In this section

- Polar Coordinates in $\mathbb{R}^2$
- Spherical Polar Coordinates $\mathbb{R}^3$
- Cylindrical Polar Coordinates $\mathbb{R}^3$
- Surfaces
- Partial Differentiation
- Tangent Planes
- Gradient
- Divergence
- Curl

Half-side formulas

Suppose that α, β, and γ are the angles of a spherical triangle. Denote by a the side of the triangle that is opposite α, and put $S=(\alpha+\beta+\gamma)/2$ and:

$$R=\sqrt{\frac{-\cos S}{\cos(S-\alpha)\cos(S-\beta)\cos(S-\gamma)}}$$

The formulas

$$\sin\left(\tfrac{1}{2}a\right)=\sqrt{\frac{-\cos S\cos(S-\alpha)}{\sin\beta\sin\gamma}}$$

$$\cos\left(\tfrac{1}{2}a\right)=\sqrt{\frac{\cos(S-\beta)\cos(S-\gamma)}{\sin\beta\sin\gamma}}$$

and $\tan\left(\frac{1}{2}a\right)=R\cos(S-\alpha)$

are called **half-side formulas**.

GENERAL SOLUTION The general solution of a problem is a solution involving a certain number of parameters from which any other solution (except for singular solutions) can be obtained, through a suitable choice of these parameters.

Specifically, in analogy to differential equations, one understands, by a general solution of a non-homogeneous linear difference equation

$$① \sum_{i=0}^{n} p_i(x)\, y(x+i) = q(x)$$

a solution in the form

$$y(x)=\sum_{i=1}^{n} c_i(x)\phi_i(x)+\psi(x)$$

where $\phi(x)$, $i=1 \ldots n$ are linearly independent solutions of the corresponding homogeneous equation. $\psi(x)$ is the non-homogeneous arbitary solution of ①, and $c_i(x)$ are arbitary periodic functions of period 1.

Non-rectangular Coordinate Systems and Surfaces | Vectors

Although up to this point we have considered only the usual rectilinear system of axes and coordinates, there are situations where a different coordinate system might be more appropriate, or more convenient.

We shall consider one other system in $\mathbb{R}^2$ and two in $\mathbb{R}^3$. But there are others which can be researched.

Polar Coordinates in $\mathbb{R}^2$

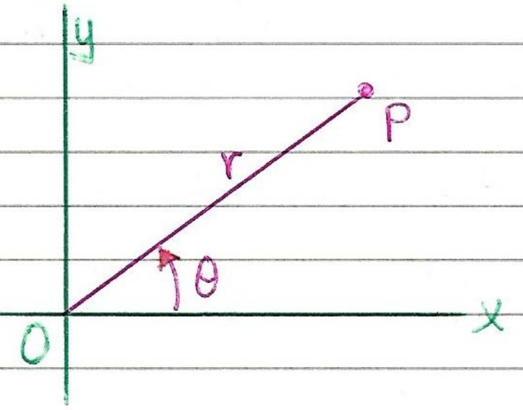
For polar coordinates in $\mathbb{R}^2$ we choose an origin 'O' together with a directed initial line through O. It is often convenient to think of the initial line as the x-axis (of a corresponding rectangular system) - but this is not strictly necessary.

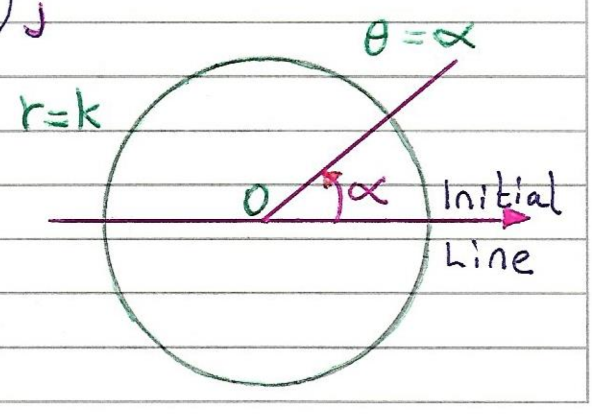
The polar coordinates of a point P in the plane are (r, θ). Where r denotes the distance from the origin, and θ denotes the angle turned through anti-clockwise from the initial line.

We choose $\mathbf{i}$ to be a unit vector in the direction of the initial line, and $\mathbf{j}$ to be a second unit vector defined so that turning from $\mathbf{i}$ through a right angle anti-clockwise would give the direction of $\mathbf{j}$.

Here $\mathbf{r}$ denotes the position vector of a point in the same way as in previous examples. Writing $|\mathbf{r}| = r$, we have:

$$\mathbf{r} = (r\cos\theta)\mathbf{i} + (r\sin\theta)\mathbf{j}$$





With coordinate systems, it is useful to consider what we get if we keep one of the variables constant, and let the other(s) range over all possible values.

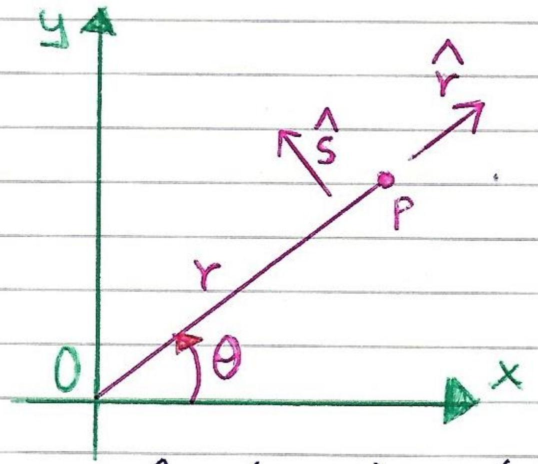
In $\mathbb{R}^2$ with the usual rectangular axes, if c is constant, the equation $x=c$ defines a lines, as does $y=c$. And similary in $\mathbb{R}^3$, $x=c$ defines a plane.

In $\mathbb{R}^2$ with polar coordinates as described previously $\mathbf{r} = (r\cos\theta)\mathbf{i} + (r\sin\theta)\mathbf{j}$. $r=c$ defines a circle for $c>0$, and $\theta=\alpha$ defines a half-line making an angle α with the initial line, as in the diagram at the bottom of the previousl page.

It is possible to allow r to be negative as well as positive. In this case $r=c$ denotes exactly the same circle as before, but $\theta=\alpha$ becomes the whole line, rather than just the half-line. Negative r would give all points on the line on the opposite side of the origin.

There are good points to both systems, but with vectors, because we define $r = |\mathbf{r}| \geq 0$, the most appropriate is the first.

Example



Suppose we set two vectors $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$ as the unit vectors in the radial and transverse directions respectivly.

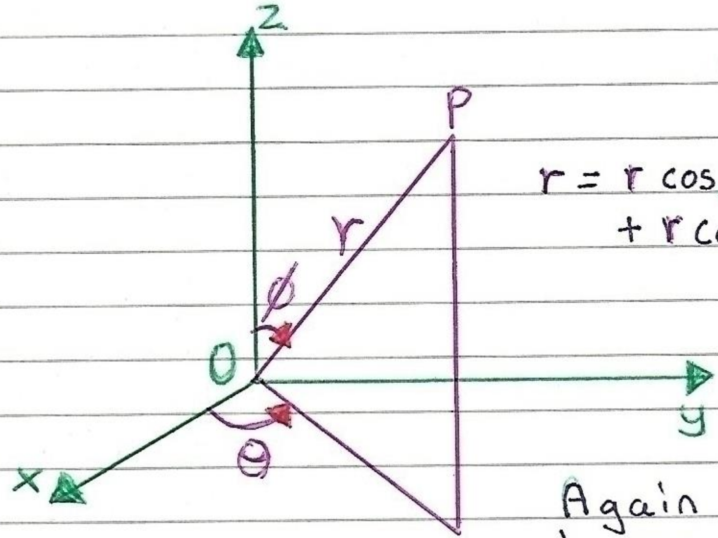
Then $\hat{\mathbf{r}} = (\cos\theta)\mathbf{i} + (\sin\theta)\mathbf{j}$

and

$\hat{\mathbf{s}} = (-\sin\theta)\mathbf{i} + (\cos\theta)\mathbf{j}$

Spherical Polar Coordinates in $\mathbb{R}^3$

In spherical polars we need to choose an origin O, and a right-handed set of rectangular axes. As in the two-dimensional case, $r = |\mathbf{r}|$ where $\mathbf{r}$ is the position vector of the point P, θ is the angle from the plane $y=0$, to the plane containing both P and the z-axis. Finally ϕ is the angle between OP and the z-axis as shown below.



- So here we have:

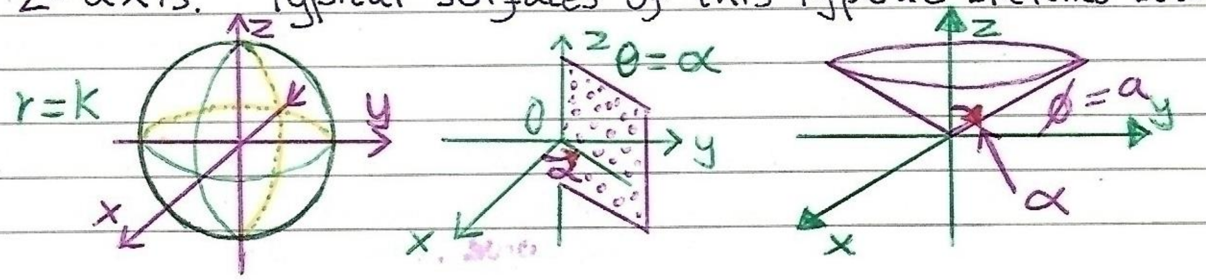
$$\mathbf{r} = r\cos\theta\sin\phi\,\mathbf{i} + r\sin\theta\sin\phi\,\mathbf{j} + r\cos\phi\,\mathbf{k}$$

Again we consider what happens when one of the three variables is constant.

Suppose k a α are fixed real numbers. Then $r = k$ ($k > 0$) denotes a sphere of radius k, which is why we refer to coordinates in this system as 'spherical polars'.

If $k = 0$, then the 'sphere' is of zero radius, and so becomes a single point.

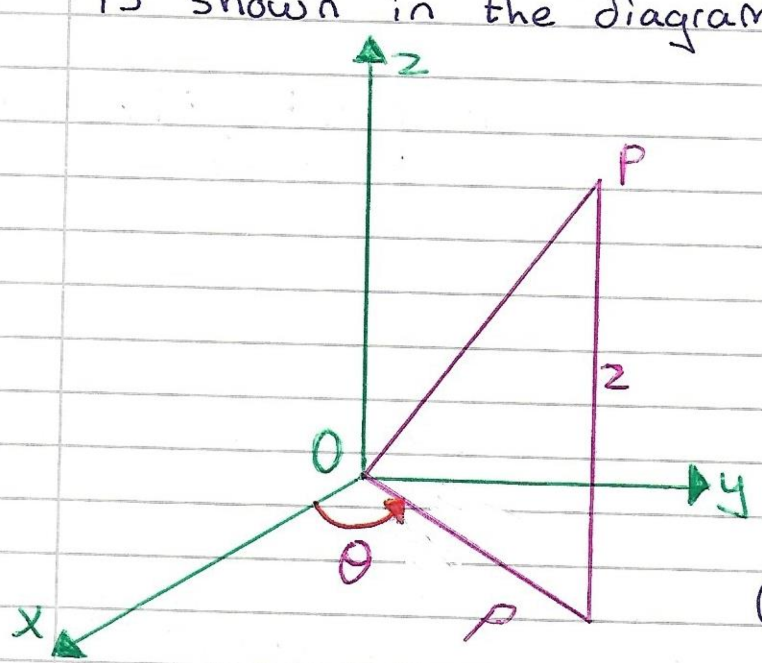
$\theta = \alpha$ is the equation of a half-plane containing the z-axis and a line in the plane $z = 0$, making an angle α with the x-axis as shown, and $\phi = \alpha$ a half-cone with vertex at the origin, and its axis along the z-axis. Typical surfaces of this type are sketched below.



Cylindrical Polar Coordinates in $\mathbb{R}^3$

Again for cylindrical polars we need to choose an origin O, and a right handed system of rectangular axes, as in the spherical polar case.

Then the coordinates of a general point P in this system are (ρ, θ, z). Where ρ is the distance of P from the z-axis, θ is the same angle as in the spherical polars, and z gives the usual distance from the plane $z=0$. All this is shown in the diagram below:



Suppose K and α are constant real numbers.

Then $\rho = K \; (K>0)$ denotes a cylinder of radius K. This explains the terminolgy here.

(If $K=0$, we simply get the z-axis)

$\theta = \alpha$ determines a half-plane as in the spherical polar case, and $z=K$ determines a plane perpendicular to the z-axis.

Cylindrical polars combine two-dimensional polar coordinates with a height z. If the 'r' is the position vector of P.

So
$$r = \rho \cos\theta \, i + \rho \sin\theta \, j + zk$$

The helix from the previous pages can be best described using cylindrical polar coordinates. Its equations become

$$\rho = a, \quad z = k\theta, \quad a, k \text{ constant}$$

(A point can be described in terms of a single parameter θ, and so the locus of the curve in the helix)

Surfaces

In the same way that a curve is given by the position vector of a general point on the curve in terms of one parameter. So a surface can be given by the position vector of a general point on the surface in terms of two parameters.

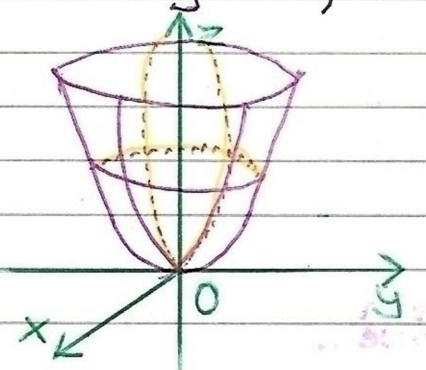
We have already just seen two such surfaces, the sphere and the cylinder. We now explore some examples of surfaces.

Consider for example, the locus of a point whose position vector is given by

$$r = xi + yj + (x^2 + y^2)k$$

The locus (literally being translated from the latin meaning 'place') is the set of all possible positions of that point under the given condition. Or as in this case, satisfying the given equation.

If the coordinates (x, y, z) are those of a point at this locus, the z denotes the height of the locus vertically above the point $(x, y, 0)$.
As x and y vary, so the point $(x, y, 0)$ can range over the whole plane of $z = 0$.
As $(x, y, 0)$ ranges over that plane, the point $(x, y, x^2 + y^2)$ ranges over the surface as shown below. If $k > 0$, the plane $z = k$ cuts the surface in a circle whose equations are given by: $x^2 + y^2 = k$, $z = k$.



whereas if $k = 0$ the intersection is a single point, and if $k < 0$ there is no real intersection.

(continued...)

(continued) If we look at the intersection with the plane $y=0$ and $x=0$ we get the respective parabolas defined by

$$z=x^2,\ y=0 \quad \text{and} \quad z=y^2,\ x=0$$

In these directions the intersection is a parabola, and the whole shape is called a circular paraboloid. Typical intersections of these types are shown in the diagram.

In general if we consider the locus of a point (x,y) in the plane which is connected by a single equation. The locus is a curve, and if we consider the locus of point (x,y,z) in $\mathbb{R}^3$, to be connected by a single equation - we have a surface.

For example: $x^2+y^2=a^2$ $(a\neq 0)$ defines a circle

where as $x^2+y^2+z^2=a^2$ $(a\neq 0)$ defines a sphere (in $\mathbb{R}^3$)

By a sphere we mean a two-dimensional surface. The three-dimensional volume contained within a sphere, is called a ball.

There are exceptions to this general rule, as we can see if we allow $a=0$. In the above equations, since in each of the $\mathbb{R}^2$ and $\mathbb{R}^3$ cases the equation will determine a single point, namely the origin.

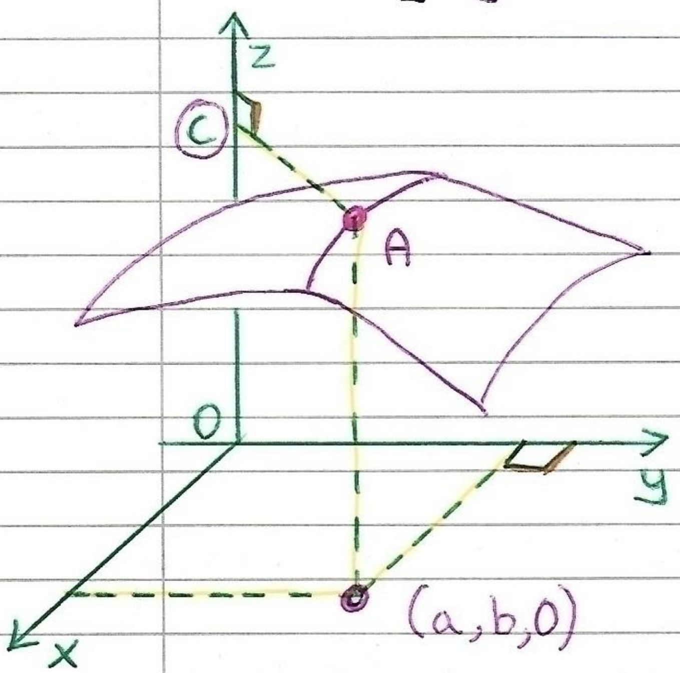
Using vectors often simplifies equations of well known surfaces. The equation of the sphere above could be expressed as $|\mathbf{r}|=a$ or as $\mathbf{r}.\mathbf{r}=a^2$. We have already met the equation of a plane in the form: $(\mathbf{r}-\mathbf{a}).\mathbf{n}=0$. A sphere of radius a centered at the point B, with position vector $\mathbf{b}$, since $|\mathbf{r}-\mathbf{b}|=a$ has the vector equation: $(\mathbf{r}-\mathbf{b}).(\mathbf{r}-\mathbf{b})=a^2$.

Partial Differentiation

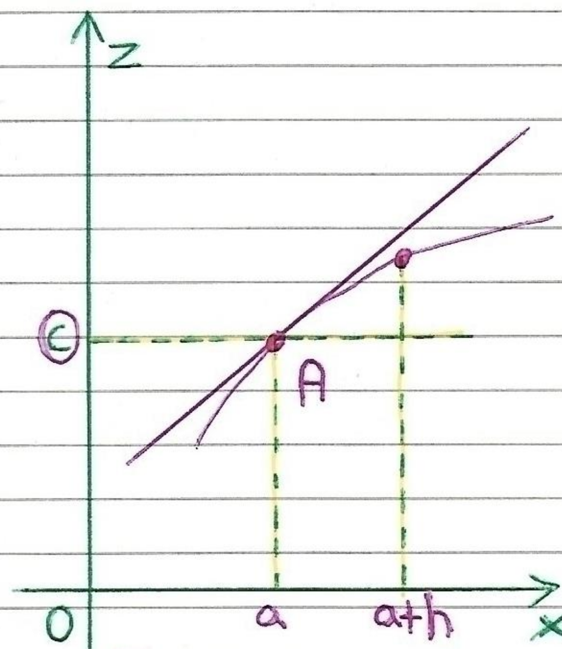
A curve has a tangent line at every point, provided the curve is defined by a differentiable function.
A surface has a tangent plane at every point, again as long as the surface is defined by a differentiable function.
We need to know what this means for surfaces before we can find tangent planes.

- Suppose we have a surface S defined by $z = f(x,y)$; where the height of the surface is given in terms of the x any y coordinates.

Let $A = (a,b,c)$ be a point on this surface (so that $c = f(a,b)$) and consider the intersection of S with the plane $y = b$.
This plane is parallel to the x-axis.
Then the plane $y=b$ cuts S in the curve $y=b$, $z = f(x,b)$. See the 2nd diagram below showing $y=b$, and its intersection with S.



[Diagram 1]



[Diagram 2]

(continued) The tangent line to this curve of intersection is also shown in Diagram 2.

The gradient of this tangent at the point A is

$$\lim_{h \to 0} \frac{f(a+h,b) - f(a,b)}{h}$$

- if this limit exists. Where this limit does exist we use the notation:

$$\lim_{h \to 0} \frac{f(a+h,b) - f(a,b)}{h} = \left.\frac{\partial f}{\partial x}\right|_{(a,b)} = f_x(a,b)$$

We shall use the later, as its more convenient to use with vectors.

Although all the examples used are differentiable. There are those functions in two variables which do not behave so well. In the same way there are functions of a single variable.

An example of such a function would be $y = 1/x$ which goes to infinity as x approaches 0. Another example would be $y = |x|$, which is not differentiable at $x = 0$ (as it is not 'smooth' there)

Similarly if we consider the intersection of S with the plane $x = a$, the gradient to the curve of intersection at the point A is:

$$\lim_{h \to 0} \frac{f(a,b+K) - f(a,b)}{K} = f_y(a,b)$$

If we are finding the partial derivatives at a general point we write f_x rather than $f_x(x,y)$, as this makes the notation simpler. In practice, to find f_x - we simply differentiate $f(x,y)$ with respect to x, treating y as a constant. (to find f_y we use x as the constant.)

Tangent Planes

Because the lines in the planes $x=a$ and $y=a$ which are tangent to the intersections of these respective planes with S, must of necessity be tangent to the surface S. -They must both lie in the tangent plane to S at A.

- Suppose the tangent plane to S at A is π whose equation is

$$px = qy + rz = s$$

then since A lies on π : $pa+qb+rc=s$
- where $c=f(a,b)$

We can now write the equation of π as $p(x-a)+q(y-b)+r(z-c)=0$

This intersects the plane $y=b$ in the line $y=b,\ p(x-a)+r(z-c)=0$

But this is the line $y=b,\ z-c=f_x(a,b)(x-a)$, so $-1:f_x(a,b)=r:p$.

Similary π intersects the plane $x=a$ in the line: $x=a,\ q(y-b)+r(z-c)=0$

which is the line $x=a,\ z-c=f_y(a,b)(y-b)$, so $-1:f_y(a,b)=r:q$. This means that the equation of the tangent plane to S at A is

$$f_x(a,b)(x-a)+f_y(a,b)(y-b)-(z-c)=0.$$

- This also means that the vector normal to the surface at A, is parallel to the vector $n=f_x(a,b)i+f_y(a,b)j-k$

Suppose the position vector of A is a, and that a general point on the surface is r: its vector equation on the tangent plane is $(r-a)\cdot n$.

If we assume f is differentiable, then z can't equal zero, if it did we would have an impossible vertical tangent.

Example

Finding the equation of the tangent plane to the surface defined by

$$z = x^2 + 2y^2 - 6xy + 4$$

at the point on the surface where $x = 0$, $y = 1$.

Solution

If we set $f(x,y) = x^2 + 2y^2 - 6xy + 4$
then $f_x = 2x - 6y$ and $f_y = 4y - 6x$

So that $f_x(0,1) = 0 - 6 = -6$, $f_y(0,1) = 4 - 0 = 4$
and $f(0,1) = 6$.

From the equation on the previous page:
$$f_x(a,b)(x-a) + f_y(a,b)(y-b) - (z-c) = 0$$

the equation of the tangent plane at a given point is
$$-6(x-0) + 4(y-1) - (z-6) = 0$$

which reduces to : $6x - 4y + z + 2 = 0$

- We can deal with surfaces defined by implicit functions, and these may involve some tangent planes parallel to any of the axes of symmetry.
 Suppose the surface is defined by an equation of the form
$$F(x,y,z) = 0$$
We can choose and two variables x, y or z, indepently, and then because of the relation defined by equation $[F(x,y,z) = 0]$ the third variable is dependent upon these two.

(continued) Suppose we choose x any y as the independent variables. Then z is dependent upon both x and y. So, if we differentiate equation $[F(x,y,z)=0]$ partially with respect to x, we get

$$F_x + F_z \frac{\partial z}{\partial x} = 0$$

and similarly, differentiating partially with respect to $y \rightarrow$ $F_y + F_z \dfrac{\partial z}{\partial y} = 0$

and provided $F_z \neq 0$, $\rightarrow \left[\dfrac{\partial z}{\partial x} = \dfrac{-F_x}{F_z} \text{ and } \dfrac{\partial z}{\partial y} = \dfrac{-F_y}{F_z} \right]$

Substituting from

$$\frac{\partial z}{\partial x} = \frac{-F_x}{F_z}, \quad \frac{\partial z}{\partial y} = \frac{-F_y}{F_z}$$

into $f_x(a,b)(x-a) + f_y(a,b)(y-b) - (z-c) = 0$

then multiplying through by F_z. We can write the equation of the tangent plane to S at A as

(Eq $*$) $\rightarrow$ $F_x(a,b,c)(x-a) + F_y(a,b,c)(y-b) + F_z(a,b,c)(z-c) = 0$

Alternatively, if we define ∇F to be the vector $F_x \mathbf{i} + F_y \mathbf{j} + F_z \mathbf{k}$, together with the usual convention for position vectors of A and the general point on S. We have

$$\nabla F(a) \cdot (r - a) = 0$$

The vector $\nabla F(a)$ is normal to the surface S at A, since it is normal (that is, perpendicular) to the tangent plane at A.

Even if $F_z = 0$, 'Eq $*$' above can be used, as, on a differentiable surface, we can't have all three partial derivatives zero at once. If $F_x \neq 0$, we could exchange the rôles of x and z. Taking y, z as independent, and x dependent on both. We end up with 'Eq $*$' anyway.

Example

Suppose a surface S is defined by

$$F(x,y,z) = x^2 + y^2 + z^2 - 9$$

Finding the equation of the tangent plane to S at the point A, whose coordinates are (1,2,2)

Solution

Firsty it is good to check that the point really does lie on S.

So $1^2 + 2^2 + 2^2 - 9 = 0$, so A is on S.

Then $F_x = 2x$, $F_y = 2y$, $F_z = 2z$

and substituting into

$$F_x(a,b,c)(x-a) + F_y(a,b,c)(y-b) + F_z(a,b,c)(z-c) = 0$$

Using the values $x=1$, $y=2$, $z=2$. We get

$$2(x-1) + 4(y-2) + 4(z-2) = 0$$

which simplified becomes: $\underline{x + 2y + 2z = 9}$

Now in this example, equation $F(x,y,z)=0$ defines a sphere with a center at the origin with radius 3.

As a check that we have arrived at the correct answer, we can note that a vector normal to the tangent plane is $i + 2j + 2k$. Which is parallel to the position vector of A. But since the center of the sphere lies at the origin, and the position of vector A lies along a radius of the sphere. It must be perpendicular to the tangent plane at the point of contact 'A'. Verifing the answer.

Gradient, Divergence and Curl

We have referred to the mechanical implications of properties of vectors many times.
To go further into this subject would be to look at three different types of derivative - grad, div, and curl.
Grad was already encounteed when we used it in the vector equation of a tangent plane to a surface at a given point.

The three derivatives are very much related, and depend upon what is called - the vector differential operator ∇. Which we call del (though some texts refer to it as nabla). It can be defined as:

$$\nabla = \frac{\partial}{\partial x} i + \frac{\partial}{\partial y} j + \frac{\partial}{\partial z} k$$

This can act on scalar or vector functions, giving the three types of derivative: Divergence, Gradient and Curl.

Gradient

Suppose $\phi : A \to \mathbb{R}$ is a differentiable function, where A is a region of $\mathbb{R}^3$ (so that $A \subseteq \mathbb{R}^3$ and at each point (x, y, z) of A the value of the function $\phi(x, y, z)$ is a real number). Then the gradient of ϕ - written gradϕ or $\nabla\phi$ is defined as

$$\nabla\phi = \frac{\partial\phi}{\partial x} i + \frac{\partial\phi}{\partial y} j + \frac{\partial\phi}{\partial z} k$$

$\nabla\phi$ defines what is called the vector field on A. That is every point in A a vector $\nabla\phi$ is defined, and small variations in position correspond to small variations in the corresponding vector.

Gradient (continued..)

We see an example of a vector field on a weather chart. The arrows on it show the direction of the wind and the size of the arrow (or sometimes with a number next to the arrow) the magnitude of the windspeed.

If we think of this idea extended to every point of a three dimensional region within $\mathbb{R}^3$, we have the concept of a 3 dimensional vector field.

If u is a unit vector, then $\nabla \phi . u$, the component of $\nabla \phi$ in the direction of u - is called the directional derivative of ϕ (in the direction of u).

The directional derivative could tell us things like, how density is increasing in a particular direction.

Divergence

Suppose V is a differentiable vector-valued function defined on a region A of $\mathbb{R}^3$, so that:

$$V(x,y,z) = \phi_1 i + \phi_2 j + \phi_3 k$$

Where ϕ_1, ϕ_2 and ϕ_3 are differentiable real-valued functions on A.
This means that $\phi_i(x,y,z)$, for $i = 1,2,3$, is a real number for all (x,y,z) in A.
The divergence of V, denoted by $\operatorname{div} V$ or $\nabla . V$ is:

$$\nabla . V = \left(\frac{\partial}{\partial x} i + \frac{\partial}{\partial y} j + \frac{\partial}{\partial z} k\right) \cdot (\phi_1 i + \phi_2 j + \phi_3 k)$$

$$= \frac{\partial \phi_1}{\partial x} + \frac{\partial \phi_2}{\partial y} + \frac{\partial \phi_3}{\partial z}$$

This behaves like a scalar product, which is why dot notation is used.

Curl

∇ is reffered to as the **del operator**

Suppose V is a differentiable vector function as the last shown for Divergence.
Then the curl of V, denoted by

curl V or $\nabla \times V$ is

$$\nabla \times V = \left(\frac{\partial}{\partial x} i + \frac{\partial}{\partial y} j + \frac{\partial}{\partial z} k\right) \times (\phi_1 i + \phi_2 j + \phi_3 k)$$

$$= \left(\frac{\partial \phi_3}{\partial y} - \frac{\partial \phi_2}{\partial z}\right) i + \left(\frac{\partial \phi_1}{\partial z} - \frac{\partial \phi_3}{\partial x}\right) j$$

$$+ \left(\frac{\partial \phi_2}{\partial x} - \frac{\partial \phi_1}{\partial y}\right)$$

This is sometimes referred to as the rotation of V, and written as 'rot V'.

Curl Example Suppose $V(x,y,z) = V_1 i + V_2 j + V_3 k$ is a differentiable vector field. Then the curl or rotation of V written $\nabla \times V$, Curl V, or rot V is defined as follows:

$$\nabla \times V = \left(\frac{\partial}{\partial x} i + \frac{\partial}{\partial y} j + \frac{\partial}{\partial z} k\right) \times (V_1 i + V_2 j + V_3 k)$$

$$= \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ V_1 & V_2 & V_3 \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ V_2 & V_3 \end{vmatrix} i - \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial z} \\ V_1 & V_3 \end{vmatrix} j + \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \\ V_1 & V_2 \end{vmatrix} k$$

$$= \left(\frac{\partial V_3}{\partial y} - \frac{\partial V_2}{\partial z}\right) i + \left(\frac{\partial V_1}{\partial z} - \frac{\partial V_3}{\partial x}\right) j + \left(\frac{\partial V_2}{\partial x} - \frac{\partial V_1}{\partial y}\right) k$$

Note the expansion of the determinant in operators $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}$ must precede V_1, V_2, V_3.

Summary

① The point with polar coordinates (r, θ) in $\mathbb{R}^2$ has position vector
$$(r\cos\theta)\mathbf{i} + (r\sin\theta)\mathbf{j}$$

② The point with spherical polar coordinates (r, θ, ϕ) in $\mathbb{R}^3$ has position vector
$$(r\cos\theta\sin\phi)\mathbf{i} + (r\sin\theta\sin\phi)\mathbf{j} + (r\cos\phi)\mathbf{k}$$

③ The point with cylindrical polar coordinates (ρ, θ, z) in $\mathbb{R}^3$, has position vector:
$$(\rho\cos\theta)\mathbf{i} + (\rho\sin\theta)\mathbf{j} + z\mathbf{k}$$

④ The tangent plane to a surface whose equation is $z = f(x, y)$ at a point A (with position vector $\mathbf{a}$) on the surface is
$$f_x(a,b)(x-a) + f_y(a,b)(y-b) - (z-c) = 0$$
where $c = f(a,b)$ and $\mathbf{a} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$

⑤ The tangent plane to a surface whose equation is $F(x,y,z) = 0$ at a point A (with position vector $\mathbf{a}$) on the surface is
$$F_x(a,b,c)(x-a) + F_y(a,b,c)(y-b) + F_z(a,b,c)(z-c) = 0$$
where $F(a,b,c) = 0$, and $\mathbf{a} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$.

⑥ The normal vector to a surface whose equation is $F(x,y,z) = 0$ at a point A (with position vector $\mathbf{a}$) on the surface is $\nabla F(\mathbf{a})$ where:
$$\nabla F(\mathbf{a}) = F_x(a,b,c)\mathbf{i} + F_y(a,b,c)\mathbf{j} + F_z(a,b,c)\mathbf{k}$$
The vector equation of the tangent plane to the surface at A is $\nabla F(\mathbf{a}) \cdot (\mathbf{r} - \mathbf{a}) = 0$.

⑦ Suppose $\phi : A \to \mathbb{R}$ is a differentiable scalar function, and $\mathbf{V}$ is a differentiable vector function, defined in A of $\mathbb{R}^3$ $\mathbf{V}(x,y,z) = \phi_1\mathbf{i} + \phi_2\mathbf{j} + \phi_3\mathbf{k}$, then grad, div, curl are:
$$\text{grad } \phi = \nabla\phi = \frac{\partial\phi}{\partial x}\mathbf{i} + \frac{\partial\phi}{\partial y}\mathbf{j} + \frac{\partial\phi}{\partial z}\mathbf{k}$$
$$\text{div } \mathbf{V} = \nabla\cdot\mathbf{V} = \frac{\partial\phi_1}{\partial x} + \frac{\partial\phi_2}{\partial y} + \frac{\partial\phi_3}{\partial z}$$
$$\text{curl } \mathbf{V} = \nabla\times\mathbf{V} = \left(\frac{\partial\phi_3}{\partial y} - \frac{\partial\phi_2}{\partial z}\right)\mathbf{i} + \left(\frac{\partial\phi_1}{\partial z} - \frac{\partial\phi_3}{\partial x}\right)\mathbf{j} + \left(\frac{\partial\phi_2}{\partial x} - \frac{\partial\phi_1}{\partial y}\right)\mathbf{k}$$

Curl and Gradient Examples

Vectors

Curl Example

Another Curl Example

Suppose $A = x^2z^2 i - 2y^2z^2 j + xy^2zk$. Find $\nabla \times A$ (on curl A) at the point $P(1,-1,1)$.

$$\nabla \times A \left(\frac{\partial}{\partial x} i + \frac{\partial}{\partial y} j + \frac{\partial}{\partial z} k\right) \times \left(x^2z^2 i - 2y^2z^2 j + xy^2zk\right)$$

$$= \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ x^2z^2 & -2y^2z^2 & xy^2z \end{vmatrix} = \left[\frac{\partial}{\partial y}(xy^2z) - \frac{\partial}{\partial z}(-2y^2z^2)\right] i - \left[\frac{\partial}{\partial x}(xy^2z) - \frac{\partial}{\partial z}(x^2z^2)\right] j +$$

$$\left[\frac{\partial}{\partial x}(-2y^2z^2) - \frac{\partial}{\partial y}(x^2z^2)\right] k$$

$$= (2xyz + 4y^2z) i - (y^2z - 2x^2z) j + 0k$$

at the point $P(1,-1,1)$, $\nabla \times A = 2i + j$

ϕ is a scalar field

Another Gradient Example

Let $\phi(x,y,z)$ be a scalar function defined and differentiable at each point (x,y,z) in a certain region of space. [That is, ϕ defines a differentiable scalar field]. Then the gradient of ϕ, written $\nabla\phi$ or grad ϕ is defined as follows.

$$\nabla\phi = \left(\frac{\partial}{\partial x} i + \frac{\partial}{\partial y} j + \frac{\partial}{\partial z} k\right)\phi = \frac{\partial\phi}{\partial x} i + \frac{\partial\phi}{\partial y} j + \frac{\partial\phi}{\partial z} k$$

Note that $\nabla\phi$ defines a vector field.

Gradient Example

Gradient Example

Suppose $\phi(x,y,z) = 3xy^3 - y^2z^2$. Find $\nabla\phi$ (or grad ϕ) at the point $P(1,1,2)$.

$\nabla\phi$ is a vector field

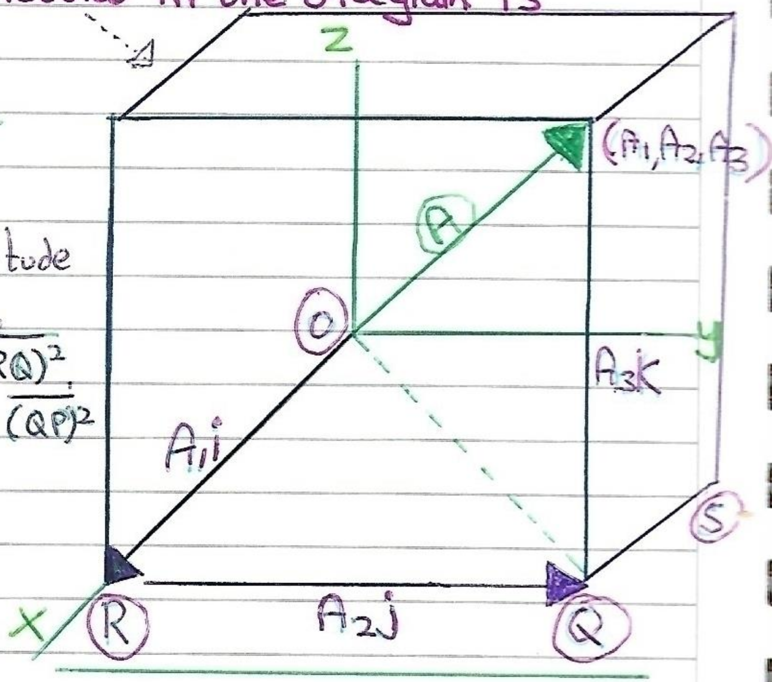
$$\nabla\phi = \left(\frac{\partial}{\partial x} i + \frac{\partial}{\partial y} j + \frac{\partial}{\partial z} k\right)\left(3xy^3 - y^2z^2\right)$$

$$= 3y^3 i + (9xy^2 - 2yz^2) j - 2y^2zk$$

Therefore $\nabla\phi(1,1,2) = 3(\underset{y}{1})^3 i + [9(\underset{x}{1})(\underset{y}{1})^2 - 2(\underset{y}{1})(\underset{z}{2})^2] j - 2(\underset{y}{1})^2(\underset{z}{2})k = 3i + j - 4k$

Picturing Vectors

① Showing that the magnitude of the vector $A = A_1 i + A_2 j + A_3 k$ pictured in the diagram is $|A| = \sqrt{A_1^2 + A_2^2 + A_3^2}$



Solution by Pytagorean theorem

$(\overline{OP})^2 = (\overline{OQ})^2 + (\overline{QP})^2$

where $\overline{OP}$ denotes magnitude of vector OP, and so on.

Similarly, $(\overline{OQ})^2 = (\overline{OR})^2 + (\overline{RQ})^2$.

Then $(\overline{OP})^2 = (\overline{OR})^2 + (\overline{RQ})^2 + (\overline{QP})^2$

or $A^2 = A_1^2 + A_2^2 + A_3^2$

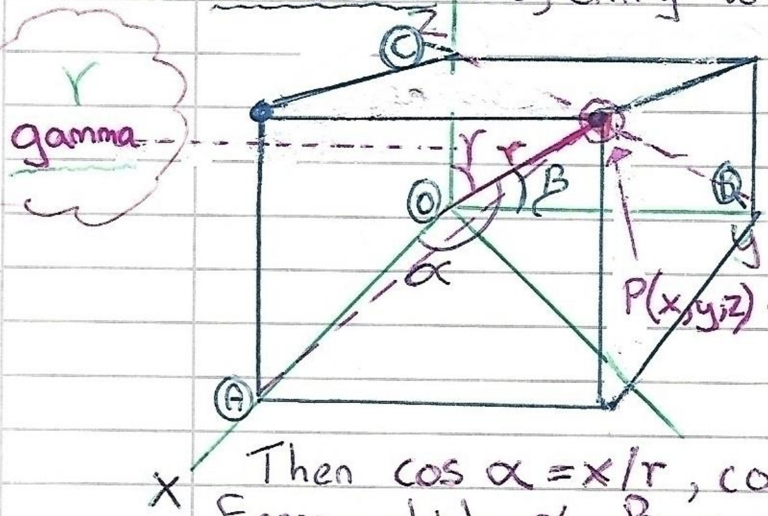
(i.e. $\sqrt{A_1^2 + A_2^2 + A_3^2} = A$)

② Determine the angles α, β, and γ that the vector $r = xi + yj + zk$ makes withe the positive directions of the coordinate axes and show that

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$$

Solution: Refering to the diagram below, triangle OAP is a right triangle with right angle at A. Then $\cos\alpha = x/|r|$. Similarly from right triangles OBP and OCP, $\cos\beta = y/|r|$ and $\cos\gamma = z/|r|$ respectively.

Also $|r| = r = \sqrt{x^2 + y^2 + z^2}$.



Then $\cos\alpha = x/r$, $\cos\beta = y/r$ and $\cos\gamma = z/r$. From which α, β and γ can be obtained. So from these it follows

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = \frac{x^2 + y^2 + z^2}{r^2} = 1$$

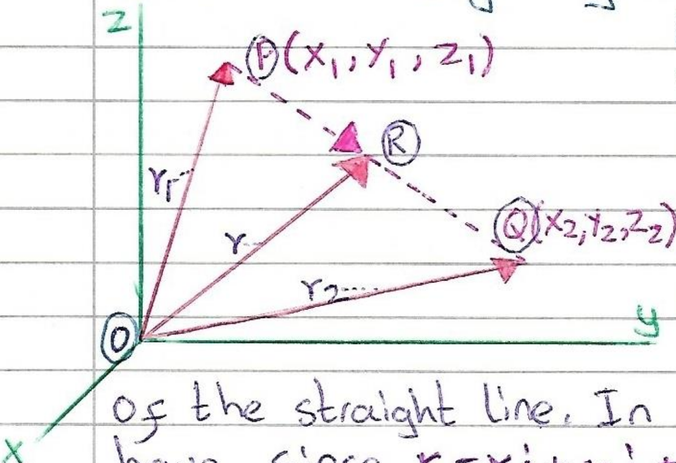
The numbers $\cos\alpha$, $\cos\beta$ and $\cos\gamma$ are called the direction cosines of the vector OP.

Picturing Vectors

Vectors

Finding the set of equations for the straight lines passing through the points $P(x_1, y_1, z_1)$ and $Q(x_2, y_2, z_2)$.

Solution: Let r_1 and r_2 be the position vectors of P and Q respectively, and r the position vector of any point R on that line joining P and Q.



As shown in the diagram:

$r_1 + PR = r$ or $PR = r - r_1$

$r_1 + PQ = r_2$ or $PQ = r_2 - r_1$

But $PR = tPQ$ where t is a scalar. Then $r - r_1 = t(r_2 - r_1)$ is the required vector equation of the straight line. In rectangular coordinates, we have, since $r = xi + yj + zk$,

$$(xi + yj + zk) - (x_1 i + y_1 j + z_1 k) = t[(x_2 i + y_2 j + z_2 k) - (x_1 i + y_1 j + z_1 k)]$$

(or)

$$(x - x_1)i + (y - y_1)j + (z - z_1)k = t[(x_2 - x_1)i + (y_2 - y_1)j + (z_2 - z_1)k]$$

Since i, j and k are non-coplaner vectors. Meaning an equation in the form $(x_1 - x_2)a + (y_1 - y_2)b + (z_1 + z_2)c = 0$ can be written $[x_1 a + y_1 b + z_1 c = x_2 a + y_2 b + z_2 c]$.

So $xa + yb + zc = 0$ which implies $x = y = z = 0$. As a side note. If $x \neq 0$ then $xa + yb + zc = 0$ would imply $xa = -yb - zc$ or $a = -(y/x)b - (z/x)c$. But $-(y/x)b - (z/x)c$ is a vector lying on the plane b, c. That is a lies on the plane b, c which is clearly in contradiction to an hypothesis that a, b and c would be non-coplaner.

Back to the problem's i, j and k vectors we have

$$x - x_1 = t(x_2 - x_1),\quad y - y_1 = t(y_2 - y_1),\quad z - z_1 = t(z_2 - z_1)$$

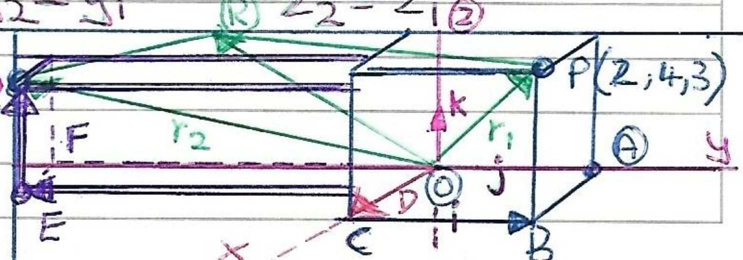
as the parametric equations of the line, t being the parameter. Eliminating t, the equations become:

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1}$$

- Consider vectors r_1, r_2 for P and Q in terms unit vectors i, j, k {Q(1,-5,2)}
- Determine graphically/analytacally results.

$r_1 = OP = OC + CB + BP = 2i + 4j + 3k$

$r_2 = OQ = OD + DE + EQ = i - 5j + 2k$



Fixed Points And Eigenvectors

Vectors

If A is a (real) eigenvector of the matrix M corresponding to an eigen~~vector~~value λ then $AM = \lambda A$ and so $[A]$ is a fixed point of the corresponding projective transformation and conversely. Since a 3×3 real matrix always has at least one real eigenvalue every projective transformation has at least one fixed point.

- Example

Let us find the fixed points of the projective transformation corresponding to the matrix
$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 4 \end{pmatrix}$ The eigenvalues are given by the equation $\begin{vmatrix} \lambda-2 & -1 & 0 \\ 0 & \lambda-1 & 1 \\ 0 & -2 & \lambda-4 \end{vmatrix} = (\lambda-2)^2(\lambda-3) = 0$

So there are two eigenvalues 2 (twice) and 3. The corresponding eigenvectors are (up to a non-zero multiple) $(0,2,1)$ and $(0,1,1)$. Hence the fixed points of the projective transformation are $[0,2,1]$ and $[0,1,1]$. Notice that the line $x=0$ is invariant in the sense that any point with x-coordinate zero is transformed into another point with x-coordinate zero.

'Worded' Example

- Perspective Example

On a straight road approaching traffic lights there are 'slow down' signs, 400m and 200m from the traffic lights. A town planner makes a perspective drawing in which the 'slow down' signs are 3cm and 1cm from the traffic lights. Where should the warning sign be placed on the drawing?

Let the slow down signs be S and S' and the warning sign be at W. If T is the traffic lights then the cross ratio is

$$\frac{(S-W)(S'-T)}{(S-T)(S'-W)} = \frac{300\times200}{400\times100} = \frac{3}{2}$$

In the drawing the warning sign is placed x cm from the traffic lights then this is equal to

$\dfrac{(3-x)(1)}{3(1-x)}$ So $9(1-x) = 2(3-x)$ which when solved gives $x = 3/7$ cm

Certain vector identities exist for vector operators. A couple of these will now be discussed in a little detail. A vector field a for which $\operatorname{div} a = \nabla \cdot a = 0$ is said to be solenoidal, and one for which curl $a = 0$ to be irrotational. If the vector a is itself derived as the gradient of a scalar then it is necessarily irrotational. To see this we simply write out the expression for curl a

$$\nabla \wedge a = \left(\frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z}, \ldots\ldots \right) = \left(\frac{\partial}{\partial y}\left(\frac{\partial \phi}{\partial z}\right) - \frac{\partial}{\partial z}\left(\frac{\partial \phi}{\partial y}\right), \ldots \right)$$

ie curl $a = (0,0,0)$ if $a = \operatorname{grad} \phi$.

This result might also be expected since curl a can be written a $\nabla \wedge \nabla \phi$ and looks like the vector product of a vector with itself, but when dealing with vector operators, such results can't be always assumed.

If a is derived as curl b for some vector b, then a is solenoidal ie $\nabla \cdot a = 0$.

The analog of the triple vector product is worth a look in the context of the vector given by curl a. So when both a and b are replaced by ∇ in the equation $c \wedge (b \wedge a)$ then the relationship is $\nabla \wedge (\nabla \wedge a) = \nabla(\nabla \cdot a) - \nabla^2 a$

or that curl(curl a) - grad (div a) -del squared a

Vector Field - Summary. Suppose to each point (x,y,z) of a region D in space there corresponds a vector $V(x,y,z)$. Then V is called a vector function of position, and we say that a vector field V has been defined on D.

① Suppose the velocity at any point within a moving fluid is known at a certain time. Then a vector field is defined.

② The function $V(x,y,z) = xy^2 i - 2yz^3 j + x^2 z k$ defines a vector field. Consider the point $P(2,3,1)$. Then $V(P) = 18i - 6j + 4k$.

A vector field V which is independent of time is called a stationary or steady-state vector field.

Gaussian Elimination / Row Reduction

Gaussian Elimination. A method for successive elimination of unknowns when solving a system of linear algebraic equations (Σ_0).

$$\sum_{j=1}^{n} a_{ij} x_j - a_{i0} = 0, \quad i = 1, \ldots, m \quad (\Sigma_0)$$

where a_{ij} are elements of some field F. Assuming that $a_{11} \neq 0$ (otherwise, renumber the equations), the key algorithmic step can be described as follows:

- Multiply the first equation by (a_{21}/a_{11}) and subtract it (term by term) from the second equation. Next, multiply the first equation by (a_{31}/a_{11}) and subtract it from the third equation, etc., until the first equation is multiplied by (a_{m1}/a_{11}) and subtracted from the last $(i=m)$ equation of the system (Σ_0).

Designate the resulting system of equations with the first equation deleted by (Σ_1), a carry out the same set of operations on (Σ_1) obtaining (Σ_2), etc. Assuming that the rank of the coefficient matrix of (Σ_0) [which is also called the rank of the system of equations (Σ_0)], $r = \text{rank}(\Sigma_0)$, is smaller than m, $r < m$, we obtain after the r^{th} step, a system (Σ_r) in which all the coefficients of the unknowns vanish. The system (Σ_0) [or (Σ_r)] is called compatible, if in (Σ_r) all the absolute terms vanish as well (i.e., when the rank of the coefficient matrix equals the rank of the augmented matrix); otherwise it is incompatible and has no solution.

To obtain a solution of a compatible system, we choose some solution $(x_r^{(0)}, \ldots, x_n^{(0)})$ of the system (Σ_{r-1}) and proceed by the back substitution to (Σ_{r-2}), (Σ_{r-3}), etc., until (Σ_0) is reached. In general, we can choose a solution for (Σ_{r-1}) by assigning arbitary values to the $n-r$ variables $x_{r+1}, \ldots, x_n$.

Gaussian Elimination - Continued

Say $X_j = C_{j-r}$ $(j = r+1, \dots n)$ so that

$$X_r^{(0)} = (a_{r0}^{(r-1)} - \sum_{j=r+1}^{n} a_{rj}^{(r-1)} C_{j-r}) / a_{rr}^{(r-1)}$$

and $x_j^{(0)} = C_{j-r}$ for $j = r+1, \dots n$. The (general) Solution will then depend on $r-n$ arbitary parameters $C_j \in F$, $j = 1, \dots, n-r$.

Once we have a solution of (Σ_{r-1}), the back substitution proceeds by assigning the values $X_r^{(0)}, \dots, X_n^{(0)}$ to the unknowns $X_r, \dots, X_n$ in the first equation of (Σ_{r-2}), obtaining $X_{r-1} = X_{r-1}^{(0)}$, and so a solution $(X_{r-1}^{(0)}, X_r^{(0)}, \dots X_n^{(0)})$ of (Σ_{r-2}). These values for $X_{r-1}, \dots, X_n$ are then substituted into the first equation of (Σ_{r-3}), obtaining $X_{r-2} = X_{r-2}^{(0)}$, etc, until a solution $(X_1^{(0)}, X_2^{(0)}, \dots, X_n^{(0)})$ of (Σ_0) is obtained. The general solution results when C_j $(j = 1, \dots n-r)$ are regarded as free parameters.

This method can be generalized in various ways. It can also be formulated in terms of a general $m \times n$ (or $m \times n+1$) Matrix A over F [representing the coefficient (or augmented) matrix of a system (Σ_0)], in which case it is normally refered to as the row reduction of A. The algorithm can then be convienently expressed through the so-called elementary row operations, which in turn can be represented by the elementary matrices of three basic types [$(I + ae_{ij})$, $i \neq j$, replacing the ith row X_i by $X_i + aX_j$, $I + e_{ij} + e_{ji} - e_{ii} - e_{jj}$, interchanging rows i and j, and $I + (c-1)e_{ii}$, $c \neq 0$, multiplying the ith row by c] acting from the left on A. Clearly, the action of the elematary matrices and of their inverses on the augmented matrix A of a system (Σ_0) produces an equivalent system of linear algebraic equations, and the process of Gauss elimination can be thus represented by a product of corresponding elementary matrices.

Row Reduction of matrices

Consider a system of m equations in n unknowns.

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

This may be written as the single equation $Ax=b$, where A is the $m \times n$ matrix (a_{rs}) and x and b are the column vectors $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ $\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$ respectively. If we can now 'divide' by A we should be able to solve for x. For matrices this corresponds to find an inverse, A^{-1}. For a true inverse to exist, A has to be square, as we shall see; nevertheless it is always possible to multiply on the left by an appropriate matrix B to put the equations into a simple form. This corresponds to the usual method of solution of simutaneous linear equations by elimination of variables. Let us illustrate this -

Consider the equations

$$\left.\begin{aligned} x + y + z + t &= 4 \\ 2x + 2y - 2z - t &= 3 \\ x \qquad + z + 2t &= 2 \end{aligned}\right\} \quad \begin{aligned} &1 \\ &2 \\ &3 \end{aligned}$$

We solve these by eliminating on unknown at a time (always retaining the equation we used for the elimination). Eliminate x using 1, subtracting it from 3 and from 2.

$$\begin{aligned} x + y + z + t &= 4 \quad [1] \\ -4z - 3t &= -5 \quad [4] \\ -y \qquad + t &= -2 \quad [5] \end{aligned}$$

Now change the order -

$$\begin{aligned} x + y + z + t &= 4 \quad [1] \\ -y \qquad + t &= -2 \quad [5] \\ -4z - 3t &= -5 \quad [4] \end{aligned}$$

Divide through by the leading coefficients -

$$\begin{aligned} x + y + z + t &= 4 \quad [1] \\ y \qquad - t &= 2 \quad [6] \\ z + (3/4)t &= 5/4 \quad [7] \end{aligned}$$

There are other routes to this system. Also it may be further simplified by eliminating z from (continued) equation (1), using equation [7]. But generally (for numerical and computational reasons) one reduces such a system no further than its echelon form, as with equations (1), [6] and [7] where no further elimination of leftmost unknowns is possible.

In this case there can be no unique solution to the system, since there are too many unknowns: t can be given any value at all, and it will still be possible to solve for x, y and z. All the same, we can give a general form of the solution, and find the important point that it is one-dimensional. This is a precise form of the equations above, with one degree of freedom.

As note the form in which we have just written the equations enables us, once we know the value of t, to obtain the values of x, y and z by back-substitution. The value of t is, however arbitrary. Writing λ for this value, the general solution is,

$$t = \lambda$$
$$z = 5/4 - (3/4)\lambda$$
$$y = 2 + \lambda$$
$$x = 4 - (2+\lambda) - (5/4 - (3/4)\lambda) - \lambda = 3/4 - (5/4)\lambda$$

The four stages above, from the original eqs. to the echelon form, is contained in the matrices-

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 2 & 2 & -2 & -1 & 3 \\ 1 & 0 & 1 & 2 & 2 \end{array}\right), \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 0 & 0 & -4 & -3 & -5 \\ 0 & -1 & 0 & 1 & -2 \end{array}\right)$$

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 0 & -1 & 0 & 1 & -2 \\ 0 & 0 & -4 & -3 & -5 \end{array}\right); \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 0 & 1 & 0 & -1 & 2 \\ 0 & 0 & 1 & 3/4 & 5/4 \end{array}\right)$$

The final column in each matrix is partitioned from the rest to remind us that it gives the right hand side of the equations; a matrix thus partitioned is called an augmented matrix.

(continued) The method of passing from one of these matrices to the next is (a combination of) so-called elementary row operations:

Here 'k' is a constant → ① Replace row r by row r − k × row s (some r ≠ s, constant)

② Multiply row r by k (some non-zero constant k)

③ permute the rows (that is, write them in a different order)

An elementary row operation of each of these types was used in the above calculation. Note ③ is required only when the variable we are trying to eliminate has coefficient 0. In order to adopt a systematic strategy we have combined rows where possible. So in passing from the first to second matrix ⑧, we have also replaced row 2 by row 2 − 2 × row 1, and row 3 with row 3 − 1 × row 1. In the third matrix we replaced row 2 by (−1) × row 2 and row 3 by $-\frac{1}{4}$ × row 3.

Now, row operations of types ①, ② and ③ may all be effected by multiplying on the left by appropriate 'elementary' matrices. To obtain the desired elementary matrix, one simply performs the given row operation on the identity matrix of the appropriate size, in this case,

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

The five elementary matrices in this instance are

$$\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{4} \end{pmatrix}$$

For example we find that.

$$\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 2 & 2 & -2 & -1 & 3 \\ 1 & 0 & 1 & 2 & 2 \end{array}\right) = \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 0 & 0 & -4 & -3 & -5 \\ 1 & 0 & 1 & 2 & 2 \end{array}\right)$$

and

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 0 & 0 & -4 & -3 & -5 \\ 1 & 0 & 1 & 2 & 2 \end{array}\right) = \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 0 & 0 & -4 & -3 & -5 \\ 0 & -1 & 0 & 1 & -2 \end{array}\right)$$

Rank and Nullity - Systems of equations

(continued) These matrices form the product below -

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&-\frac14\end{pmatrix}\begin{pmatrix}1&0&0\\0&-1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\-1&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\-2&1&0\\0&0&1\end{pmatrix}=\begin{pmatrix}1&0&0\\1&0&-1\\\frac12&-\frac14&0\end{pmatrix}$$

We find that these row operations could all be done in one go by multiplication on the left.

$$\begin{pmatrix}1&0&0\\1&0&-1\\\frac12&-\frac14&0\end{pmatrix}\left(\begin{array}{cccc|c}1&1&1&1&4\\2&2&-2&-1&3\\1&0&1&2&2\end{array}\right)=\left(\begin{array}{cccc|c}1&1&1&1&4\\0&1&0&-1&2\\0&0&1&\frac34&\frac54\end{array}\right)$$

Thus the matrix - $\begin{pmatrix}1&0&0\\1&0&-1\\\frac12&-\frac14&0\end{pmatrix}$ acts like a 'left inverse' for the original matrix

New Topic

Rank and Nullity

Before giving a more precise description of the algorithm we have sketched for solving systems of linear equations, we take a look at linear transforms. Since $m\times n$ matrices can be regarded as linear transformations from an n-dimensional space to an m-dimensional space, it is natural to view the solution of systems of equations in this light. Normally we'll work with vector spaces F^n and F^m of column vectors over a field F, but some results can be formulated in greater generality.

If $f: V\to W$ is a linear transformation, the null space (or kernel) N and range R of 'f' are given by $N=\{x\in V: f(x)=0\}$, $R=\{f(x): x\in V\}$. It is easily checked that N is a subspace of V, and R is a subspace of W. Their dimensions are called the nullity and rank of f, respectively.

Lemma rank f + nullity f = dim V.

Intuitively this lemma says that, the more vectors are destroyed by f, the smaller it's range is.

Example Let 'f' have the matrix $\begin{pmatrix}2&1&2&1\\1&-1&3&0\\3&3&1&2\end{pmatrix}$ Then $\begin{pmatrix}x\\y\\z\\t\end{pmatrix}\in N \Leftrightarrow \left\{\begin{array}{l}2x+y+2z+t=0\\x-y+3z=0\\3x+3y+z+2t=0\end{array}\right\}$

Using row reductions this is equiv. to $\left.\begin{array}{l}2x+y+2z+t=0\\-(3/2)y+2z-(1/2)t=0\\(3/2)y-2z+(1/2)t=0\end{array}\right\}$ → and so to → $\left\{\begin{array}{l}x+(\frac12)y+z+(\frac12)t=0\\y-(4/3)z+(\frac13)t=0\end{array}\right\}$

Rank and Nullity Continued

(continued) In the case at the bottom of the previous page, there are two degrees of freedom (z and t may be arbitrarily chosen), so dim N, the nullity, is equal to 2. More precisely, N has as the basis the vectors $\begin{pmatrix} -5/3 \\ 4/3 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -1/3 \\ -1/3 \\ 0 \\ 1 \end{pmatrix}$ The typical member of R is $\begin{pmatrix} 2 & 1 & 2 & 1 \\ 1 & -1 & 3 & 0 \\ 3 & 3 & 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} 2x+y+2z+t \\ x-y+3z \\ 3x+3y+z+2t \end{pmatrix}$

$$= \begin{pmatrix} (1/2)x-(1/2)y+(3/2)z \\ x-y+3z \\ 0 \end{pmatrix} + \begin{pmatrix} (3/2)x+(3/2)y+(1/2)z+t \\ 0 \\ 3x+3y+z+2t \end{pmatrix}$$

$$= (x-y+3z)\begin{pmatrix} 1/2 \\ 1 \\ 0 \end{pmatrix} + (3x+3y+z+2t)\begin{pmatrix} 1/2 \\ 0 \\ 1 \end{pmatrix}$$

And so $\left\{ \begin{pmatrix} 1/2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/2 \\ 1 \\ 0 \end{pmatrix} \right\}$ is a basis for R, and is also 2 dimensional. Note: $2+2=\dim V$, verifying the Lemma.

[Theorem] The set of solutions of a system of linear equations over a field F in 'n' unknowns is either empty or is a translated subspace of F^n.

coset/ translated space: If U is a subspace of the vector space V and $x \in V$, then $U+x$, the set of all vectors of the form $u+x$ with $u \in U$, is called a translated space of V (or coset of U). Proof - As indicated above, the system may be written in the form $Ax=b$, where $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ is the vector of unknowns. If the set of solutions is non-empty, let v be a solution. Thus $Av=b$. Let N be the null-space of A. Then $Ax=b \Leftrightarrow Ax=Av \Leftrightarrow A(x-v)=0 \Leftrightarrow x-v \in N \Leftrightarrow x \in N+v$ so that the set of solutions is the translated subspace $N+v$.

[corollary] [Corollary] A system of linear equations over an infinite field has 0, 1 or infinitely many solutions.

[Banded Matrix] A matrix A is said to be 'banded' if all it's nonzero elements are confined within a band formed by diagonals parallel to the main diagonal. Thus $A_{ij}=0$ when $|i-j|>\beta$, and $A_{k,k-\beta} \neq 0$ or $A_{k,k+\beta} \neq 0$ for at least one value of k, where β is the half-bandwidth and $2\beta+1$ is the bandwidth. The band of the matrix is the set of elements for which $|i-j| \leq \beta$. In other words, for a certain (continued) row i, all elements having column indices in the range $i-B$ to $i+B$, a total of $2B+1$ elements per row, belong to the band. This can be much smaller than the order of the matrix.

Another Gauss Example - Gauss Elimination by Columns

We will assume that A, the matrix of the linear system is already prepared for elimination, in the sense that the diagonal elements can be used directly as pivots in the same order in which they are given. In general, pivots are selected among the nonzero elements of A and brought to the main diagonal by means of row and column permutations, the selection being done in such a way that sparsity is preserved and numerical stability is ensured. The question of pivot selection or ordering is left in the following discussion.

Gauss elimination by columns consists of n steps. The purpose of the k^{th} step is to eliminate all the nonzero elements of the matrix which lie on the column k where this is below the diagonal. At the first step, the nonzeros of Column 1 of $A \equiv A^{(1)}$ are eliminated by subtracting convenient multiples of row 1, element by element, from each of the remaining rows with a nonzero in column 1. The element A_{11}, belonging to the row that is going to be subtracted from other rows (row 1 in this case) and to the column that will be eliminated (column 1 in this case),

[Pivot] is called the pivot and assumed to be nonzero. Previous to the elimination, row 1 is normalized by dividing all its nonzero elements by the pivot. A matrix $A^{(2)}$ is obtained, with $A^{(2)}_{i1}=0$ for $i>1$ and $A^{(2)}_{11}=1$.

At the second step, $A^{(2)}_{22}$ is selected to be the pivot. Again we assume $A^{(2)}_{22} \neq 0$. Row 2 is normalized, and all nonzeros of the second column below the diagonal are eliminated by subtraction of convenient multiples of the normalized second row from the

Gauss Elimination by Columns

(continued) corresponding rows. Note that, since $A_{21}^{(2)} = 0$, the elements of column 1 will not be affected. A matrix $A^{(3)}$ is obtained with $A_{i1}^{(3)} = 0$ for $i > 1$, $A_{i2}^{(3)} = 0$ for $i > 2$, and $A_{11}^{(3)} = A_{22}^{(3)} = 1$. In other words, $A^{(3)}$ is upper triangular unit diagonal in its first two columns.

[upper triangle]

At the beginning of the k^{th} step we have a matrix $A^{(k)}$ with zeros on its first $k-1$ columns below the diagonal and ones on the $k-1$ initial positions of the diagonal. The following example shows $A^{(k)}$ for the case $n=6$, $k=3$.

[pivot]

$$\begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 1 & X & X & X & X & X \\ 2 & & 1 & X & X & X & X \\ 3 & & & X & X & X & X \\ 4 & & & X & X & X & X \\ 5 & & & X & X & X & X \\ 6 & & & X & X & X & X \end{array}$$

At the k^{th} step, $A_{kk}^{(k)}$ is selected to be the pivot. Row k is normalized and the elements on column k below the diagonal of $A^{(k)}$ are eliminated by subtraction of convenient multiples of the normalized row k from all those rows which have a nonzero in column k below the diagonal. The matrix $A^{(k+1)}$ is obtained with zeros in its first k columns below the diagonal and ones on the first k positions of the diagonal.

This process is continued until, at the end of step n the matrix $A^{(n+1)}$ is obtained, which has only zeros below the diagonal and ones on the diagonal, and is thus upper triangular unit diagonal.

[Upper triangle diagonal]

This process is continued until, at the end of step n the matrix $A^{(n+1)}$ is obtained, which has only zeros below the diagonal and ones on the diagonal, and is thus upper triangular unit diagonal.

A convenient notation for the algorithm is obtained by using the elementary matrices D_k and L_k^c. The k^{th} step of Gauss elimination by columns is equivalent to pre-multiplication of $A^{(k)}$ by $(L_k^c)^{-1} D_k^{-1}$:

$$A^{(k+1)} = (L_k^c)^{-1} D_k^{-1} A^{(k)}$$

where $(D_k)_{kk} = A_{kk}^{(k)}$, $(L_k^c)_{ik} = A_{ik}^{(k)}$ for $i > k$

(continued) ... and $A^{(1)} = A$. Note that by property of the off-diagonal elements of $(L_K^c)^{-1}$ are those of L_K^c with their signs reversed, ie $-A_{ik}^{(k)}$.

Therefore, if for the sake of completeness we add a trivial pre-multiplication of $D_n^{-1} A^{(n)}$ by $(L_n^c) \equiv I$ (the identity matrix) to obtain $A^{(n+1)}$ we have

$$(L_n^c)^{-1} D_n^{-1} \ldots (L_2^c)^{-1} D_2^{-1} (L_1^c) D_1^{-1} A = U$$

where $U \equiv A^{(n+1)}$ is upper triangular unit diagonal with

$$U_{kj} = (D_k)_{kk}^{-1} A_{kj}^{(k)} \equiv A_{kj}^{(k+1)} \quad \text{for } j > k$$

[Lower triangular] From this expression the factorized form of A is $A = LU$, where $L = D_1 L_1^c D_2 L_2^c \ldots D_n L_n^c$ is a lower triangular matrix, with nonzero diagonal elements. Since the products in the last equation are calculated by superposition, the elements of L can easily be obtained from $[(D_k)_{kk} = A_{kk}^{(k)}]$ as $L_{ik} = A_{ik}^{(k)}$ for $i \geq k$.

Using $U_{kj} = (D_k)_{kk}^{-1} A_{kj}^{(k)} \equiv A_{kj}^{(k+1)}$ and $L_{ik} = A_{ik}^{(k)}$ the operations performed on $A^{(k)}$ to obtain $A^{(k+1)}$ can be indicated: $A_{ij}^{(k+1)} = A_{ij}^{(k)} - L_{ik} U_{kj} \quad i, j > k$

So that the complete expression of the elements of $A^{(k)}$ can be written:

$$A_{ij}^{(k)} = A_{ij} - \sum_{m=1}^{k} L_{im} U_{mj} \quad i, j > k$$

The elimination ~~form~~ or of the inverse is obtained from $A = LU$: $A^{-1} = U^{-1} L^{-1}$

In practice, the calculation of U^{-1} and L^{-1} is never explicitly made. The results of the elimination are usually overwritten on the storage initially occupied by A, in the form of a table which represents two different matrices and is sometimes called 'a table of factors':

$$\begin{matrix} (D_1)_{11}^{-1} & U_{12} & U_{13} & \ldots \\ (L_1^c)_{21} & (D_2)_{22}^{-1} & U_{23} & \\ (L_1^c)_{31} & (L_2^c)_{32} & (D_3)_{33}^{-1} & \\ \vdots & & & \end{matrix}$$

First, let us examine how this table is formed. By $(D_k)_{kk} = A_{kk}^{(k)}$, $(L_k^c)_{ik} = A_{ik}^{(k)}$ for $i > k$, the lower triangle is formed simply by leaving the (continued) elements $A_{ik}^{(k)}$ where they are just before elimination; these elements are precisely those which will be eliminated in step k, The diagonal of the table is formed by leaving, at each step k, the reciprocal of the diagonal element of row k, The reciprocal is obtained at the time that row k is normalized, because it is computationally faster to calculate the reciprocal and to multiply all row elements by it, than to divide by the diagonal element, Besides, it is convenient to have the reciprocals directly available when the table is used for back-substitution. In the upper triangle, the elements of U are left as they are obtained in the course of the elimination, as indicated by

$$U_{kj} = (D_k)_{kk}^{-1} A_{kj}^{(k)} \equiv A_{kj}^{(k+1)} \quad \text{for } j > k.$$

Now let us examine how the table of factors is used. Its most common application is the solution of the system $Ax = b$, by means of $X = A^{-1} b$, which, using $A^{-1} = U^{-1} L^{-1}$, is $X = U^{-1} L^{-1} b$. From $L = D_1 L_1^c D_2 L_2^c \dots D_n L_n^c$, we have $L^{-1} = (L_n^c)^{-1} D_n^{-1} \dots (L_2^c)^{-1} D_2^{-1} (L_1^c)^{-1} D_1^{-1}$. The matrices D_k^{-1} are directly available from the table.

Charlotte Ameil