \underline{b}\bullet$ (unit vector in $\underline{a}$ direction)

Notice that the <u>coefficient</u> of $\text{pro}_{\underline{a}}\underline{b}$ (i.e. the coefficient of the projection of $\underline{b}$ in the direction of $\underline{a}$) can be negative if the angle θ between $\underline{a}$, $\underline{b}$ is an obtuse angle.

<u>Example</u>

A force $\underline{F} = 2\underline{i} + 3\underline{j} + \underline{k}$ acts on a particle which is displaced though $\underline{d} = \underline{i} - \underline{j} + 2\underline{k}$. Find the coefficient of $\text{pro}_{\underline{d}}\underline{F}$ (i.e. the coefficient of the projection of $\underline{F}$ in the direction of $\underline{d}$) and the work done by the force.

Solution:

Coefficient of $\text{pro}_{\underline{d}}\underline{F}$ is $\underline{F} \bullet \dfrac{\underline{d}}{|\underline{d}|} = \dfrac{2-3+2}{\sqrt{6}} = \dfrac{1}{\sqrt{6}}$

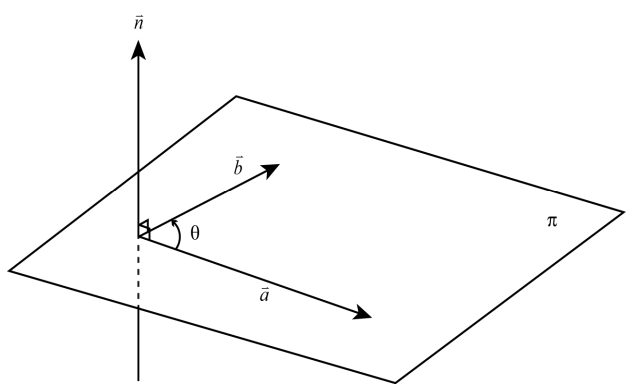
Work done by $\underline{F}$ = (Coefficient of $\underline{F}$ in the direction of $\underline{d}$) multiplied by $|\underline{d}|$

$$= \left(\underline{F} \cdot \frac{\underline{d}}{|\underline{d}|}\right)|\underline{d}| = \underline{F} \cdot \underline{d} = 1$$

Let $\underline{a}$, $\underline{b} \in R^3$ be two three component vectors, the <u>vector product</u> or <u>cross product</u> of $\underline{a}$ and $\underline{b}$, written $\underline{a} \times \underline{b}$, is defined as:

$$\underline{a} \times \underline{b} = \begin{cases} |\underline{a}||\underline{b}|\sin\theta\underline{v} & \text{if } \underline{a} \neq 0, \underline{b} \neq 0 \\ \underline{0} & \text{otherwise} \end{cases}$$

, where v a unit vector such that $\underline{a}$, $\underline{b}$, $\underline{v}$ form a right-handed triple, and θ the angle between $\underline{a}$, $\underline{b}$.



Notice that cross product is defined only for 3-component vectors.

Properties of the cross product:

(i) The result is a vector and $\underline{a} \times \underline{b}$ is zero *iff* $\underline{a} = \underline{0}$ or $\underline{b} = \underline{0}$ or $\underline{a}$ and $\underline{b}$ are parallel.

(ii) $\underline{a} \times \underline{a} = \underline{0}$

(iii) $\underline{a} \times \underline{b} = -\underline{b} \times \underline{a}$

(iv) $|\underline{a} \times \underline{b}| = |\underline{a}||\underline{b}|\sin\theta$= area of parallelogram with sides $\underline{a}$ and $\underline{b}$.

$$\underline{a} \times \underline{b} = \begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \underline{i}(a_2b_3\text{-}a_3b_2) + \underline{j}(a_3b_1\text{-}a_1b_3) + \underline{k}(a_1b_2\text{-}a_2b_1)$$

(Prove using results (vii) and (viii) below)

(v) $\underline{a} \times (\underline{b} + \underline{c}) = (\underline{a} \times \underline{b}) + (\underline{a} \times \underline{c})$

(vi) $m\ (\underline{a} \times \underline{b}) = (m\underline{a} \times \underline{b}) = (\underline{a} \times m\underline{b}) = (\underline{a} \times \underline{b})m$

(vii) $\underline{i} \times \underline{i} = \underline{j} \times \underline{j} = \underline{k} \times \underline{k} = \underline{0}$, $\quad \underline{i} \times \underline{j} = \underline{k},\ \underline{j} \times \underline{k} = \underline{i},\ \underline{k} \times \underline{i} = \underline{j}$

<u>Example</u>

If $\underline{a} = 5\underline{i} + 4\underline{j} + 2\underline{k}$ and $\underline{b} = 4\underline{i} - 5\underline{j} + 3\underline{k}$, find $\underline{a} \times \underline{b}$ and a unit vector perpendicular to both $\underline{a}$ and $\underline{b}$.

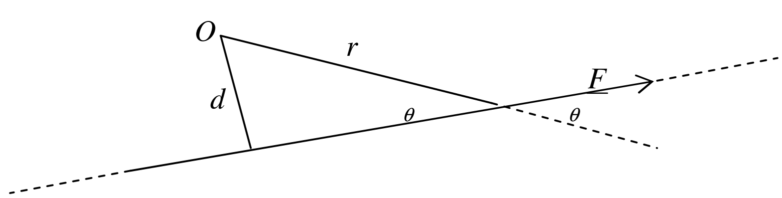
$$\underline{a} \times \underline{b} = \begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ 5 & 4 & 2 \\ 4 & -5 & 3 \end{vmatrix} = \underline{i}\begin{vmatrix} 4 & 2 \\ -5 & 3 \end{vmatrix} - \underline{j}\begin{vmatrix} 5 & 2 \\ 4 & 3 \end{vmatrix} + \underline{k}\begin{vmatrix} 5 & 4 \\ 4 & -5 \end{vmatrix}$$

$$= 22\underline{i} + \ 7\underline{j} + 41\underline{k}$$

A unit vector is $\pm \dfrac{\underline{a} \times \underline{b}}{|\underline{a} \times \underline{b}|} = \pm \dfrac{1}{\sqrt{2214}}(22\underline{i} - 7\underline{j} - 41\underline{k})$

<u>Example</u> – Moment of a force

In mechanics the moment, m, of a force $\underline{F}$ about a point 0 is defined as the magnitude of $\underline{F}$ times the perpendicular distance (d) from 0 to the line of action, L, of $\underline{F}$. Let $\underline{r}$ be the vector from 0 to any point on L.



$m = |\underline{F}||\underline{r}|\sin\theta$

$m = |\underline{r} \times \underline{F}|$

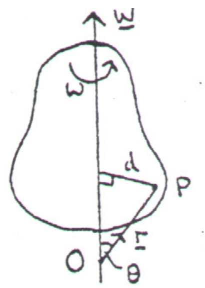
The vector $\underline{m} = \underline{r} \times \underline{F}$ is called the vector moment of $\underline{F}$ about 0. Its direction is along the axis about which $\underline{F}$ has a tendency to produce a rotation.

Example – Magnetic Field

The force $\underline{F}$ experienced by a point charge q moving with velocity $\underline{v}$ in a magnetic field of flux density $\underline{B}$ is given by $\underline{F} = q\,\underline{v} \times \underline{B}$

Example – Rotation

Consider a rigid body rotating with angular speed w about an axis. Let $\underline{w}$ be the vector with magnitude w and direction along the axis such that the rotation of the body appears clockwise looking along this direction.



Let P be any point in the body and d its distance from the axis. Then P has speed wd.

Let P have position vector $\underline{r}$ with respect to some point 0 on the axis. Then

$$d = |\underline{r}|\sin\theta$$

$$wd = |\underline{w}||\underline{r}|\sin\theta = |\underline{w}\times\underline{r}|$$

And the velocity $\underline{v}$ of P is given by $\underline{v} = \underline{w}\times\underline{r}$.

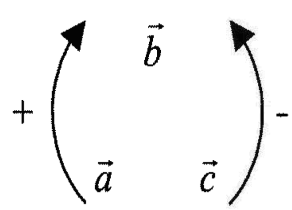
Products of three or more vectors follow naturally:

Consider the triple scalar product $\underline{a}\bullet(\underline{b}\times\underline{c})$. We observe that $\underline{a}\bullet(\underline{b}\times\underline{c})$ is a scalar.

Properties:

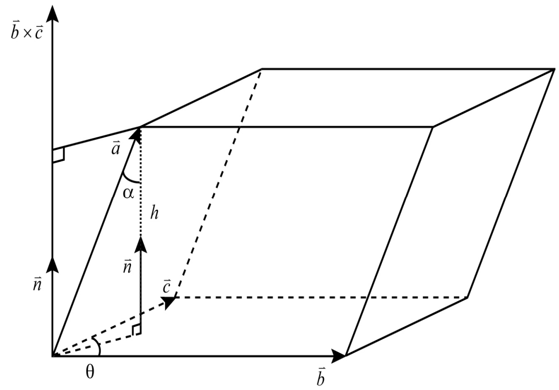
(i) $$\underline{a}\bullet(\underline{b}\times\underline{c}) = (a_1\underline{i} + a_2\underline{j} + a_3\underline{k})\bullet\begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

(ii) $\underline{a}\bullet(\underline{b}\times\underline{c}) = (\underline{a}\times\underline{b})\bullet\underline{c}$ - property of determinant and the triple scalar product is usually written $(\underline{a}, \underline{b}, \underline{c})$



(iii) $(a,\ b,\ c) = -(b,\ a,\ c) = (a,\ b,\ c) = (b,\ c,\ a) = (c,\ a,\ b)$

(iv) Geometrically the absolute value of $(\underline{a}, \underline{b}, \underline{c})$ equals the volume of the parallelepiped with $\underline{a}$, $\underline{b}$ and $\underline{c}$ as adjacent edges.



Proof:

Observe that $\vec{b}\times\vec{c}=\left(|\vec{b}||\vec{c}|\sin\theta\right)\vec{n}$, where $\vec{n}$ is a unit vector in the same direction as $\vec{b}\times\vec{c}$ such that $\vec{b},\vec{c},\vec{n}$ form a right-hand triple.

Area of the base of the parallelepiped $=|\vec{b}\times\vec{c}|=|\vec{b}||\vec{c}|\sin\theta\ (>0)$.

Perpendicular height of the parallelepiped $= h=|\vec{a}|\cos\alpha=|\vec{a}\cdot\vec{n}|$.

$\therefore$ Volume of the parallelepiped = Base area of the parallelepiped$\times$its height

$$= |\vec{b}\times\vec{c}||\vec{a}\cdot\vec{n}|=\left(|\vec{b}||\vec{c}|\sin\theta\right)|\vec{a}\cdot\vec{n}|=\left|\vec{a}\cdot\left(|\vec{b}||\vec{c}|\sin\theta\right)\vec{n}\right|=|\vec{a}\cdot\vec{b}\times\vec{c}|.$$

(v) Three vectors are coplanar *iff* their triple scalar product is zero.

Consider the triple vector product $\underline{a}\times(\underline{b}\times\underline{c})$. We observe that $\underline{a}\times(\underline{b}\times\underline{c})$ is a vector.

Properties:

(i) Note that $\underline{a}\times(\underline{b}\times\underline{c})\neq(\underline{a}\times\underline{b})\times\underline{c}$ in general e.g. $\underline{i}\times(\underline{j}\times\underline{j})=\underline{0}$ whereas $(\underline{i}\times\underline{j})\times\underline{j}=-\underline{i}$

(ii) $\underline{a}\times(\underline{b}\times\underline{c})=(\underline{a}\bullet\underline{c})\underline{b}-(\underline{a}\bullet\underline{b})\underline{c}$ (prove by expanding both sides in components – straightforward but tedious)

Some Vector Identities:

(a) $(\underline{a}\times\underline{b})\bullet(\underline{c}\times\underline{d})=(\underline{a}\bullet\underline{c})(\underline{b}\bullet\underline{d})-(\underline{a}\bullet\underline{d})(\underline{b}\bullet\underline{c})$

(b) $(\underline{a}\times\underline{b})\times(\underline{c}\times\underline{d})=(\underline{a},\underline{b},\underline{d})\underline{c}-(\underline{a},\underline{b},\underline{c})\underline{d}$

(c) $(\underline{a}\times\underline{b})\bullet(\underline{b}\times\underline{c})\times(\underline{c}\times\underline{a})=(\underline{a},\underline{b},\underline{c})^2$

Example

Prove identity (a) above.

$$\begin{aligned}(\underline{a}\times\underline{b})\bullet(\underline{c}\times\underline{d}) &= \underline{a}\bullet\left[\underline{b}\times(\underline{c}\times\underline{d})\right] && \text{triple scalar product of } \underline{a}, \underline{b},\ \underline{c}\times\underline{d}\\ &= \underline{a}\bullet\left[(\underline{b}\bullet\underline{d})\underline{c}-(\underline{b}\bullet\underline{c})\underline{d}\right] && \text{property (ii)}\\ &= (\underline{a}\bullet\underline{c})(\underline{b}\bullet\underline{d})-(\underline{a}\bullet\underline{d})(\underline{b}\bullet\underline{c})\end{aligned}$$

4. Linear Dependence and Independence

If $\underline{a}_1, \underline{a}_2, \ldots., \underline{a}_k$ are any k n-component vectors, then an expression of the form $\sum_{i=1}^{k} m_i \underline{a}_i$, ($m_1, m_2, \ldots.., m_k$ are any k scalars) is called a linear combination of $\underline{a}_1, \underline{a}_2, \ldots.,\ \underline{a}_k$.

$\underline{a}_1, \underline{a}_2, \ldots., \underline{a}_k$ is linearly dependent if at least one of the vectors can be represented as a linear combination of the others. Otherwise the set is linearly independent.

Examples

The vectors $\underline{a} = 3\underline{i} + 5\underline{j} - 2\underline{k}$, $\underline{b} = 4\underline{j} + 2\underline{k}$ and $\underline{c} = \underline{i} + \underline{j} - \underline{k}$ are linearly dependent since $\underline{a} = \frac{1}{2}\underline{b} + 3\underline{c}$

Hence vector $\underline{a}$ lies in the plane of the vectors $\underline{b}$ and $\underline{c}$. However, the vectors $\underline{i}$, $\underline{j}$ and $\underline{k}$ are linearly independent.

An equivalent definition is: A set of k n-components vectors is linearly independent *iff* $\sum_{i=1}^{k} m_i \underline{a}_i = \underline{0}$ implies $m_1 = m_2 = \cdots = m_k = 0$, that is, the vector equation $\sum_{i=1}^{k} m_i \underline{a}_i = \underline{0}$ has the trivial solution $m_1 = m_2 = \cdots = m_k = 0$ only.

Proof of equivalence :

Assume $m_p \neq 0$ for some $1 \le p \le k$, then $\sum_{i=1}^{k} m_i \underline{a}_i = \underline{0}$ *iff* $m_p \underline{a}_p = -\sum_{\substack{i=1\\ i\neq p}}^{k} m_i \underline{a}_i$

iff $\underline{a}_p = -\sum_{\substack{i=1\\ i\neq p}}^{k} \frac{m_i}{m_p} \underline{a}_i$ *iff* $\{\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_k\}$ is linearly dependent.

If two vectors in 3-D space are linearly dependent they must be collinear. If three vectors in 3-D space are linearly dependent they must either be collinear of coplanar. Hence three vectors form a linearly independent set *iff* their triple scalar product is not zero.

Four or more vectors in 3-D space will always be linearly dependent.

Example

If $\underline{a} = 3\underline{i} + 5\underline{j} - 2\underline{k}$, $\underline{b} = 4\underline{j} + 2\underline{k}$, $\underline{c} = \underline{i} + \underline{j} - \underline{k}$, $\underline{c} = \underline{i} + \underline{j} - \underline{k}$

$$(\underline{a},\underline{b},\underline{c}) = \begin{vmatrix} 3 & 5 & -2 \\ 0 & 4 & 2 \\ 1 & 1 & -1 \end{vmatrix} = 0$$

And hence $\underline{a}$, $\underline{b}$ and $\underline{c}$ are linearly dependent.

Example

Show that the four 4-component vectors, $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, e_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ in R^4 are linearly independent.

Proof:

Consider the vector equation $x_1 e_1 + x_2 e_2 + x_3 e_3 + x_4 e_4 = \underline{0}$, that is,

$$x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Then

$$x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow x_1 = x_2 = x_3 = x_4 = 0$$

Example

Given any four 3-component vectors, $\underline{v}_1 = \begin{pmatrix} v_{11} \\ v_{21} \\ v_{31} \end{pmatrix}, \underline{v}_2 = \begin{pmatrix} v_{12} \\ v_{22} \\ v_{32} \end{pmatrix}, \underline{v}_3 = \begin{pmatrix} v_{13} \\ v_{23} \\ v_{33} \end{pmatrix}, \underline{v}_4 = \begin{pmatrix} v_{14} \\ v_{24} \\ v_{34} \end{pmatrix}$, show that they must be dependent.

Proof:

$$x_1 \underline{v}_1 + x_2 \underline{v}_2 + x_3 \underline{v}_3 + x_4 \underline{v}_4 = x_1 \begin{pmatrix} v_{11} \\ v_{21} \\ v_{31} \end{pmatrix} + x_2 \begin{pmatrix} v_{12} \\ v_{22} \\ v_{32} \end{pmatrix} + x_3 \begin{pmatrix} v_{13} \\ v_{23} \\ v_{33} \end{pmatrix} + x_4 \begin{pmatrix} v_{14} \\ v_{24} \\ v_{34} \end{pmatrix} = \underline{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} x_1 v_{11} + x_2 v_{12} + x_3 v_{13} + x_4 v_{14} \\ x_1 v_{21} + x_2 v_{22} + x_3 v_{23} + x_4 v_{24} \\ x_1 v_{31} + x_2 v_{32} + x_3 v_{33} + x_4 v_{34} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} v_{11} x_1 + v_{12} x_2 + v_{13} x_3 + v_{14} x_4 = 0 \\ v_{21} x_1 + v_{22} x_2 + v_{23} x_3 + v_{24} x_4 = 0 \\ v_{31} x_1 + v_{32} x_2 + v_{33} x_3 + v_{34} x_4 = 0 \end{cases}$$

$$\begin{cases} v_{11}x_1 + v_{12}x_2 + v_{13}x_3 + v_{14}x_4 = 0 \\ v_{21}x_1 + v_{22}x_2 + v_{23}x_3 + v_{24}x_4 = 0 \\ v_{31}x_1 + v_{32}x_2 + v_{33}x_3 + v_{34}x_4 = 0 \end{cases}$$ is a homogeneous system in unknowns x_1, x_2, x_3, x_4 and since there are more unknowns than equations, there must exist infinitely many solutions for x_1, x_2, x_3, x_4, thus, there must exist non-trivial solutions for x_1, x_2, x_3, x_4. It therefore follows that

$$\underline{v}_1 = \begin{pmatrix} v_{11} \\ v_{21} \\ v_{31} \end{pmatrix}, \underline{v}_2 = \begin{pmatrix} v_{12} \\ v_{22} \\ v_{32} \end{pmatrix}, \underline{v}_3 = \begin{pmatrix} v_{13} \\ v_{23} \\ v_{33} \end{pmatrix}, \underline{v}_4 = \begin{pmatrix} v_{14} \\ v_{24} \\ v_{34} \end{pmatrix}$$ must be linearly dependent.

In R^n, $n \ge 1$ there are *n* linearly independent *n*-component vectors , for instance, $e_1, \cdots, e_n$,

whereas any set of *n* + 1 or more *n*-component vectors is linearly dependent.

Consider $\{\underline{v}_1, \underline{v}_2, ..., \underline{v}_m\}$ in R^n, $n \ge 1$, if $\underline{v}_1, \underline{v}_2, ..., \underline{v}_m$ are linearly independent and each *n*-component vector $\underline{v} \in R^n$ is a linear combination of $\underline{v}_1, \underline{v}_2, ..., \underline{v}_m$ then $\{\underline{v}_1, \underline{v}_2, ..., \underline{v}_m\}$ is called a <u>basis</u> of R^n, $n \ge 1$, for instance, $\{e_1, \cdots, e_n\}$ is a basis of R^n. Note, however, that a basis is not unique.

If $\{\underline{v}_1, \underline{v}_2, ..., \underline{v}_m\}$ is a basis of R^n, then $m = n$, that is, every basis of R^n contains *n* vectors and we say that R^n has <u>dimension *n*</u> .

$\{\underline{v}_1, \underline{v}_2, ..., \underline{v}_m\}$ in R^n is said to be <u>orthogonal</u> if $\underline{v}_i \bullet \underline{v}_j = 0$ if $i \ne j$. $\{\underline{v}_1, \underline{v}_2, ..., \underline{v}_m\}$ in R^n is <u>orthonormal</u> if $\{\underline{v}_1, \underline{v}_2, ..., \underline{v}_m\}$ in R^n is <u>orthogonal</u> and $\|\underline{v}_i\|^2 = \underline{v}_i \bullet \underline{v}_i = 1$ for $i = 1, 2, ..., m$.

n-component non-zero vectors which are orthogonal are also linearly independent (can you prove this?) but the converse is not true (give an example).

<u>Example</u>

(i) For R^3, the dimension of R^3 is 3, as expected, and the 3 vectors $\underline{i}$, $\underline{j}$ and $\underline{k}$, which are linearly independent, form an orthonormal basis for R^3. Any vector $\underline{v}$ may be written as a linear combination of $\underline{i}$, $\underline{j}$ and $\underline{k}$.

(ii) The vectors $\underline{i} + \underline{j}$, $2\underline{i} - \underline{j}$ and $\underline{k}$ also form a basis for R^3 since they are linearly independent and any vector in R^3 may be expressed as a linear combination of these vectors, eg. $-4\underline{i} + 5\underline{j} + 6\underline{k} = 2(\underline{i} + \underline{j}) - 3(2\underline{i} - \underline{j}) + 6\underline{k}$. However they are not useful in practice since they are not orthogonal.

(iii) The vectors $\underline{a} = 3\underline{i} + 5\underline{j} - 2\underline{k}$, $\underline{b} = 4\underline{j} + 2\underline{k}$ and $\underline{c} = \underline{i} + \underline{j} - \underline{k}$ of the previous example do not form a basis for R^3 since they are linearly dependent. The vector $\underline{d} = \underline{i} + \underline{j} + \underline{k}$, for example, cannot be expressed as a linear combination of $\underline{a}$, $\underline{b}$ and $\underline{c}$.

(iv) In R^n, the n vectors $\{e_1, e_2, \cdots, e_n\}$, that is, $\left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\}$ are linearly independent and thus form a basis of R^n, called the <u>standard basis</u> and also they are orthonormal therefore form an orthonormal basis of R^n.

Matrix Algebra

1. Introduction

A matrix of order $m \times n$ or an $m \times n$ matrix is a rectangular array of numbers having m rows and n columns. It can be written

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ . & & & . \\ . & & & . \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = \left(a_{ij}\right)$$

The mn numbers $a_{11}, \dots, a_{mn}$ are called the elements (entries) of the matrix. The notation a_{ij} denotes the element of A which is in the i^{th} row and j^{th} column.

A matrix with only one row is called a row matrix or row vector while a matrix with only one column is a column matrix or column vector. These will be written $\underline{x}^T$ and $\underline{x}$ respectively.

A matrix with n rows and n columns is a square matrix of order n. The elements $a_{11}, a_{22}, \dots, a_{nn}$ are the diagonal elements.

Two matrices A and B are equal *iff* they have the same number of rows and columns and $a_{ij} = b_{ij}$ for all i and j.

Addition: The sum of two $m \times n$ matrices $A = \left(a_{ij}\right)$ and $B = \left(b_{ij}\right)$ is the $m \times n$ matrix $C = \left(c_{ij}\right)$ where

$c_{ij} = a_{ij} + b_{ij}$ for $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$. We write $C = A + B$.

The zero matrix 0 has all elements zero.

Scalar multiplication: The product of an $m \times n$ matrix $A = \left(a_{ij}\right)$ with a scalar (number) q is

$qA = Aq = \left(qa_{ij}\right)$ i.e. every element of A is multiplied by q.

The matrix $-A = \left(-a_{ij}\right)$ is the negation of A.

We note that with these definitions, the following properties hold:

1. $A+B=B+A$
2. $A+(B+C)=(A+B)+C$
3. $A+O=A$
4. $A+(-A)=O$
5. $q(A+B)=qA+qB$
6. $(p+q)A=pA+qA$
7. $(pq)A=p(qA)$
8. $1A=A$

The transpose A^T of an $m\times n$ matrix A is the $n\times m$ matrix obtained by interchanging the rows and columns of A.

Let A be a real square matrix it is said to be symmetric if $A^T=A$ and skew-symmetric if $A^T=-A$.

A diagonal matrix has all elements not on the diagonal zero i.e. $a_{ij}=0$ if $i\neq j$.

The $n\times n$ unit matrix I or $I_n=\left(\delta_{ij}\right)$ is a diagonal matrix with all its diagonal elements unity.

The product of an $m \times n$ matrix A with an $n\times p$ matrix B is the $m\times p$ matrix $C = AB$ with elements

$$c_{ij}=\sum_{k=1}^{n}a_{ik}b_{kj}=\text{(Row } i \text{ of } A)\bullet\text{(Column } j \text{ of } B),\quad i=1,2,...,m\ ;\ \ j=1,2,.....,p$$

Note that the matrix multiplication is only possible if the number of columns of A is the same as the number of rows of B. i.e. the matrices are conformable. The element c_{ij} is the dot product of the i^{th} row of A and the j^{th} column of B considering them as vectors in the vector space R^n.

Example

Let $A=\begin{pmatrix}1&1&2\\2&3&-1\end{pmatrix}, B=\begin{pmatrix}1&0&1\\2&1&0\\-1&1&0\end{pmatrix}$, find AB.

Solution:

$$AB=\begin{pmatrix}1&1&2\\2&3&-1\end{pmatrix}\begin{pmatrix}1&0&1\\2&1&0\\-1&1&0\end{pmatrix}=\begin{pmatrix}\begin{pmatrix}1\\1\\2\end{pmatrix}\bullet\begin{pmatrix}1\\2\\-1\end{pmatrix} & \begin{pmatrix}1\\1\\2\end{pmatrix}\bullet\begin{pmatrix}0\\1\\1\end{pmatrix} & \begin{pmatrix}1\\1\\2\end{pmatrix}\bullet\begin{pmatrix}1\\0\\0\end{pmatrix}\\ \begin{pmatrix}2\\3\\-1\end{pmatrix}\bullet\begin{pmatrix}1\\2\\-1\end{pmatrix} & \begin{pmatrix}2\\3\\-1\end{pmatrix}\bullet\begin{pmatrix}0\\1\\1\end{pmatrix} & \begin{pmatrix}2\\3\\-1\end{pmatrix}\bullet\begin{pmatrix}1\\0\\0\end{pmatrix}\end{pmatrix}=\begin{pmatrix}1&3&1\\9&2&2\end{pmatrix}$$

Suppose $B=\left(\underline{b}_1\middle|\underline{b}_2\middle|\underline{b}_3\right)$ where $\underline{b}_1,\underline{b}_2,\underline{b}_3$ are the correspondent columns of B. We observe that the first column of AB is $A\underline{b}_1$, the second column of AB is $A\underline{b}_2$, the third column of AB is $A\underline{b}_3$, that is, $AB=\left(A\underline{b}_1\middle|A\underline{b}_2\middle|A\underline{b}_3\right)$.

Then $AB=\left(\begin{pmatrix}1&1&2\\2&3&-1\end{pmatrix}\begin{pmatrix}1\\2\\-1\end{pmatrix}\quad\begin{pmatrix}1&1&2\\2&3&-1\end{pmatrix}\begin{pmatrix}0\\1\\1\end{pmatrix}\quad\begin{pmatrix}1&1&2\\2&3&-1\end{pmatrix}\begin{pmatrix}1\\0\\0\end{pmatrix}\right)=\begin{pmatrix}1&3&1\\9&2&2\end{pmatrix}$

In addition, Let $A=\left(\underline{a}_1\middle|\underline{a}_2\middle|\underline{a}_3\right)$ where $\underline{a}_1,\underline{a}_2,\underline{a}_3$ are the correspondent columns of A, we also observe

$$AB=\begin{pmatrix}1&1&2\\2&3&-1\end{pmatrix}\begin{pmatrix}1&0&1\\2&1&0\\-1&1&0\end{pmatrix}=\left(\underline{a}_1\middle|\underline{a}_2\middle|\underline{a}_3\right)\begin{pmatrix}b_{11}&b_{12}&b_{13}\\b_{21}&b_{22}&b_{23}\\b_{31}&b_{32}&b_{33}\end{pmatrix}$$
$$=\left(b_{11}\underline{a}_1+b_{21}\underline{a}_2+b_{31}\underline{a}_3\middle|b_{12}\underline{a}_1+b_{22}\underline{a}_2+b_{32}\underline{a}_3\middle|b_{13}\underline{a}_1+b_{23}\underline{a}_2+b_{33}\underline{a}_3\right)$$
$$=\left(1\begin{pmatrix}1\\2\end{pmatrix}+2\begin{pmatrix}1\\3\end{pmatrix}+(-1)\begin{pmatrix}2\\-1\end{pmatrix}\middle|0\begin{pmatrix}1\\2\end{pmatrix}+1\begin{pmatrix}1\\3\end{pmatrix}+1\begin{pmatrix}2\\-1\end{pmatrix}\middle|1\begin{pmatrix}1\\2\end{pmatrix}+0\begin{pmatrix}1\\3\end{pmatrix}+0\begin{pmatrix}2\\-1\end{pmatrix}\right)=\begin{pmatrix}1&3&1\\9&2&2\end{pmatrix}$$

Matrix multiplication has the following properties:

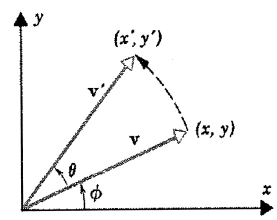
1. $(qA)B=q(AB)=A(qB)$, normally written as qAB
2. $A(BC)=(AB)C$, normally written as ABC
3. $A(B+C)=AB+AC$
4. $(A+B)C=AC+BC$

Note however that

1. $AB\neq BA$ in general
2. $AB=O\ \not\Rightarrow\ A=O$ or $B=O$

<u>Example</u> – Rotation of Axes

Let θ be a fixed angle and suppose a point $\underline{v}=\begin{pmatrix}x\\y\end{pmatrix}$ in the plane transforms to the point $\underline{v}'=\begin{pmatrix}x'\\y'\end{pmatrix}$ according to the rule $\begin{cases}x'=x\cos\theta-y\sin\theta\\y'=x\sin\theta+y\cos\theta\end{cases}$. This is known as a <u>linear transformation</u> and is called the rotation of R^2 through the angle θ clockwise. $\begin{cases}x'=x\cos\theta-y\sin\theta\\y'=x\sin\theta+y\cos\theta\end{cases}$ may be represented by the matrix equation $\begin{pmatrix}x'\\y'\end{pmatrix}=\begin{pmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$. Geometrically, $\underline{v}'=\begin{pmatrix}x'\\y'\end{pmatrix}$ is the vector that results if the coordinate axes are rotated clockwise through an angle θ. To see this, let ϕ be the angle between $\underline{v}=\begin{pmatrix}x\\y\end{pmatrix}$ and the positive x axis. Suppose the coordinate axes are rotated clockwise through an angle θ, we observe that the resultant effect is equivalent to fixing the coordinate axes and $\underline{v}$ is rotated anticlockwise through an angle θ. Let $\underline{v}'=\begin{pmatrix} x' \\ y' \end{pmatrix}$ be the vector that results when $\underline{v}=\begin{pmatrix} x \\ y \end{pmatrix}$ is rotated anticlockwise through an angle θ. If r denotes the length of $\underline{v}$, then $x=r\cos\phi,\ y=r\sin\phi$. Similarly, since $\underline{v}'$ has the same length as $\underline{v}$, we have $x'=r\cos(\theta+\phi),\ y'=r\sin(\theta+\phi)$.



Therefore,

$$\begin{pmatrix} x' \\ y' \end{pmatrix}=\begin{pmatrix} r\cos(\theta+\phi) \\ r\sin(\theta+\phi) \end{pmatrix}=\begin{pmatrix} r\cos\theta\cos\phi-r\sin\theta\sin\phi \\ r\sin\theta\cos\phi+r\cos\theta\sin\phi \end{pmatrix}=\begin{pmatrix} x\cos\theta-y\sin\theta \\ x\sin\theta+y\cos\theta \end{pmatrix}$$
$$=\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

For instance, if the coordinate axes are rotated clockwise through an angle $\theta=\dfrac{\pi}{6}$, then the point (2,3) has the resultant coordinates (x',y') and $\begin{pmatrix} x' \\ y' \end{pmatrix}=\begin{pmatrix} \cos\frac{\pi}{6} & -\sin\frac{\pi}{6} \\ \sin\frac{\pi}{6} & \cos\frac{\pi}{6} \end{pmatrix}=\begin{pmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix}=\begin{pmatrix} \sqrt{3}-3/2 \\ 1+3\sqrt{3}/2 \end{pmatrix}$

If the point $\underline{v}=\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ in the plane first transforms to the point $\underline{u}=\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ according to the rule

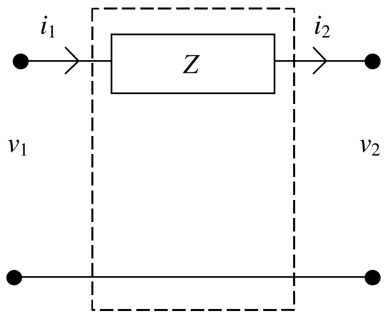
$\begin{cases} x_2=a_{11}x_1+a_{12}y_1 \\ y_2=a_{21}x_1+a_{22}y_1 \end{cases}$, that is, $\underline{u}=A\underline{v}$ where $A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, and then transforms to the point $\underline{w}=\begin{pmatrix} x_3 \\ y_3 \end{pmatrix}$

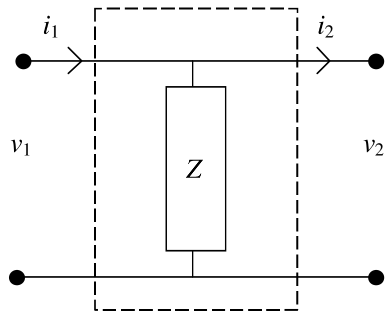
according to the rule $\begin{cases} x_3=b_{11}x_2+b_{12}y_2 \\ y_3=b_{21}x_2+b_{22}y_2 \end{cases}$, that is, $\underline{w}=B\underline{u}$ where $B=\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$, then

$\underline{w}=\begin{pmatrix} x_3 \\ y_3 \end{pmatrix}, \underline{v}=\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ are related by $\underline{w}=BA\underline{v}$.

Example Two-port networks

Consider the following 2-port (4-terminal) networks with an impedance Z in series and parallel respectively:





v=potential
i=current
Z=impedance

$$\begin{cases} i_1 = i_2 \\ v_1 - v_2 = i_2 Z \end{cases} \qquad \begin{cases} v_1 = v_2 \\ v_2 = (i_1 - i_2) Z \end{cases}$$

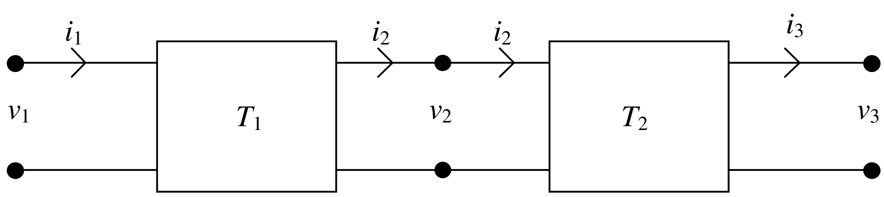
(Ohm's Law)

or in matrix terms

$$\begin{pmatrix} v_1 \\ i_1 \end{pmatrix} = \begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_2 \\ i_2 \end{pmatrix} \qquad \begin{pmatrix} v_1 \\ i_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1/Z & 1 \end{pmatrix} \begin{pmatrix} v_2 \\ i_2 \end{pmatrix}$$

In each ease the 2×2 matrix is known as the transmission matrix for the network.

If two two-point networks are connected in cascade with transmission matrices T_1 and T_2 respectively:



then the combined transmission matrix will be $T_1 T_2$ since

$$\begin{pmatrix} v_1 \\ i_1 \end{pmatrix} = T_1 \begin{pmatrix} v_2 \\ i_2 \end{pmatrix} \text{ and } \begin{pmatrix} v_2 \\ i_2 \end{pmatrix} = T_2 \begin{pmatrix} v_3 \\ i_3 \end{pmatrix} \text{ hence } \begin{pmatrix} v_1 \\ i_1 \end{pmatrix} = T_1 T_2 \begin{pmatrix} v_3 \\ i_3 \end{pmatrix}$$

You may check that this is true in the particular case of (a) – take $Z = Z_1$ and (b) – take $Z = Z_2$ connected in cascade.

Similarly any number of networks connected in cascade may be analyzed in this way.

2. Determinants

Every square matrix A has a number associated with it called the <u>determinant</u> of A, written as $\det A$ or $|A|$.

For a 2×2 matrix $A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, $|A|=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}=a_{11}a_{22}-a_{12}a_{21}$

For a 3×3 matrix $A=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$,

$$|A|=\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}=a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}-a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}+a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}=a_{11}A_{11}+a_{12}A_{12}+a_{13}A_{13}$$

where A_{ij} is the cofactor of element a_{ij}, and $A_{ij}=(-1)^{i+j}M_{ij}$ where M_{ij} is the <u>minor</u> of the element a_{ij} and is the determinant of that submatrix of A obtained by deleting the i^{th} row and j^{th} column of A.

$|A|$ above was <u>expanded by the first row</u> although we could have similarly expanded by any row or column to give the same result.

For an $n\times n$ matrix, we have

$$\begin{aligned} &\text{by row :} && |A|=a_{i1}A_{i1}+a_{i2}A_{i2}+\ldots+a_{in}A_{in} && i=1,2,\ldots\ldots \text{ or } n \\ &\text{by column :} && |A|=a_{1j}A_{1j}+a_{2j}A_{2j}+\ldots+a_{nj}A_{nj} && j=1,2,\ldots\ldots \text{ or } n \end{aligned}$$

, which defines $|A|$ in terms of $(n-1)\times(n-1)$ determinants, each of which is then defined in terms of $(n-2)\times(n-2)$ determinants etc.

Properties

(i) $|A|=|A^T|$ i.e. rows and columns may be interchanged.

(ii) If all the elements in a row (or column) are zero, then $|A|=0$.

(iii) Interchanging any two rows (or columns) reverses the sign of $|A|$.

(iv) If corresponding elements in any two rows (or columns) are proportional then $|A| = 0$.

(v) The value of a determinant is unchanged if a multiple of one row (or column) is added to another row (or column).

(vi) $|AB| = |A||B|$, both A and B must be square matrices.

(vii) If the elements in any row (or column) are multiplied by a number, then $|A|$ is multiplied by that number. Note that $|kA| = k^n |A| \quad (\neq k|A|)$.

(viii) $|A| = 0$ if the rows (or columns) are linearly dependent.

Determinants are not of great practical use as they are expensive to compute, but are of theoretical value.

3. Systems of Linear Equations

A set of equations of the form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases}$$

is called a <u>system of *m* linear equations in *n* unknowns</u>, $x_1, x_2, \ldots, x_n$. The numbers a_{ij} are the <u>coefficients</u> of the system and are given. The "right-hand side" numbers $b_1, b_2, \ldots, b_m$ are also given. If all the b_i are zero the system is <u>homogeneous</u> otherwise it is <u>inhomogeneous</u>. A <u>solution</u> is a set of numbers $x_1, x_2, \ldots, x_n$ satisfying all m equations.

The system may be written in matrix form $A\underline{x} = \underline{b}$, where

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

<u>coefficient matrix</u> <u>solution vector</u> <u>right-hand side vector</u>

The $m \times (n+1)$ matrix

$$B = (A, \underline{b}) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ . & & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}$$

is known as the <u>augmented matrix</u> of the system. The solution of a system of linear equations is not changed by

(a) interchanging any two equations,
(b) multiplying any equation by a non-zero constant,
(c) adding a constant multiple of one equation to another equation.

If instead of equations we consider the augmented matrix of the system, we may define the following <u>elementary row operations</u>

(a) interchanging any two rows (type 1),
(b) multiplying any row by a non-zero constant (type 2),
(c) adding a constant multiple of one row to another row (type 3).

Two matrices are <u>row equivalent</u> if one may be obtained from the other in a finite number of elementary row operations. Clearly two augmented matrices which are row equivalent may represent systems of linear equations with the <u>same</u> solution.

<u>Gaussian Elimination</u>

This method (and variations of it) is the most popular method of solving system of linear equations on a computer.

<u>Example</u>

Solve the system $\begin{cases} x_1 + 4x_2 - 2x_3 = 3 \\ 2x_1 - 2x_2 + x_3 = 1 \\ 3x_1 + x_2 + 2x_3 = 11 \end{cases}$ by Gaussian Elimination.

Solution:

$$\begin{array}{cc} \begin{pmatrix} 1 & 4 & -2 & : & 3 \\ 2 & -2 & 1 & : & 1 \\ 3 & 1 & 2 & : & 11 \end{pmatrix} & \begin{array}{c} r_1 \\ r_2 \\ r_3 \end{array} \end{array}$$

Elimination Stage:

$$\underset{\substack{r_2-2r_1\\ r_3-3r_1}}{\sim}\left(\begin{array}{cccccc}1 & 4 & -2 & : & 3\\ 0 & -10 & 5 & : & -5\\ 0 & -11 & 8 & : & 2\end{array}\right)\begin{array}{l}r_1\\ r_{2a}\\ r_{3a}\end{array} \underset{r_{3a}-(-11/-10)r_{2a}}{\sim}\left(\begin{array}{cccccc}1 & 4 & -2 & : & 3\\ 0 & -10 & 5 & : & -5\\ 0 & 0 & 2.5 & : & 7.5\end{array}\right)\begin{array}{l}r_1\\ r_{2a}\\ r_{3b}\end{array}$$

$$\Rightarrow\begin{cases} x_1+4x_2-2x_3=3\\ \quad -10x_2+5x_3=-5\\ \qquad\qquad 2.5x_3=7.5\end{cases}$$

The system is now in upper triangular form and we proceed with the back substitution finding x_1, x_2, x_3 in reverse order: $x_3=3, x_2=2, x_1=1$

For $n>3$ the method proceeds in the same way, reducing those elements below the diagonal in a column to zero by subtracting multiples of a row.

At an intermediate stage of the elimination we have,

$$\begin{array}{cc} & \text{column } r \\ \text{row } r & \left(\begin{array}{ccccccc} a_{11} & a_{12} & \dots & a_{1r} & \dots & a_{1n} & . & b_1\\ 0 & a_{22}^{(2)} & & a_{2r}^{(2)} & \dots & a_{2n}^{(2)} & . & b_2^{(2)}\\ 0 & 0 & & .. & & & . & \\ . & & 0 & a_{rr}^{(r)} & \dots & a_{rn}^{(r)} & . & b_r^{(r)}\\ . & & 0 & a_{r+1r}^{(r)} & \dots & . & . & .\\ . & & . & & & . & . & .\\ 0 & \dots & 0 & a_{nr}^{(r)} & \dots & a_{nn}^{(r)} & . & b_n^{(r)}\end{array}\right)\end{array}$$

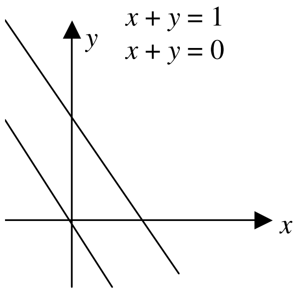
$a_{rr}^{(r)}$ is known as the pivot. $a_{r+1r}^{(r)}$ is reduced to zero by subtracting $(a_{r+1r}^{(r)}/a_{rr}^{(r)})$ times row r from row r+1, and similarly for the other elements $a_{r+2r}^{(r)},\dots,a_{nr}^{(r)}$ in the r^{th} column.

If $a_{rr}^{(r)}=0$, then row r cannot be used as the pivot and r^{th} is interchanged with some row below it. In practice, in order to minimize the growth of rounding errors in a computer algorithm for this method, row r is interchanged with that row below it for which $\left|a_{ir}^{(r)}\right| \quad (i=r, r+1,\dots,n)$ is largest. This ensures that the multipliers $(a_{ir}^{(r)}/a_{rr}^{(r)})$ are all ≤ 1 in modulus and is known as partial pivoting.

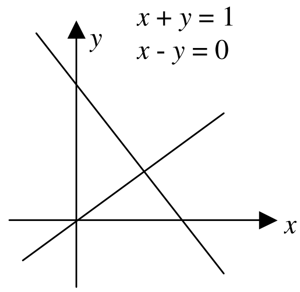
In the above we have assumed n equations in the n unknowns and that there is a unique solution. There are other possibilities.
If a system of equations has at least one solution, we say the equations are consistent, otherwise they are inconsistent.

Example

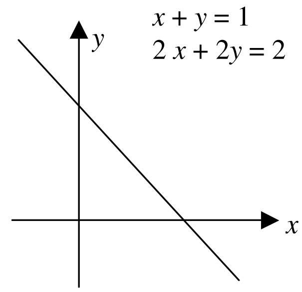
Take $m = n = 2$



no solution
(inconsistent)



one solution
(consistent)



infinitely many solutions
(consistent)

Let us view these sets of equations in 2 different ways which will give us some insight into the more general results given later.

A. In each case the equations are equivalent to:

(a) $x\begin{pmatrix}1\\1\end{pmatrix}+y\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}$, (b) $x\begin{pmatrix}1\\1\end{pmatrix}+y\begin{pmatrix}1\\-1\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}$, (c) $x\begin{pmatrix}1\\2\end{pmatrix}+y\begin{pmatrix}1\\2\end{pmatrix}=\begin{pmatrix}1\\2\end{pmatrix}$

For a solution to exist the right-hand side vector $\underline{b}$ must be a linear combination of the vectors formed by the columns of A. If the columns of A are themselves linearly dependent then that solution will not be unique.

B. Performing elementary row operations:

(a) $\begin{pmatrix}1&1&1\\1&1&0\end{pmatrix}$ (b) $\begin{pmatrix}1&1&1\\1&-1&0\end{pmatrix}$ (c) $\begin{pmatrix}1&1&1\\2&2&2\end{pmatrix}$

$\begin{pmatrix}1&1&1\\0&0&-1\end{pmatrix}$ $\begin{pmatrix}1&1&1\\0&-2&-1\end{pmatrix}$ $\begin{pmatrix}1&1&1\\0&0&0\end{pmatrix}$

In case (a) the last equation gives $0=-1$ which shows that there is no solution. In case (b) there is a unique solution y = 0.5, x = 0.5. In case (c) a row of zeros has been produced which means that y may be assigned arbitrarily, say $y = k$ and then x = 1-k for any value of k.

A matrix is said to be in <u>row echelon form</u> if (i) the first k rows of it are nonzero and the rest are zero; (ii) the *pivot* a_{in_i} of the i^{th} (i=1,…,k) row (i.e. the first nonzero entry of each nonzero row) is to the right of the pivots of the preceding rows.

In addition, a matrix is said to be in <u>reduced row echelon form</u> if (iii) the pivot a_{in_i} of each nonzero row is 1; (iv) it is in row echelon form and the pivots are the only nonzero entries in their columns.

Example

The matrices,

$$\begin{pmatrix}1 & 4 & -2 & 3\\ 0 & -10 & 5 & -5\\ 0 & 0 & 2.5 & 7.5\end{pmatrix},\ \begin{pmatrix}1 & 1 & 1\\ 0 & 0 & -1\end{pmatrix},\ \begin{pmatrix}1 & 1 & 1\\ 0 & -2 & -1\end{pmatrix},\ \begin{pmatrix}1 & 1 & 1\\ 0 & 0 & 0\end{pmatrix},\ \begin{pmatrix}1 & 2 & 0 & 3\\ 0 & 0 & 1 & 2\\ 0 & 0 & 0 & 0\end{pmatrix}$$ are all in row echelon form and $\begin{pmatrix}1 & 1 & 1\\ 0 & 0 & 0\end{pmatrix}, \begin{pmatrix}1 & 2 & 0 & 3\\ 0 & 0 & 1 & 2\\ 0 & 0 & 0 & 0\end{pmatrix}$ are in reduced row echelon form.

Recall that a set of vectors is linearly independent if none of the vectors can be written as a linear combination of the other vectors in the set.

Example

$\begin{pmatrix}1\\4\\-2\\3\end{pmatrix}, \begin{pmatrix}2\\-2\\1\\1\end{pmatrix}, \begin{pmatrix}3\\1\\2\\11\end{pmatrix}$ are linearly independent since $a\begin{pmatrix}1\\4\\-2\\3\end{pmatrix} + b\begin{pmatrix}2\\-2\\1\\1\end{pmatrix} + c\begin{pmatrix}3\\1\\2\\11\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\end{pmatrix}$ gives

$$\begin{pmatrix}a + 2b + 3c\\ 4a-2b+c\\ -2a+b+2c\\ 3a+b+11c\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\end{pmatrix} \Leftrightarrow \begin{cases} a+2b+3c=0\\ 4a-2b+c=0\\ -2a+b+2c=0\\ 3a+b+11c=0\end{cases}.$$

Using Gaussian elimination the only solution is $a=b=c=0$.

Example

Consider the matrix $A = \begin{pmatrix}1 & 4 & -2 & 3\\ 0 & -10 & 5 & -5\\ 0 & 0 & 2.5 & 7.5\end{pmatrix}$. Show that the row vectors of A $\begin{pmatrix}1\\4\\-2\\3\end{pmatrix}, \begin{pmatrix}0\\-10\\5\\-5\end{pmatrix}, \begin{pmatrix}0\\0\\2.5\\7.5\end{pmatrix}$ are linearly independent.

Proof:

$$a\begin{pmatrix}1\\4\\-2\\3\end{pmatrix} + b\begin{pmatrix}0\\-10\\5\\-5\end{pmatrix} + c\begin{pmatrix}0\\0\\2.5\\7.5\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\end{pmatrix} \text{ gives } \begin{cases} a=0\\ 4a-10b=0\\ -2a+5b+2.5c=0\\ 3a-5b+7.5c=0\end{cases}$$

Directly we see that the only solution is $a = b = c = 0$.

Note that the non-zero rows of the echelon form of a matrix are independent.

Example

Suppose the following matrix in reduced row echelon form is the augmented matrix of some inhomogeneous system. Solve the system and express the solutions in vector form.

$$\left(\begin{array}{ccccc|c} 0 & 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Solution:

As there are no pivots in columns 1 & 3 , x_1, x_3 are therefore free i.e. you can assign parameters to them. However, there are pivots in columns 2 , 4 , 5, x_2, x_4, x_5 are therefore restricted i.e. they must be expressed in terms of the free variables or the constants. Let $x_1 = s, x_3 = t$; where s, t are any real numbers.

From row three we have $x_5 = 2$. From row two we have $x_4 = 0$. From row one we have $x_2 = 1 - 2x_3 = 1 - 2t$.

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} s \\ 1-2t \\ t \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 2 \end{pmatrix} + \begin{pmatrix} s \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -2t \\ t \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 2 \end{pmatrix} + s\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$, where s, t are any real numbers.

The rank of a matrix A, denoted by rank A, is the maximum number of linearly independent row vectors.

Example

$$A = \begin{pmatrix} 1 & 4 & -2 & 3 \\ 0 & -10 & 5 & -5 \\ 0 & 0 & 2.5 & 7.5 \end{pmatrix}, \text{ rank } A = 3; \quad A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix}, \text{ rank } A = 1$$

Theorem

Row equivalent matrices have the same rank.

Proof: Not required.

Theorem

Matrices A and A^T have the same rank.

Proof: Not required.

As a result of the theorem, the maximum number of linearly independent row vectors of A is the same as the maximum number of linearly independent columns vectors of A.

The rank of a matrix A may also be defined as the maximum number of linearly independent column vectors of A.

Example

Show that $\begin{pmatrix} 1 & 4 & -2 & 3 \\ 2 & -2 & 1 & 1 \\ 3 & 1 & 2 & 11 \end{pmatrix}$, $\begin{pmatrix} 1 & 4 & -2 & 3 \\ 0 & -10 & 5 & -5 \\ 0 & 0 & 2.5 & 7.5 \end{pmatrix}$ both have rank 3.

Proof:

$$\begin{pmatrix} 1 & 4 & -2 & 3 \\ 2 & -2 & 1 & 1 \\ 3 & 1 & 2 & 11 \end{pmatrix} \underset{\substack{r_2-2r_1 \\ r_3-3r_1}}{\sim} \begin{pmatrix} 1 & 4 & -2 & 3 \\ 0 & -10 & 5 & -5 \\ 0 & -11 & 8 & 2 \end{pmatrix} \underset{r_3-\frac{11}{10}r_2}{\sim} \begin{pmatrix} 1 & 4 & -2 & 3 \\ 0 & -10 & 5 & -5 \\ 0 & 0 & 2.5 & 7.5 \end{pmatrix}$$

We observe $\begin{pmatrix} 1 & 4 & -2 & 3 \\ 0 & -10 & 5 & -5 \\ 0 & 0 & 2.5 & 7.5 \end{pmatrix}$ has rank 3. Since $\begin{pmatrix} 1 & 4 & -2 & 3 \\ 2 & -2 & 1 & 1 \\ 3 & 1 & 2 & 11 \end{pmatrix}$, $\begin{pmatrix} 1 & 4 & -2 & 3 \\ 0 & -10 & 5 & -5 \\ 0 & 0 & 2.5 & 7.5 \end{pmatrix}$ are row equivalent. Thus $\begin{pmatrix} 1 & 4 & -2 & 3 \\ 2 & -2 & 1 & 1 \\ 3 & 1 & 2 & 11 \end{pmatrix}$, $\begin{pmatrix} 1 & 4 & -2 & 3 \\ 0 & -10 & 5 & -5 \\ 0 & 0 & 2.5 & 7.5 \end{pmatrix}$ both have rank 3.

This result gives us a practical method of finding the rank of a matrix – we reduce it to echelon form.

The idea of rank enables us to determine the existence and uniqueness or otherwise of solutions of systems of linear equations.

Theorem

Suppose we have a system of m equations in n unknowns represented by $A\underline{x} = \underline{b}$ with augmented matrix $B = (A \vdots b)$.

The equations have:

(a) a unique solution *iff* rank A = rank B = n;

(b) an infinite number of solutions *iff* rank A = rank B < n;

(c) no solution *iff* rank A < rank B.

Proof:

We perform elementary row operations to reduce B to echelon form (Gaussian elimination).

The results will be illustrated in the case of 5 (=m) equations in 4 (=n) unknowns.

(a) $\begin{pmatrix} x & x & x & x & . & x \\ 0 & x & x & x & . & x \\ 0 & 0 & x & x & . & x \\ 0 & 0 & 0 & x & . & x \\ 0 & 0 & 0 & 0 & . & 0 \end{pmatrix}$, x – denotes a non-zero element

Then rank A = rank B = n (= 4 in this case) and the system has a unique solution.

(b) $\begin{pmatrix} x & x & x & x & . & x \\ 0 & x & x & x & . & x \\ 0 & 0 & x & x & . & x \\ 0 & 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & . & 0 \end{pmatrix}$

Here rank A = rank B < n (rank A = rank B = 3), x_4 may be assigned arbitrarily and x_1, x_2 and x_3 determined in terms of x_4. Hence the solution is not unique. More generally if rank A = rank B = $r < n$ then $n - r$ unknowns may be assigned arbitrarily.

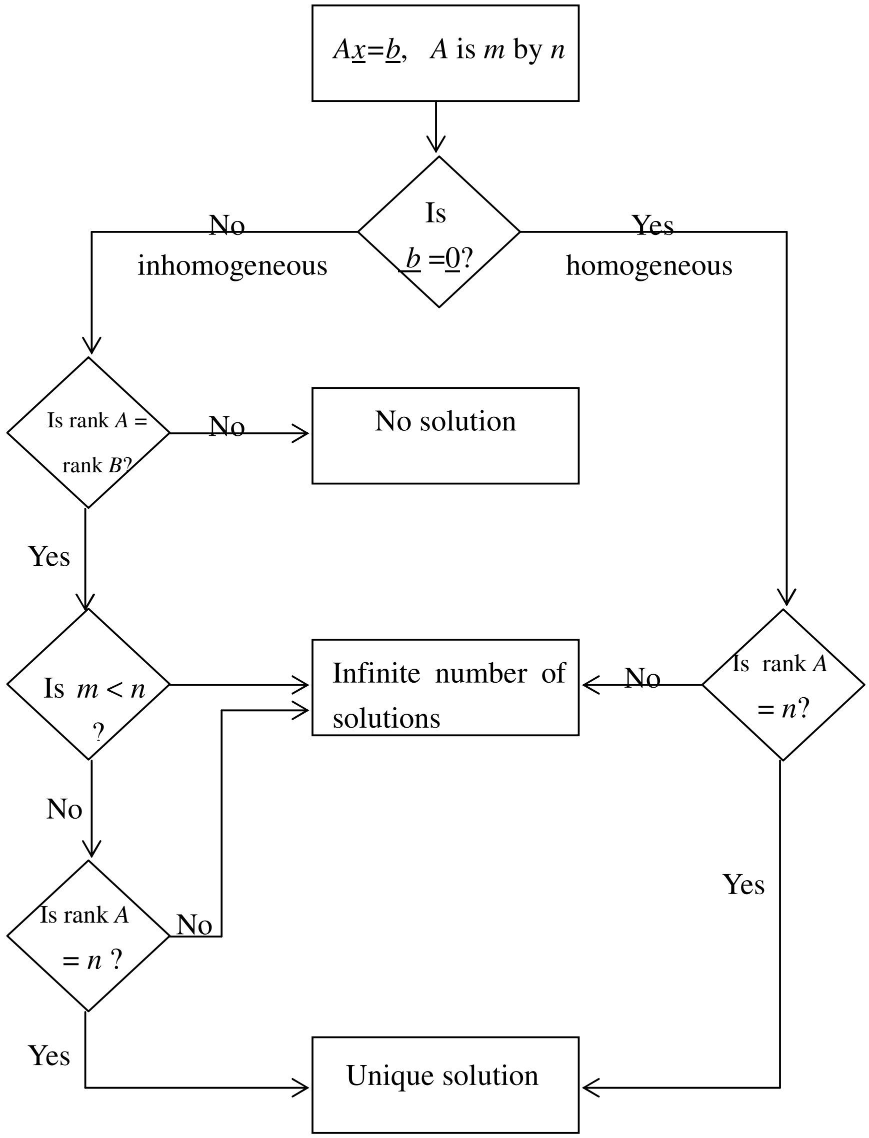
(c) $\begin{pmatrix} x & x & x & x & . & x \\ 0 & x & x & x & . & x \\ 0 & 0 & 0 & 0 & . & x \\ 0 & 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & . & 0 \end{pmatrix}$

Here rank A < rank B (2 < 3) and the system is inconsistent.

Homogeneous Equations

(i) a homogeneous system ($\underline{b} = \underline{0}$) always has rank A = rank B and the trivial solution $\underline{x} = \underline{0}$.
If rank $A < n$ this solution is not unique.

(ii) If there are fewer equations than unknowns ($m < n$) a homogeneous system always has non-trivial solutions, since rank A = rank $B \leq m < n$.



Example

Consider the system $\begin{cases} x_1 + 2x_2 - 3x_3 = -1 \\ 3x_1 - x_2 + 2x_3 = 7 \\ 5x_1 + 3x_2 + ax_3 = b \end{cases}$ for various values of a and b.

$$\begin{pmatrix} 1 & 2 & -3 & . & -1 \\ 3 & -1 & 2 & . & 7 \\ 5 & 3 & a & . & b \end{pmatrix} \begin{matrix} r_1 \\ r_2 \\ r_3 \end{matrix} \underset{\substack{r_2 - 3r_1 \\ r_3 - 5r_1}}{\sim} \begin{pmatrix} 1 & 2 & -3 & . & -1 \\ 0 & -7 & 11 & . & 10 \\ 0 & -7 & a+15 & . & b+5 \end{pmatrix} \begin{matrix} r_1 \\ r_{2a} \\ r_{2b} \end{matrix} \underset{r_{3a} - r_{2a}}{\sim} \begin{pmatrix} 1 & 2 & -3 & . & -1 \\ 0 & -7 & 11 & . & 10 \\ 0 & 0 & a+4 & . & b-5 \end{pmatrix} \begin{matrix} r_1 \\ r_{2a} \\ r_{3b} \end{matrix}$$

(i) If $a \neq -4$, then rank A = rank B = 3 and there is a unique solution.

i.e. $a = 0, b = 9$ gives $x_3 = 1, x_2 = 1/7, x_1 = 12/7$

(ii) If $a = -4$ and $b = 5$, then rank A = rank B = 2 and there are infinitely many solutions.

$x_3 = k$ (say), $x_2 = (11k - 10)$, $x_1 = (13 - k)/7$.

(iii) If $a = -4$ and $b \neq 5$, then rank A = 2 < 3 = rank B and there is no solution.

Efficiency

Cramer's rule may be used for solving $A\underline{x} = \underline{b}$ when n is small, say $n = 2$ or $n = 3$, but is not efficient for large n. If we count the number of multiplications and division required for a method, we have for large n:

Gaussian elimination – about $n^3/3$; Cramer's rule –- about $(n+1)!$

Example

For a 25×25 system of equations (small for engineering and science problems) we have:

Method	Operations	Time on a CRAY 2
Gaussian Elimination	5208	3×10^{-6} secs
Cramer's Rule	4×10^{26}	8×10^{9} years!!!!

Determinants

The fastest way to evaluate a determinant is to reduce the matrix to upper triangular form, U, using Gaussian elimination. Then $\det A = (-1)^r \times \text{product of diagonal elements of } U$ where r = number of row interchanges. Even using this method of evaluation, Cramer's rule requires $(n+1)$ determinants ie $(n+1)$ reductions to upper triangular form, whereas Gaussian elimination requires only one reduction followed by a relatively cheap back substitution.

4. Inverse Matrix

The inverse of a square $n \times n$ matrix A is denoted by A^{-1} and is an $n \times n$ matrix such that $AA^{-1} = A^{-1}A = I$, where I is the $n \times n$ unit matrix. If A has an inverse, then A is called a non-singular matrix, otherwise A is a singular matrix.

Properties

(i) If A has an inverse it is unique

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

(iii) $(A^T)^{-1} = (A^{-1})^T$

(iv) The inverse of a symmetric matrix is symmetric

(v) $\det(A^{-1}) = (\det(A))^{-1}$

(vi) A^{-1} exists *iff* rank $A = n$

(vii) A^{-1} exists *iff* $\det(A) \neq 0$

(viii) $(A^{-1})^{-1} = A$

Although A^{-1} often appears during the manipulation of matrix expressions, it is not usually the case that A^{-1} is needed explicitly. For example, in the solution of the linear equations $A\underline{x} = \underline{b}$. The solution $\underline{x}$ may be expressed as $\underline{x} = A^{-1}\underline{b}$. However $\underline{x}$ is actually computed using Gaussian elimination and we never need to know A^{-1}.

If A^{-1} is needed, it may be found by performing elementary row operations simultaneously on A and I. We are seeking the matrix $X \quad (= A^{-1})$ such that $AX = I$, $\qquad X = A^{-1}I = A^{-1}$. Thus the columns of X are the solutions of $A\underline{x} = \underline{i}_j$, where $\underline{i}_j$ is the j^{th} column of the unit matrix I. Hence we simultaneously solve n sets of linear equations, each with the same matrix A but with a different column of I as the right-hand-side vector.

<u>Example</u>

Find the inverse of $A = \begin{pmatrix} 3 & -1 & 1 \\ -15 & 6 & -5 \\ 5 & -2 & 2 \end{pmatrix}$.

Solution:

$$\begin{array}{c} \qquad A \qquad\qquad\qquad I \\ \left(\begin{array}{ccccccc} 3 & -1 & 1 & . & 1 & 0 & 0 \\ -15 & 6 & -5 & . & 0 & 1 & 0 \\ 5 & -2 & 2 & . & 0 & 0 & 1 \end{array}\right)\begin{array}{l} r_1 \\ r_2 \\ r_3 \end{array} \end{array} \underset{\substack{r_2+5r_1 \\ r_3-5r_1/3}}{\sim} \left(\begin{array}{ccccccc} 3 & -1 & 1 & . & 1 & 0 & 0 \\ 0 & 1 & 0 & . & 5 & 1 & 0 \\ 0 & -1/3 & 1/3 & . & -5/3 & 0 & 1 \end{array}\right)\begin{array}{l} r_1 \\ r_{2a} \\ r_{3a} \end{array}$$

$$\underset{r_{3a}+r_{2a}/3}{\sim} \left(\begin{array}{ccccccc} 3 & -1 & 1 & . & 1 & 0 & 0 \\ 0 & 1 & 0 & . & 5 & 1 & 0 \\ 0 & 0 & 1/3 & . & 0 & 1/3 & 1 \end{array}\right)\begin{array}{l} r_1 \\ r_{2a} \\ r_{3b} \end{array}$$

We may now carry out 3 separate back substitutions to find the three columns of A^{-1}. Alternatively, but equivalently, we may continue to use row operations to further reduce A to the unit matrix I as follows:

$$\underset{r_1-3r_{3b}}{\sim} \left(\begin{array}{ccccccc} 3 & -1 & 0 & . & 1 & -1 & -3 \\ 0 & 1 & 0 & . & 5 & 1 & 0 \\ 0 & 0 & 1/3 & . & 0 & 1/3 & 1 \end{array}\right)\begin{array}{l} r_{1a} \\ r_{2a} \\ r_{3a} \end{array} \underset{r_{1a}+r_{2a}}{\sim} \left(\begin{array}{ccccccc} 3 & 0 & 0 & . & 6 & 0 & -3 \\ 0 & 1 & 0 & . & 5 & 1 & 0 \\ 0 & 0 & 1/3 & . & 0 & 1/3 & 1 \end{array}\right)\begin{array}{l} r_{1b} \\ r_{2b} \\ r_{3b} \end{array} \underset{\substack{r_{1b}/3 \\ 3r_{3b}}}{\sim}$$

$$\begin{array}{c} \quad I \qquad\qquad A^{-1} \\ \left(\begin{array}{ccccccc} 1 & 0 & 0 & . & 2 & 0 & -1 \\ 0 & 1 & 0 & . & 5 & 1 & 0 \\ 0 & 0 & 1 & . & 0 & 1 & 3 \end{array}\right) \end{array}$$

The method of finding the solutions with which the matrix will be row reduced into reduced row echelon form is called **Gauss Jordan Method**.

Check $AA^{-1} = A^{-1}A = I$. Note if

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad A^{-1} = \frac{1}{\det A}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \text{ and if } A = \begin{pmatrix} a_{11} & & & & \\ & a_{22} & & O & \\ & & . & & \\ & O & & . & \\ & & & & a_{nn} \end{pmatrix}, \text{ a diagonal matrix}$$

with $a_{11}, a_{22}, \cdots, a_{nn} \neq 0$, then $A^{-1} = \begin{pmatrix} 1/a_{11} & & & & \\ & 1/a_{22} & & O & \\ & & . & & \\ & O & & . & \\ & & & & 1/a_{nn} \end{pmatrix}$

<u>Example</u>

Let $A = \begin{pmatrix} 1 & 1 & 3 \\ -2 & -2 & -3 \\ 3 & 1 & 4 \end{pmatrix}$. Find A^{-1} using elementary row operations to reduce A to I. (i.e. to use the Gauss Jordan Method.)

Solution:

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 3 & 1 & 0 & 0 \\ -2 & -2 & -3 & 0 & 1 & 0 \\ 3 & 1 & 4 & 0 & 0 & 1 \end{array}\right) \underset{\substack{r_2+2r_1 \\ r_3-3r_1}}{\sim} \left(\begin{array}{ccc|ccc} 1 & 1 & 3 & 1 & 0 & 0 \\ 0 & 0 & 3 & 2 & 1 & 0 \\ 0 & -2 & -5 & -3 & 0 & 1 \end{array}\right) \underset{r_2 \leftrightarrow r_3}{\sim} \left(\begin{array}{ccc|ccc} 1 & 1 & 3 & 1 & 0 & 0 \\ 0 & -2 & -5 & -3 & 0 & 1 \\ 0 & 0 & 3 & 2 & 1 & 0 \end{array}\right)$$

$$\underset{-\frac{1}{2}r_2}{\sim} \left(\begin{array}{ccc|ccc} 1 & 1 & 3 & 1 & 0 & 0 \\ 0 & 1 & 5/2 & 3/2 & 0 & -1/2 \\ 0 & 0 & 3 & 2 & 1 & 0 \end{array}\right) \underset{\frac{1}{3}r_3}{\sim} \left(\begin{array}{ccc|ccc} 1 & 1 & 3 & 1 & 0 & 0 \\ 0 & 1 & 5/2 & 3/2 & 0 & -1/2 \\ 0 & 0 & 1 & 2/3 & 1/3 & 0 \end{array}\right)$$

$$\underset{\substack{r_2-\frac{5}{2}r_3 \\ r_1-3r_3}}{\sim} \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 0 & -1/6 & -5/6 & -1/2 \\ 0 & 0 & 1 & 2/3 & 1/3 & 0 \end{array}\right) \underset{r_1-r_2}{\sim} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -5/6 & -1/6 & 1/2 \\ 0 & 1 & 0 & -1/6 & -5/6 & -1/2 \\ 0 & 0 & 1 & 2/3 & 1/3 & 0 \end{array}\right)$$

So

$$A^{-1} = \begin{pmatrix} -5/6 & -1/6 & 1/2 \\ -1/6 & -5/6 & -1/2 \\ 2/3 & 1/3 & 0 \end{pmatrix}$$

5. Block Matrices

It can be convenient to partition a matrix into submatrices by horizontal and vertical lines as illustrated by:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & . & a_{14} & . & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & . & a_{24} & . & a_{25} & a_{26} \\ . & . & . & . & . & . & . & . \\ a_{31} & a_{32} & a_{33} & . & a_{34} & . & a_{35} & a_{36} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix}$$

where $A_{11} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$, $A_{22} = (a_{34})$, etc.

All the usual matrix operations of addition, multiplication etc may be performed on partitioned matrices as if the submatrices were single elements, as long as the matrices are partitioned such that it is permissible to form the various operations.

Example

If A is as above, $A = \begin{array}{c c} & \begin{array}{ccc} 3 & 1 & 2 \end{array} \\ \begin{array}{c} 2 \\ 1 \end{array} & \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix} \end{array}$

(1,2,3 – denote numbers of rows/columns in each submatrix) then A may be added to a (3×6) matrix B partitioned in the same way or multiplied by a $(6 \times n)$ matrix C partitioned as

$$C = \begin{array}{c c} \begin{array}{c} 3 \\ 1 \\ 2 \end{array} & \begin{pmatrix} C_{11} \\ C_{21} \\ C_{31} \end{pmatrix} \end{array}, \quad AC = \begin{array}{c c} \begin{array}{c} 2 \\ 1 \end{array} & \begin{pmatrix} A_{11}C_{11} + A_{12}C_{21} + A_{13}C_{31} \\ A_{21}C_{11} + A_{22}C_{21} + A_{23}C_{31} \end{pmatrix} \end{array}$$

In fact C may be partitioned vertically in any way to give a similar vertical partition of AC.

Example

$$A=\begin{pmatrix}1 & 2 & . & 5\\ 3 & 4 & . & 6\end{pmatrix}=\begin{pmatrix}A_{11} & A_{12}\end{pmatrix} \quad , \; B=\begin{pmatrix}6 & 7 & . & 8\\ 4 & 5 & . & 9\\ . & . & . & .\\ 1 & 2 & . & 3\end{pmatrix}=\begin{pmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{pmatrix} \text{ then}$$

$$\begin{aligned}AB &= \left(A_{11}B_{11}+A_{12}B_{21} : A_{11}B_{12}+A_{12}B_{22}\right)\\ &= \left(\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}\begin{pmatrix}6 & 7\\ 4 & 5\end{pmatrix}+\begin{pmatrix}5\\ 6\end{pmatrix}\begin{pmatrix}1 & 2\end{pmatrix} \begin{matrix}.\\ .\end{matrix} \begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}\begin{pmatrix}8\\ 9\end{pmatrix}+\begin{pmatrix}5\\ 6\end{pmatrix}\begin{pmatrix}3\end{pmatrix}\right)\\ &= \left(\begin{pmatrix}14 & 17\\ 34 & 41\end{pmatrix}+\begin{pmatrix}5 & 10\\ 6 & 12\end{pmatrix} \begin{matrix}.\\ .\end{matrix} \begin{pmatrix}26\\ 60\end{pmatrix}+\begin{pmatrix}15\\ 18\end{pmatrix}\right)=\begin{pmatrix}19 & 27 & . & 41\\ 40 & 53 & . & 78\end{pmatrix}\end{aligned}$$

Least Squares Approximations

In a system $A\underline{x} = \underline{b}$, that is,

$$\begin{bmatrix}1 & 0\\ 5 & 4\\ 2 & 4\end{bmatrix}\begin{bmatrix}u\\ v\end{bmatrix}=\begin{bmatrix}b_1\\ b_2\\ b_3\end{bmatrix},$$

or

$$u\begin{bmatrix}1\\ 5\\ 2\end{bmatrix}+v\begin{bmatrix}0\\ 4\\ 4\end{bmatrix}=\begin{bmatrix}b_1\\ b_2\\ b_3\end{bmatrix}$$

we would like to find the solution u, v. Note that this system $A\underline{x} = \underline{b}$ is solvable if and only if $\underline{b}$ can be expressed as a linear combination of the columns of A.

A $m \times n$ system $A\underline{x} = \underline{b}$ with $m > n$ either has a solution or not. In practice, inconsistent equations arise and have to be solved. One possibility is to determine $\underline{x}$ from a part of the system, and ignore the rest; this is hard to justify if all equations of the system come from the same source. Rather than expecting no error in some equations and large errors in the others, it is more reasonable to choose $\underline{x}$ so as to minimize the average error in all the equations.

Consider the system of equations $A\underline{x} = \underline{b}$, where A is a $m \times n$ matrix, $\underline{x} \in \Re^n$ and $\underline{b} \in \Re^m$. We can form a residual

$$\underline{r} = \underline{b} - A\underline{x}.$$

The distance between $\underline{b}$ and $A\underline{x}$ is given by

$$\| \underline{b} - A\underline{x} \| = \| \underline{r} \|.$$

We wish to find a vector $\underline{x}$ for which $\underline{r}$ will be a minimum. In other words, $\hat{\underline{x}}$ is a least square solution if $\| A\hat{\underline{x}} - b \| \leq \| A\underline{x} - b \| \forall \underline{x} \in \Re^n$. This is equivalent to minimizing $\| \underline{r} \|^2$.

We now use a geometric approach to find a least squares solution to $A\underline{x} = \underline{b}$ and let $A\underline{x}$ lies on a plane W. If our equations $A\underline{x} = \underline{b}$ are consistent, then $\underline{x}$ has exact solution and $\underline{b}$ lies on W. However, we now have an inconsistent set $A\underline{x} = \underline{b}$. If $\underline{c} = A\hat{\underline{x}}$, then $\underline{c}$ is a vector in W that is closest to $\underline{b}$. As shown in Figure 1,

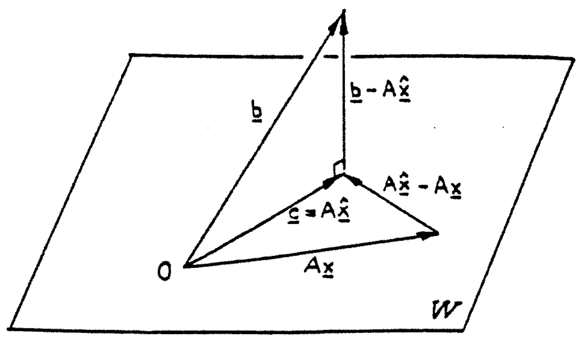


Figure 1

To make $\underline{b} - A\hat{x}$ to be the shortest distance from the plane W, $\underline{b} - A\hat{\underline{x}}$ must be orthogonal (perpendicular) to all vectors $A\hat{\underline{x}} - A\underline{x}$ in W.

Now
$$(A\hat{\underline{x}} - A\underline{x}) \cdot (A\hat{\underline{x}} - \underline{b}) = 0$$
$$\Rightarrow \quad (A\hat{\underline{x}} - A\underline{x})^T (A\hat{\underline{x}} - \underline{b}) = 0$$
$$\Rightarrow \quad (\hat{\underline{x}} - \underline{x})^T A^T (A\hat{\underline{x}} - b) = 0 \qquad \text{for all } \underline{x}.$$

Since this equation must hold for all $\underline{x}$, since $\underline{e}_i = \hat{\underline{x}} - \underline{x}$ which is non-zero so that

$$\underline{x} = \hat{\underline{x}} - \underline{e}_i \text{ for } i = 1, \ldots, n.$$

Now $\hat{\underline{x}}$ is a least squares solution if and only if

$$A^T (A\hat{\underline{x}} - \underline{b}) = \underline{0} ,$$
$$\Leftrightarrow \quad A^T A\hat{\underline{x}} = A^T \underline{b}. \tag{1}$$

The equations of the system (1) are called the normal equations for $A\underline{x} = \underline{b}$.

Theorem

Let A be an $m \times n$ matrix. The least squares solutions to

$$A\underline{x} = \underline{b}$$

are precisely the solutions to the system of normal equations

$$A^T A\underline{x} = A^T\underline{b}.$$

Further, this system has at least one solution.

Proof (For reference only)

It can be shown that

$$\text{rank}\,[A^TA] = \text{rank}\,[A^TA \mid A^T\underline{b}]$$

Thus, $A^TA\underline{x} = A^T\underline{b}$ always has a solution, and we only need to show that the set of least squares solutions is identical with the set of solutions to the normal equations.

We can also prove that if $\underline{z}\cdot\underline{w} = 0$, then

$$\| \underline{z} - \underline{w} \|^2 = \| \underline{z} \|^2 + \| \underline{w} \|^2$$

Now let $\underline{u}$ be a solution to $A^TA\underline{x} = A^T\underline{b}$. Let $\underline{y}$ be any vector in $\Re^n$, and set

$$\underline{z} = A\underline{y} - A\underline{u} \text{ and } \underline{w} = \underline{b} - A\underline{u}.$$

Then
$$\underline{z}\cdot\underline{w} = (A\underline{y} - A\underline{u})^T\,(\underline{b} - A\underline{u})$$

$$= (\underline{y} - \underline{u})^T A^T (\underline{b} - A\underline{u})$$

$$= (\underline{y} - \underline{u})^T\,\underline{0}$$

$$= 0.$$

Hence,

$$\| Ay - \underline{b} \|^2 = \| A\underline{y} - A\underline{u} \|^2 + \| \underline{b} - A\underline{u} \|^2 \geq \| \underline{b} - A\underline{u} \|^2$$

Since this results holds for any $y \in \Re^n$, $\underline{u}$ is a least squares solution to $A\underline{x} = \underline{b}$.

On the other hand, let $\underline{x}_0$ be a least squares solution to $A\underline{x} = \underline{b}$ and let $\underline{v}$ be any solution to $A^TA\underline{x} = A^T\underline{b}$.

Then, as above,

$$\| A\underline{x}_0 - \underline{b} \|^2 = \| A\underline{x}_0 - A\underline{v} \|^2 + \| \underline{b} - A\underline{v} \|^2 \geq \| A\underline{v} - \underline{b} \|^2.$$

This means that $\| A\underline{x}_0 - A\underline{v} \|^2 = 0$,

or $A\underline{x}_0 = A\underline{v}$.

Hence $A^TA\underline{x}_0 = A^TA\underline{v} = A^T\underline{b}$

and $\underline{x}_0$ is a solution to the normal equations. □

Example To find the least squares solutions to

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \underline{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

we form the system of normal equations

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \underline{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

or

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \underline{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Solving this equation yields $\underline{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ as the least squares solution. Further, as depicted in Figure 2, the column space W is the x_1, x_3 – plane and

$$A\underline{x} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

is the point of W closest to $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

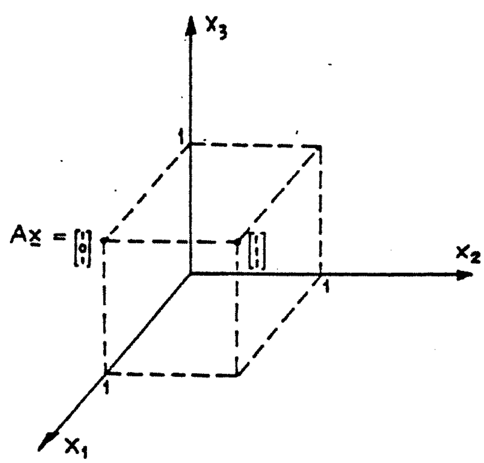


Figure 2

Least Squares Polynomial Approximations to given data

Suppose we obtain a set of data and we estimate that the data should theoretically be a polynomial P such that

$$P(x) = a_n x^n + \cdots + a_1 x + a_0$$

for some choice of a_0, a_1, ..., a_n. We define the errors between polynomial values $P(x_1), \ldots, P(x_n)$ and data values $y_1, \ldots, y_m$ as

$$\begin{aligned} e_1 &= P(x_1) - y_1 = a_n x_1^n + \ldots + a_1 x_1 + a_0 - y_1 \\ &\vdots \\ e_m &= P(x_m) - y_m = a_n x_m^n + \ldots + a_1 x_m + a_0 - y_m \end{aligned}$$

Such errors are illustrated in Figure 3.

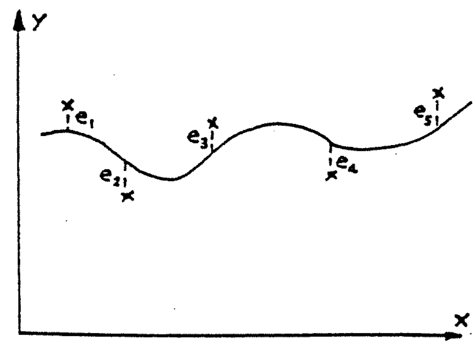


Figure 3

For

$$\underline{r} = \begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix}, A = \begin{bmatrix} 1 & x_1 & \cdots & x_1^n \\ \vdots & \vdots & & \vdots \\ 1 & x_m & & x_m^n \end{bmatrix}, \underline{x} = \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix}, \underline{b} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix},$$

the e equations can be written as

$$\underline{r} = A\underline{x} - \underline{b}.$$

A least square polynomial approximation to the points will be a polynomial for which the error

$$(e_1^2 + \ldots + e_m^2)^{1/2} = \| \underline{r} \|$$

is smallest. Such a polynomial can be found by finding a least squares solution to

$$A\underline{x} = \underline{b}.$$

Using the normal equations, we have

$$A^T A\underline{x} = A^T \underline{b}$$

i.e.

$$\begin{bmatrix} m & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^n \\ & \sum x_i^2 & \sum x_i^3 & & \sum x_i^{n+1} \\ & & \sum x_i^4 & & \vdots \\ & & & \ddots & \\ & \text{sym.} & & & \sum x_i^{2n} \end{bmatrix} \underline{x} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \vdots \\ \sum x_i^n y_i \end{bmatrix}$$

Example

Find the least squares polynomial approximation of degree 1 or less to the points [0, 1], [1, 1.4], and [2, 1.9].

We need to find $P(x) = a_1 x + a_0$ using least squares. To find this polynomial, we set

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.4 \\ 1.9 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1.4 \\ 1.9 \end{bmatrix}$$

or

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 4.3 \\ 5.2 \end{bmatrix}.$$

Solving this system gives $a_0 = 0.98$ and $a_1 = 0.45$ to 2 decimal places. So

$$P(x) = 0.45x + 0.98. \qquad \blacksquare$$

Example

Find the least squares polynomial $P(x) = a_2 x^2 + a_1 x + a_0$ that approximates the data [-1, 1], [0, 0], [1, 1] and [2, 5]. To find this polynomial, we set

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 5 \end{bmatrix}$$

The system of normal equations gives

$$\begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 10 \\ 22 \end{bmatrix}$$

which yields $a_0 = -0.15$, $a_1 = 0.05$ and $a_2 = 1.25$. Thus

$$P(x) = 1.25x^2 + 0.05x - 0.15\,.$$

MA0136/MA2170.doc/sk

27.8.02