

Lecture Notes for

# MTH5102: CALCULUS III

**Dr. Will Sutherland**

**These notes are updated from previous versions by Malcolm McCallum and M.J. Thompson, with figures by C.D. Murray, and by P. Saha.**

**School of Mathematical Sciences,  
Queen Mary, University of London  
September 2010**

“It is one of the most unnatural features of science that the abstract language of mathematics should provide such a powerful tool for describing the behaviour of systems both inanimate, as in physics, and living, as in biology. Why the world should conform to mathematical descriptions is a deep question. Whatever the answer, it is astonishing.”

Lewis Wolpert (1992)

Students on the module MTH5102 Calculus III are welcome to download, print and photocopy these notes, in whole or in part, for their personal use. The notes are intended to supplement rather than to replace the lectures.

© Will Sutherland 2009-10, M.A.H. MacCallum 2006-8, 2001, P. Saha 2005, M.J. Thompson 1999.

# Chapter 1

## Introductory material

Last revised: 29 Sep 2010

This chapter gives a quick review of the key parts of the prerequisite courses (Calculus I and II, and Geometry I) which we will actually use in Calculus III, adding some extra material. Those parts which are revision will be without examples.

### 1.1 Trigonometric functions

#### 1.1.1 Values

(See Thomas 1.6)

We can quickly obtain the value of a trigonometric function for any argument in terms of values for  $x \in [0, \frac{1}{2}\pi]$  by remembering a few things. First we have the table

	$0^\circ$	$30^\circ = \frac{\pi}{6}$ radians	$45^\circ = \frac{\pi}{4}$ rad.	$60^\circ = \frac{\pi}{3}$ rad.	$90^\circ = \frac{\pi}{2}$ rad.
cos	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0
sin	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1

To get the sign for other values we can use the mnemonic table

Radians	Degrees	sin	cos	tan	Positive functions
$(0, \frac{1}{2}\pi)$	$(0^\circ, 90^\circ)$	+	+	+	All
$(\frac{1}{2}\pi, \pi)$	$(90^\circ, 180^\circ)$	+	-	-	Sin
$(\pi, \frac{3}{2}\pi)$	$(180^\circ, 270^\circ)$	-	-	+	Tan
$(\frac{3}{2}\pi, 2\pi)$	$(270^\circ, 360^\circ)$	-	+	-	Cos

sometimes called the ‘Add Sugar To Coffee’ rule – or use Thomas’ variant “All Students Take Calculus”. (Note: to be entirely accurate we should have special rows in this table for the values  $\frac{1}{2}\pi$  etc because at those points one or more of the functions will be zero or unbounded.)

Then we remember what happens when we replace  $x$  by  $-x$ ,  $x + \pi/2$  or  $x + \pi$ :

$$\begin{aligned}\cos(-x) &= \cos x, & \sin(-x) &= -\sin x, \\ \cos(x + \frac{\pi}{2}) &= -\sin x, & \sin(x + \frac{\pi}{2}) &= \cos x, \\ \cos(x + \pi) &= -\cos x, & \sin(x + \pi) &= -\sin x.\end{aligned}\tag{1.1}$$

These are very easy to derive from  $e^{ix} = \cos x + i \sin x$ , remembering that  $e^{i\pi/2} = i$ ,  $e^{i\pi} = -1$ . Using them in combination we can get

$$\cos(\pi - x) = -\cos x, \quad \sin(\pi - x) = \sin x.$$

and so on.

More generally

$$\cos(x + (m + \frac{1}{2})\pi) = (-1)^{(m+1)} \sin x, \quad \sin(x + (m + \frac{1}{2})\pi) = (-1)^m \cos x,\tag{1.2}$$

$$\cos(x + n\pi) = (-1)^n \cos x, \quad \sin(x + n\pi) = (-1)^n \sin x.\tag{1.3}$$

where  $m$  and  $n$  are integers. These identities enable us to relate the value we want to a value in the first quadrant (i.e. the range  $[0, \frac{1}{2}\pi]$ ). Remember the special cases for  $x = 0$ ,

$$\cos(n\pi) = (-1)^n, \quad \sin((n + \frac{1}{2})\pi) = (-1)^n.\tag{1.4}$$

$$\cos((n + 1/2)\pi) = 0, \quad \sin(n\pi) = 0.\tag{1.5}$$

which will turn up regularly later on.

## 1.1.2 Identities for the trigonometric functions

The most important formulae to remember are

$$\sin^2 A + \cos^2 A = 1\tag{1.6}$$

$$\cos(A + B) = \cos A \cos B - \sin A \sin B\tag{1.7}$$

$$\sin(A + B) = \sin A \cos B + \cos A \sin B.\tag{1.8}$$

If you have trouble remembering which of the last two is which, and which has the minus in it, try substituting some special values such as  $A = 0$  or  $B = \frac{1}{2}\pi$  and checking the result. For example, taking  $A = 0$  in the last equation gives  $\sin B = 0 + \sin B$ , consistent, whereas if you had tried  $\sin(A + B) = \sin A \cos B - \cos A \sin B$  you would get  $\sin B = 0 - \sin B$ , clearly wrong. From these and the earlier results Eq. 1.1 we get

$$\cos(A - B) = \cos A \cos B + \sin A \sin B$$

$$\sin(A - B) = \sin A \cos B - \cos A \sin B.$$

and by adding or subtracting various pairs of the above equations, we get

$$\cos A \cos B = \frac{1}{2}(\cos(A + B) + \cos(A - B))\tag{1.9}$$

$$\sin A \sin B = \frac{1}{2}(\cos(A - B) - \cos(A + B))\tag{1.10}$$

$$\sin A \cos B = \frac{1}{2}(\sin(A + B) + \sin(A - B)),\tag{1.11}$$

which we will find very useful in doing integrations like  $\int \cos(nx) \cos(mx) dx$  which turn up later on.

The double angle cases

$$\begin{aligned}\sin 2x &= 2 \sin x \cos x \\ \cos 2x &= \cos^2 x - \sin^2 x = 2 \cos^2 x - 1 = 1 - 2 \sin^2 x \\ \cos^2 x &= \frac{1}{2}(1 + \cos 2x) \\ \sin^2 x &= \frac{1}{2}(1 - \cos 2x)\end{aligned}$$

come up often; we get the “half angle” cases by just substituting in  $y = 2x, x = y/2$  in the above.

Using the basic identities we can easily derive plenty more, such as

$$\begin{aligned}\sec^2 A &= 1 + \tan^2 A \\ \cos C + \cos D &= 2 \cos \frac{1}{2}(C + D) \cos \frac{1}{2}(C - D).\end{aligned}$$

We should also note (see Thomas 3.4) that for any constant  $k$ ,

$$\frac{d(\sin(kx))}{dx} = k \cos(kx), \quad \frac{d(\cos(kx))}{dx} = -k \sin(kx).$$

Both  $\sin(kx)$  and  $\cos(kx)$  therefore obey<sup>1</sup>

$$\frac{d^2 y}{dx^2} = -k^2 y.$$

and it can be proved that these give all solutions, i.e.

$$\frac{d^2 y}{dx^2} = -k^2 y \quad \Leftrightarrow \quad y = a \cos(kx) + b \sin(kx) \quad (1.12)$$

for some constants  $a$  and  $b$ . We could also write the right side as a combination of  $e^{ix}$  and  $e^{-ix}$ .

## 1.2 Ln, or $\log_e$ , exp, and hyperbolic functions

(See Thomas section 7.2)

The natural logarithm  $\ln x$  can be defined as

$$\ln x = \int_1^x \frac{dt}{t}.$$

This implies  $\ln 1 = 0$ . Note that this is not a good definition if  $x < 0$ , but it is easy to show that for negative  $x$ ,  $\int^x du/u = \ln|x| + \text{constant}$ . The number  $e$  (Euler’s number) is then defined by  $\ln e = 1$ . (After  $\pi$ , this is the second most important constant in maths).

From the definition it is obvious that

$$\frac{d \ln x}{dx} = \frac{1}{x}.$$

One also finds:

$$\ln(ab) = \ln a + \ln b,$$

Repeated application of this shows  $\ln(a^n) = n \ln a$  for integer  $n$ , and it turns out this is true for any power  $p$  i.e.

$$\ln x^p = p \ln x,$$

---

<sup>1</sup>Those who have done applied maths. at A-level or later may recognize this as an equation for simple harmonic motion.

In particular, either putting  $p = -1$  in the above, or  $b = 1/a$  in the previous equation, gives us

$$\ln(a^{-1}) = \ln(1/a) = -\ln(a)$$

and hence

$$\ln(a/b) = \ln a - \ln b .$$

**Note** that  $\ln(a+b) \neq \ln a + \ln b$  (unless  $a+b=ab$ ).

We can define  $\exp$  (see Thomas 7.3) to be the inverse function to  $\ln$ , so that  $\exp(\ln x) = \ln(\exp x) = x$ . Then  $\exp 1 = e$  and  $\exp r = e^r$ . Note that  $\exp(a+b) = e^a e^b$  NOT  $e^a + e^b$ . For any number  $a$ ,  $a = e^{\ln a}$  and hence  $a^x = (e^{\ln a})^x = e^{x \ln a}$ . In particular, this enables us to relate the usual logarithms (base 10) to natural logarithms since if  $x = \log_{10} y$ ,  $y = 10^x = e^{x \ln 10}$ , so  $\ln y = x \ln 10$  and  $x = \ln y / \ln 10$ . For a general  $a$  we can define  $\log_a y = x$  to be such that  $y = a^x$ , so  $\ln y = \log_e y$ .

One can show that

$$\frac{d \exp x}{dx} = \exp x ;$$

to prove that, take  $y = \exp x$ , take  $\ln$  of both sides so  $\ln y = x$ , then differentiate giving  $(1/y)dy/dx = 1$  (by the chain rule), and rearrange.

We can now use  $e^x$  to define the hyperbolic functions (see Thomas 7.8)

$$\cosh x = \frac{1}{2}(e^x + e^{-x}), \quad \sinh x = \frac{1}{2}(e^x - e^{-x}) .$$

These functions have identities and derivative properties that run closely parallel to those of  $\sin$  and  $\cos$ . If you know the trigonometric identities, the identities for hyperbolic functions can be recovered by substituting  $\cosh$  for  $\cos$  and  $i \sinh$  for  $\sin$ , where  $i^2 = -1$ .

From differentiating  $e^x$  we find

$$\frac{d \sinh kx}{dx} = k \cosh kx, \quad \frac{d \cosh kx}{dx} = k \sinh kx .$$

Thence

$$\frac{d^2 y}{dx^2} = k^2 y \Leftrightarrow y = a \cosh(kx) + b \sinh(kx) \tag{1.13}$$

for some constants  $a$  and  $b$  (we can also write  $y$  as a combination of  $e^x$  and  $e^{-x}$ ).

Comparing Eq. 1.13 with 1.12, we now see how to solve  $d^2 y/dx^2 = Cy$  for any constant  $C$ : if  $C$  is positive, we define  $k = \sqrt{C}$  and get 1.13, while if  $C$  is negative we define  $k = \sqrt{-C}$  and get 1.12; finally if  $C = 0$  we easily integrate twice to get  $y = ax + b$ .

### 1.3 Double and triple integrals

(See Thomas 15.1 and 15.4)

First let us revise the idea of 2-D integration.

**Example 1.1.** Integrate the function  $f(x,y) = x^2 y^2$  over the triangular area  $\mathcal{R}: 0 \leq x \leq 1, 0 \leq y \leq x$ .

We can write this integral as

$$\int_{\mathcal{R}} f(x,y) dA,$$

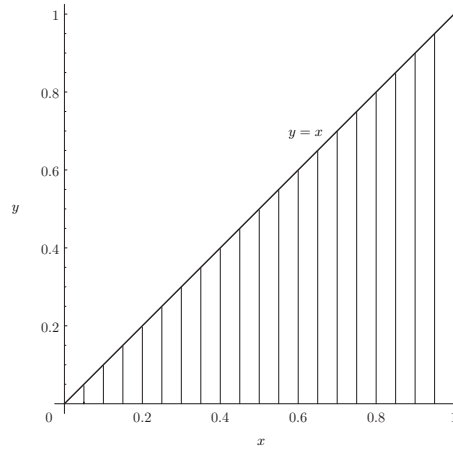


Figure 1.1: Integrating over the triangular region  $\mathcal{R} : 0 \leq x \leq 1, 0 \leq y \leq x$ .

where  $dA$  is an area element. But the area of a little rectangle of length  $\delta x$  in the  $x$ -direction and length  $\delta y$  in the  $y$ -direction is  $\delta A = \delta x \delta y$ ; hence we can rewrite  $dA$  as  $dA = dx dy$ . Thus the integral we want (cf. Fig. 1.1) is

$$\begin{aligned} \iint f(x,y) dx dy &= \int_{x=0}^1 \left( \int_{y=0}^x x^2 y^2 dy \right) dx \\ &= \int_{x=0}^1 \left( x^2 \int_{y=0}^x y^2 dy \right) dx \\ &= \int_0^1 x^2 \left( \frac{1}{3} x^3 \right) dx \\ &= \left[ \frac{1}{18} x^6 \right]_0^1 = \frac{1}{18}. \end{aligned}$$

Here there are two key points to note: the limits on the (inner)  $y$ -integral depend on  $x$ , and in the second step we have moved the  $x^2$  outside the  $y$  integral because it does not depend on  $y$ ; so, the  $x^2$  behaves like a “constant” inside the  $y$ -integral, but not for the  $x$  integral.

---

An area integral such as this is often called a double integral (because it can be rewritten as two 1-D integrations). Some authors use two integration signs, to remind you that it is an area integral: thus they would write  $\iint f(x,y) dA$ . In this course, **whenever it is obvious that an integral is over area, we shall generally just write  $\int f(x,y) dA$ .**

Similarly some books write  $\iiint f(x,y,z) dV$  for a volume integral: **where no confusion will arise, we shall just write  $\int f(x,y,z) dV$ .**

We shall need to put in all the integral signs when obtaining a value by doing the two or three integrations with respect to coordinates.

**Exercise 1.1.** Calculate  $\int_{\mathcal{R}} f(x,y) dA$  for

$$f(x,y) = 1 - 6x^2y \quad \text{and} \quad \mathcal{R} : 0 \leq x \leq 2, -1 \leq y \leq 1.$$

[Answer: 4]

□

Note that in that exercise, the region of integration is a rectangle, so the limits of both the  $x$ - and  $y$ -integrations were constants so one could do the  $x$ - or the  $y$ -integration first – the answer will be the same. This holds for any rectangular region in 2-D, or for a cuboid when we come to 3-D integration.

In example 1.1, we had a triangle; now the upper limit of the “inside” integral depends on the “outer” variable. upper limit of the  $y$ -integral was  $x$ , so the  $y$ -integration had to be performed first with the limits as given. Otherwise the answer would have read

$$\int \int f(x,y) dx dy = \int_{y=0}^x \left( \int_{x=0}^1 x^2 y^2 dx \right) dy = \frac{x^3}{9}.$$

This depends on  $x$ , which is ridiculous as the answer is for a whole area, not some value of  $x$ . If we want to change the order of integration we need to take  $y$  going from 0 to 1, then  $x$  runs from  $y$  to 1 (check the sketch); now we have to put the  $dx$  integral on the inside, and we get

$$\int \int f(x,y) dx dy = \int_{y=0}^1 \left( \int_{x=y}^1 x^2 y^2 dx \right) dy.$$

Check that this does give the same answer.

It’s important that you understand how to get these limits: when doing a numerical evaluation of a multiple integral, there are several rules to remember :

- i) Work out the limits on each variable from a sketch.
- ii) The limits on each integral may depend on the variables  $x, y$  etc appearing as  $dx, dy$  **outside** that integral, but should not depend on those inside it. So the limits on the outermost integral sign should not depend on any of  $x, y, z$ ; if  $dx$  is the outermost integral
- iii) The limits on each integral apply to the “matching” variable, again working from inside to outside. So the last integral sign matches the first one of  $dx, dy, dz$  etc.
- iv) Evaluate the resulting multiple integral from the “inside out”, so you evaluate the innermost integration first. Putting in brackets can be helpful here, as in the example above.

It is a straightforward step from double integrals to volume integrals (triple integrals) of the form  $\int_V f(x, y, z) dV$ . In Cartesian coordinates we have  $dV = dx dy dz$  (the volume of a 3-D rectangular box) and so

$$\int_V f(x, y, z) dV = \int \int \int_V f(x, y, z) dx dy dz.$$

Sometimes the geometry of the volume will make other choices of coordinate system preferable. In Thomas 15.3 and 15.6, which were studied in Calculus II, two-dimensional plane integrals in polar coordinates, and triple integrals in spherical and cylindrical polars are discussed: you will find it very useful to revise those sections. For a general change of coordinate system from Cartesians  $(x, y, z)$  to  $(u, v, w)$ ,

$$\int \int \int_V f dx dy dz = \int \int \int_V f J du dv dw$$

where  $J$  is the Jacobian determinant of  $(x, y, z)$  with respect to  $(u, v, w)$ . This determinant  $J$  is the volume ratio of the two coordinate systems: if we take an infinitesimal cuboid in  $(u, v, w)$  space of volume  $du dv dw$ , this will map to a parallelepiped in  $x, y, z$  space, and  $J$  is the ratio of those volumes (if you need to revise this in more detail, see Thomas 15.7).

**Exercise 1.2.** Evaluate  $\int \int \int e^x dx dy dz$  over the volume  $V$  of the tetrahedron bounded by the four planes  $x = 0, y = 0, z = 0$  and  $x + y + z = a$  ( $a > 0$ ). [Answer:  $e^a - \frac{1}{2}a^2 - a - 1$ .] □

## 1.4 Curves and surfaces

We shall use various geometrical shapes in examples, so we need the equations for them. The main ones are so-called ‘conic sections’ in two dimensions, and related three-dimensional surfaces. Other courses also discuss more complicated shapes (see e.g. Thomas 10.6 and 10.7).

First we discuss curves in two dimensions. There are three main ways to specify a curve:

One way is to give an equation  $y = f(x)$ ; a second way is to give an equation  $g(x,y) = 0$ : the curve is then the set of points  $(x, y)$  obeying the equation. Given the first form, we can get the second by defining  $g(x,y) = y - f(x)$ , but not necessarily the converse

A third way is the **parametric form** in terms of two functions of some variable  $t$ :  $x = a(t)$ ,  $y = b(t)$  (see Thomas 3.5). Sometimes we can take  $t = x$  itself. The parametrized form carries extra information, about which direction and how fast we go along the curve as  $t$  changes.<sup>2</sup> We will see a lot more examples of the parametrised form in Chapter 2

Using the second way, some standard curves are:

$$x^2 + y^2 = a^2 \quad \text{circle, centre } (0, 0), \text{ radius } a \quad (1.14)$$

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad \text{ellipse, centre } (0, 0), \text{ semi-major axes } a \text{ and } b \quad (1.15)$$

$$y = ax^2 + b \quad \text{parabola, symmetric about } x = 0 \quad (1.16)$$

$$cy^2 - kx^2 = a^2, (ck > 0) \quad \text{hyperbola, symmetric about } x = 0 \text{ and } y = 0 \quad (1.17)$$

(See Thomas 1.2, 1.5.) The special case of a hyperbola with  $a = 0$  is just a pair of straight lines. These curves involving only constants and powers up to  $x^2$  and  $y^2$  are known as the **conic sections**.

To recognize these, first look for the coefficients of the  $x^2$  and  $y^2$ :

if one is 0, but the corresponding variable appears linearly, it’s a parabola;

if they have the same sign it’s an ellipse (or as a special case a circle), and

if they have opposite sign it’s a hyperbola

(assuming the remaining constants allow there to be some points:  $x^2 + y^2 = -5$  has no real points).

What if the equation is quadratic but not one of these standard forms? Given

$$x^2 + 6x + y^2 + 8y = 0$$

we can carry out a process called ‘completing the square’ to write it as

$$(x + 3)^2 + (y + 4)^2 = 25$$

which we now recognize as a circle radius 5, centre  $(-3, -4)$ : this circle passes through the origin. Similar methods can be used to recognize the other standard curves if they are given relative to origins different from the ones used in the most standard forms below (cf. Thomas 1.5).

We can also recognize the case where the axes have been transformed, in a similar way. For example,  $xy = b^2 \Leftrightarrow (y+x)^2 - (y-x)^2 = 4b^2$ , so it’s a hyperbola where the symmetry axes are at  $45^\circ$  to those used in (1.17) with  $c = k = 1$  and with  $4b^2 = a^2$ . In general we have to complete the square on the terms quadratic in  $x$  and  $y$ : for example the rearrangement

$$x^2 + 4xy + 3y^2 = (x + 2y)^2 - y^2$$

shows the curve  $x^2 + 4xy + 3y^2 = 6$  is not an ellipse, as you might think from the fact the coefficients of  $x^2$  and  $y^2$  are both positive, but a hyperbola.

---

<sup>2</sup>The latter approach is used heavily in Geometry II.



Parametrized curves are also useful, especially when calculating **line integrals** along curves later on. Here are some standard parametrizations for the circle, ellipse and hyperbola:

$$x^2 + y^2 = a^2 \quad (x, y) = (a \cos \theta, a \sin \theta) \quad (1.18)$$

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (x, y) = (a \cos \theta, b \sin \theta) \quad (1.19)$$

$$y^2 - x^2 = a^2 \quad (x, y) = (a \sinh \theta, a \cosh \theta). \quad (1.20)$$

These work because of the identity (1.6) and its hyperbolic counterpart  $\cosh^2 x - \sinh^2 x = 1$ . (See Thomas 10.4 for further or alternative parametrizations.) We shall use these, especially the first two, later.

## 1.5 Surfaces in 3-D

[Here we meet material you may not have seen before.]

For surfaces in 3 dimensions, there are similarly three main ways to give the equations. One is to give one coordinate in terms of the other two, e.g.  $z = h(x, y)$ . Another is to use a single equation  $V(x, y, z) = 0$ . The third is by a parametrization in terms of two variables e.g.  $(x(u, v), y(u, v), z(u, v))$  (see Thomas 16.6, and more details in Chapter 2).

We shall again focus on surfaces described by quadratics in  $x$ ,  $y$  and  $z$  at worst. To work out what the surface is like, one good way is to consider letting one coordinate be constant, for example  $z = d$ , which means we are considering a “slice” through the surface  $V = 0$  at the plane  $z = d$ . The intersection of a curved surface and a plane is generally a 1-D curve, which we should be able to identify from the previous section. Then we just stack those curves for varying  $d$ .

One simple case is

$$x^2 + y^2 = a^2$$

The equation is the same as for a circle, but as we are now in 3 dimensions, it implies  $z$  can take any value. In each plane  $z = d$  we have a circle. Hence, this is an infinite circular cylinder along the  $z$ -axis. Very often some bounding values of  $z$  are given, e.g.  $0 \leq z \leq 2$ . Then we have a finite cylinder, the shape of a drinks can.

**Example 1.2.** What is the surface  $\frac{y^2}{a^2} + \frac{z^2}{b^2} = 1$ ?

It is an infinite elliptical cylinder along the  $x$  axis.

We can also have parabolic and hyperbolic “cylinders”, using (1.16) and (1.17).

Another simple three-dimensional surface is that of a **sphere** of radius  $a$  centred at the origin:

$$x^2 + y^2 + z^2 = a^2. \quad (1.21)$$

**Example 1.3.** What is the surface  $x^2 + y^2 + z^2 = a^2, x \geq 0$ ?

The hemisphere to the right of the plane  $x = 0$ .

We can put together cases where we get one of the standard types of curve listed earlier in planes  $z = d$  and different ones in planes  $x = k$  or  $y = m$  say.

For example, we can generalize the ellipse (1.15) to

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad (1.22)$$

(see Thomas, 12.6). In this case each of the three types of cut-plane gives an ellipse as the curve. The surface is an *ellipsoid*. This shape, and the ones that follow, are shown in diagrams 12.48-12.52 in Thomas. (also, the Wikipedia article on “Quadrics” has some pretty graphics).

If instead we had taken

$$\frac{z}{c} = \frac{x^2}{a^2} + \frac{y^2}{b^2}, \quad (1.23)$$

we then have an ellipse in each plane  $z = d$  but a parabola in each plane  $x = k$  or  $y = m$ . This is an *elliptic paraboloid*. Changing the plus to a minus in this equation gives a hyperbolic paraboloid.

Similarly we can obtain an (*elliptic*) *hyperboloid* as

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = +1. \quad (1.24)$$

Here we have ellipses in planes  $z = d$ , and hyperbolae in the planes  $x = 0$  and  $y = 0$ . Moving the  $z^2$  term over to the RHS, we see the RHS is positive for any  $z$ , so there is an ellipse for any fixed value of  $z$  and the surface has just one piece (we say ‘one sheet’).

However, if instead we had a  $-1$  on the right, i.e.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1. \quad (1.25)$$

we can rearrange into

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z^2}{c^2} - 1. \quad (1.26)$$

It’s now clear that if  $z^2/c^2 > 1$ , i.e.  $z < -c$  or  $z > c$ , we again get an ellipse in the  $xy$  plane; but if  $-c < z < c$  the RHS is negative and there are no solutions for  $x, y$ . This is a hyperboloid of two sheets.

The elliptic paraboloid and hyperboloid have circular special cases where  $a = b$ . Note also that we can swap  $x, y$  and  $z$  around in these forms so we have different choices of axes for the same shapes.

There is also the special case of the hyperboloid equation where the constant on the RHS is zero, i.e.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0 \quad (1.27)$$

This is a cone through the origin. Taking a plane through the origin such as  $x = 0$ , we get two straight lines, while taking planes perpendicular to the axes but not through the origin gives ellipses or parabolae. In fact all the quadratic curves (ellipses, circles, parabolae and hyperbolae) can be obtained by intersecting the circular cone  $z^2 = x^2 + y^2$  with planes (not necessarily perpendicular to the axes): this is why they are called conic sections (see Thomas chapter 10).

If we are given a quadratic surface in a different form, we can first rearrange it into one of the forms above: rearrange so that all the  $x, y, z$  terms are on the left, and the constant on the right; if the constant is not zero, divide by it to get a  $+1$  on the right; then look at the  $x^2, y^2$  and  $z^2$  parts: if two or three of these have negative coefficients, just multiply by  $-1$  to make at least two of the coefficients of  $x^2, y^2, z^2$  positive. Then,

If all 3 are positive, it’s an ellipsoid (or a sphere).

If two are positive and one negative it’s a hyperboloid, and we need to check the constant term

to see if it's one sheet or two

If one of the three  $x^2, y^2$  or  $z^2$  terms is zero, but there is a linear term in the corresponding variable, it's a paraboloid: the relative sign of the other two show if it is elliptic or hyperbolic.

If one variable is missing completely, it's a "cylinder" given by the matching 2-D curve.

As in the case of curves, we can work out what the shape is if the equations are not in standard form but have shifted origins or rotated axes, by completing the square. For this course, we'll keep it simple though, so we will only be looking at surfaces which are aligned with the coordinate axes.

Now for some parametrized versions (see Thomas 16.6)

	Implicit form	Parametric form	
Cylinder	$x^2 + y^2 = a^2$	$(x, y, z) = (a \cos \theta, a \sin \theta, z)$	Parameters $\theta, z$ (1.28)
Sphere	$x^2 + y^2 + z^2 = a^2$	$(a \sin \theta \cos \phi, a \sin \theta \sin \phi, a \cos \theta)$	Parameters $\theta, \phi$ (1.29)
Ellipsoid	$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = +1$	$(a \sin \theta \cos \phi, b \sin \theta \sin \phi, c \cos \theta)$	Parameters $\theta, \phi$ (1.30)
Hyperboloid	$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$	$(a \cos u, b \sin u \cosh v, c \sin u \sinh v)$	Parameters $u, v$ (1.31)

## 1.6 Vectors

(Note: this is in chapter 12 in Thomas but in Geometry I you used Hirst)

Vectors can be introduced as displacements in space, called **position vectors**. To describe a position vector, we need to specify its direction and its length or magnitude (to say how far we go in the given direction). This is a geometric definition. A vector is different from a **scalar**, a quantity which has only a magnitude but no direction.

One can draw a vector as an arrow of the appropriate length and direction. Vectors are usually notated in print by boldface type, e.g. **a**, and in handwriting by under- or over-lining such as  $\underline{a}$ ,  $\vec{a}$ , or  $\bar{a}$ .

**Warning:** When writing, it is tempting to miss off the under/overlines to save time. This is a bad idea, because if you confuse what's a scalar and what's a vector in your working, you immediately get nonsense.

To define a vector algebraically, i.e. in a formula, we can use the Cartesian coordinates of the point to which it displaces the origin, e.g.

$$\mathbf{r} = (x, y, z). \tag{1.32}$$

**Note:** As you saw in Geometry I, we can write vectors either as row or column vectors. The column vector form is useful if you are multiplying by matrices (like rotation matrices), but in this course we shall mainly use the row vector form which is more compact.)

Here  $x, y$  and  $z$  are called the *components* of  $\mathbf{r}$ . We may refer to  $(x, y, z)$  as the point  $\mathbf{r}$ . From now on we shall use the notation  $\mathbf{r}$  only for this vector.

The length of a vector  $\mathbf{v}$  is denoted by  $|\mathbf{v}|$  or sometimes just  $v$ ; this is a scalar. The vector  $\mathbf{r}$  has length  $r = \sqrt{x^2 + y^2 + z^2}$ , by Pythagoras' theorem in 3 dimensions.

To add vectors  $\mathbf{a}$  and  $\mathbf{b}$  we simply take the displacement obtained by displacing first by  $\mathbf{a}$  and then by  $\mathbf{b}$  (the result can be defined as the diagonal of the parallelogram with sides  $\mathbf{a}$  and  $\mathbf{b}$ ). In components this says that  $\mathbf{v} = (v_1, v_2, v_3)$  and  $\mathbf{w} = (w_1, w_2, w_3)$  have the sum

$$\mathbf{v} + \mathbf{w} = (v_1 + w_1, v_2 + w_2, v_3 + w_3).$$

Subtraction can then be defined similarly. The zero vector  $\mathbf{0}$  is the one with zero magnitude (and no well-defined direction!).

It is now easy to show this obeys the usual rules of addition (and subtraction).<sup>3</sup>

The displacement from a point  $\mathbf{r}_1$  to a point  $\mathbf{r}_2$  is  $\mathbf{r}_2 - \mathbf{r}_1$ .

We can multiply a vector by a scalar (a number)  $\lambda$ , simply by multiplying its magnitude, preserving the direction. In components, if  $\mathbf{v} = (v_1, v_2, v_3)$  then we have

$$\lambda \mathbf{v} = (\lambda v_1, \lambda v_2, \lambda v_3).$$

This operation also obeys very simple and obvious rules.<sup>4</sup> This multiplication gives us a way to define the unit vector (the vector of length 1) in the same direction as  $\mathbf{v}$ , denoted by  $\hat{\mathbf{v}}$ , by  $\hat{\mathbf{v}} \equiv \mathbf{v}/|\mathbf{v}|$  (strictly, we should write the number first so we would have to write  $(1/|\mathbf{v}|)\mathbf{v}$ , but in practice it's obvious what we mean).

These rules give us another common way of writing a vector. We note that we can arrive at the same total displacement by first moving along the  $x$ -axis, then parallel to the  $y$ -axis then parallel to the  $z$ -axis; and we can express this by defining the unit vectors  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$  along the directions of the three axes by

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}.$$

This way of writing (1.32) has the advantage of making it clearer how the components change if we change our choice of axes: if we rotate our axes to a different system  $x', y', z'$ , we will get 3 new unit vectors e.g.  $\mathbf{i}'$ ,  $\mathbf{j}'$  and  $\mathbf{k}'$ , and converting vectors between systems looks like a matrix multiplication - more on this later.

Note that all of these statements about position vectors in 3 dimensions can very simply be applied in 2 dimensions also, with obvious minor changes.

Although we have motivated vectors by introducing them as displacements, they can represent, or be interpreted as, many other things: for example, a force, a velocity, inputs and outputs in an economic model, and so on.

A parametric equation of the type

$$\mathbf{r} = \mathbf{p} + t\mathbf{q}, \quad -\infty < t < \infty \quad (1.33)$$

defines a line through point  $\mathbf{p}$  parallel to direction  $\mathbf{q}$ . For example  $\mathbf{r} = t\mathbf{k}$ ,  $-\infty < t < \infty$  is the  $z$  axis.

Using this, we can get the straight line going through two given points  $\mathbf{r}_1$  and  $\mathbf{r}_2$ : the vector from  $\mathbf{r}_1$  to  $\mathbf{r}_2$  is  $\mathbf{r}_2 - \mathbf{r}_1$ , so the (infinite) line through them is

$$\mathbf{r} = \mathbf{r}_1 + t(\mathbf{r}_2 - \mathbf{r}_1), \quad -\infty < t < \infty \quad (1.34)$$

If instead we take a range  $0 \leq t \leq 1$  in the above, this gives us the finite line segment with end-points at the two given points. This will be very useful later on, *memorise it*.

---

<sup>3</sup>This means that for any vectors  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$ ,

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}, \quad (\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}), \quad \exists \mathbf{0} \text{ such that } \mathbf{a} + \mathbf{0} = \mathbf{a},$$

and given  $\mathbf{a}$ ,  $\exists(-\mathbf{a})$  such that  $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$ . These rules are purely abstract and make no reference to displacements or three dimensions, and are part of the general definition of a vector space which is given in Linear Algebra I. Those who have encountered groups will recognise that they ensure that the space of vectors is an additive group under vector addition.

<sup>4</sup>More precisely, for any vectors  $\mathbf{a}$  and  $\mathbf{b}$ , and numbers  $\lambda$  and  $\mu$ , we have

$$\lambda(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}, \quad (\lambda + \mu)\mathbf{a} = \lambda\mathbf{a} + \mu\mathbf{a}, \quad (\lambda\mu)\mathbf{a} = \lambda(\mu\mathbf{a})$$

and  $1\mathbf{a} = \mathbf{a}$ . For a general vector space, as defined in Linear Algebra I, the scalars are elements of a general field but here we shall only use the real numbers  $\mathbb{R}$ . However, these rules do apply when  $\lambda$  and  $\mu$  are elements of a general field, for instance the complex numbers  $\mathbb{C}$ .

---

---

**Example 1.4.** Medians of a triangle

Vectors can often be used to derive geometrical results very concisely, as this example shows.

Let  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  be the corners of a triangle. The midpoint of the side connecting  $\mathbf{b}$  and  $\mathbf{c}$  will be  $\frac{1}{2}(\mathbf{b} + \mathbf{c})$ . A line through this midpoint and  $\mathbf{a}$  is

$$\mathbf{r} = \mathbf{a} + t\left(\frac{1}{2}\mathbf{b} + \frac{1}{2}\mathbf{c} - \mathbf{a}\right), \quad -\infty < t < \infty$$

which is called the median through  $\mathbf{a}$ . Putting  $t = \frac{2}{3}$  (note: here this choice is a rabbit out of the hat, but we can find it by writing down a second median and solving for the intersection point) we get the point  $\frac{1}{3}(\mathbf{a} + \mathbf{b} + \mathbf{c})$ . Since this point is symmetric in  $\mathbf{a}, \mathbf{b}, \mathbf{c}$ , the medians through  $\mathbf{b}$  and  $\mathbf{c}$  will also pass through it. Hence the three medians of a triangle intersect at a single point.

---

---

If we write out the components of (1.33), with notation  $\mathbf{r} = (x, y, z)$ ,  $\mathbf{p} = (p_1, p_2, p_3)$ ,  $\mathbf{q} = (q_1, q_2, q_3)$  we find

$$x = p_1 + tq_1 \quad y = p_2 + tq_2, \quad z = p_3 + tq_3,$$

from which we can eliminate  $t$  to get

$$\frac{x - p_1}{q_1} = \frac{y - p_2}{q_2} = \frac{z - p_3}{q_3},$$

giving the two independent linear equations (e.g. for  $y$  and  $z$  in terms of  $x$ ) needed for a line in three-dimensional space.

We can now write functions of 3-dimensional position  $f(x, y, z)$  more compactly as functions  $f(\mathbf{r})$ . Equations of the form  $f(\mathbf{r}) = \text{constant}$  define surfaces, the constant surfaces of  $f$ . A simple example is  $r^2 = 1$ , which is a sphere of unit radius centred at the origin. (Recall our notation allows  $r \equiv |\mathbf{r}|$ .)

---

---

**Example 1.5.** A sphere

The geometrical interpretation of

$$|\mathbf{r} - \mathbf{k}| = 1$$

as a sphere of unit radius centred at  $(0, 0, 1)$  is obvious. Equivalent expressions are  $x^2 + y^2 + (z - 1)^2 = 1$  and  $x^2 + y^2 + z^2 - 2z = 0$ .

---

---

**Warning:** One of the commonest errors made by students is to confuse vectors and scalars, in particular to start adding together the components of a vector. The vector  $(3, 1, 2)$  is not the same as the scalar 6. This may seem obvious now, but the mistake is more easily made when using basis vectors like  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$ ; then it somehow seems to be easier to make the mistake  $3\mathbf{i} + \mathbf{j} + 2\mathbf{k} = 6$ .

## 1.7 Scalar and vector products

We have defined vector addition and subtraction, but not multiplication of vectors. This is more complicated because to obtain another vector we need to define both a magnitude and a direction (and in general, vector division cannot be defined at all; we can divide a vector by a scalar  $\lambda$  just by multiplying by  $1/\lambda$ , but we cannot divide anything by a vector).

We first define the dot product, or scalar product<sup>5</sup>, whose result is not a vector but a scalar. For vectors  $\mathbf{v}$  and  $\mathbf{w}$ , this is defined by

$$\mathbf{v} \cdot \mathbf{w} \equiv |\mathbf{v}||\mathbf{w}|\cos\theta, \quad (1.35)$$

where  $\theta$  is the angle between  $\mathbf{v}$  and  $\mathbf{w}$ . An alternative definition in terms of the components  $(v_1, v_2, v_3)$  and  $(w_1, w_2, w_3)$  of  $\mathbf{v}$  and  $\mathbf{w}$  is

$$\mathbf{v} \cdot \mathbf{w} \equiv v_1w_1 + v_2w_2 + v_3w_3 = \sum_{i=1}^3 v_iw_i.$$

One can prove that the two definitions are the same by applying Pythagoras' theorem to a triangle constructed as follows. Take sides  $\mathbf{v}$ ,  $\mathbf{w}$  and  $\mathbf{v} + \mathbf{w}$ . Draw the perpendicular from  $\mathbf{v} + \mathbf{w}$  to the line in direction  $\mathbf{v}$ . It has height  $|\mathbf{w}|\sin\theta$  and meets the direction  $\mathbf{v}$  at a distance  $|\mathbf{v}| + |\mathbf{w}|\cos\theta$ . Now write out Pythagoras with the lengths in terms of  $|\mathbf{v}|$ ,  $|\mathbf{w}|$  and  $\theta$  and again in terms of components and compare the results. The details are left as an exercise (if you have trouble, look in the online notes for MAS114 Geometry I or in A.E. Hirst, Vectors in 2 or 3 dimensions, Arnold 1995, chapter 3).

We note in particular that two non-zero vectors  $\mathbf{v}$  and  $\mathbf{w}$  are perpendicular ( $\theta$  is a right angle) if and only if  $\mathbf{v} \cdot \mathbf{w} = 0$ .

---

---

**Example 1.6.** (This example was used in Geometry I.)

Find  $\cos\theta$  where  $\theta$  is the angle between  $\mathbf{v} = (1, 3, -1)$  and  $\mathbf{w} = (2, 2, 1)$ .

$$\begin{aligned} \mathbf{v} \cdot \mathbf{w} &= 1 \cdot 2 + 3 \cdot 2 + (-1) \cdot 1 = 7 = |\mathbf{v}||\mathbf{w}|\cos\theta, \\ |\mathbf{v}|^2 &= 1^2 + 3^2 + (-1)^2 = 11, \\ |\mathbf{w}|^2 &= 2^2 + 2^2 + 1^2 = 9, \text{ so} \\ \cos\theta &= \frac{7}{\sqrt{11}\sqrt{9}} = \frac{7}{3\sqrt{11}} \end{aligned}$$

---

<sup>5</sup>In a more abstract setting (such as in Linear Algebra I) this may also be called the inner product.

---

From either form of the definition we can easily derive various algebraic rules.<sup>6</sup>

A geometrical application of the dot product is in giving the equation of a plane. The plane through a fixed point  $\mathbf{p}$  perpendicular to a fixed vector  $\mathbf{v}$  is given by the set of all points  $\mathbf{r}$  which have  $\mathbf{r} - \mathbf{p}$  perpendicular to  $\mathbf{v}$ , as is easily seen from a sketch. Since two perpendicular vectors have a dot product of zero, this gives

$$(\mathbf{r} - \mathbf{p}) \cdot \mathbf{v} = 0 \quad (1.36)$$

This easily rearranges to  $\mathbf{r} \cdot \mathbf{v} = \mathbf{p} \cdot \mathbf{v}$  and the right-hand side is just a constant for given  $\mathbf{p}, \mathbf{v}$ .

In components, if  $\mathbf{v} = (a, b, c)$  and  $\mathbf{p} \cdot \mathbf{v} = d$  the equation for a plane reads  $ax + by + cz = d$ . In practice people often choose a unit vector  $\mathbf{n}$  when specifying a plane in this form, so that  $\mathbf{p} \cdot \mathbf{n}$  becomes the perpendicular distance of the plane from the origin. Then, the distance of any other point  $\mathbf{r}_1$  from that plane is given by  $(\mathbf{r}_1 - \mathbf{p}) \cdot \mathbf{n} = \mathbf{r}_1 \cdot \mathbf{n} - \mathbf{p} \cdot \mathbf{n} = \mathbf{r}_1 \cdot \mathbf{n} - d$  (the sign here tells one which side of the plane  $\mathbf{r}_1$  is on).

**The vector product:** To define a product of two vectors which is a third vector, we need to define a direction from two vectors  $\mathbf{u}$  and  $\mathbf{v}$ . The only way to do this which treats the two vectors equally is to take the perpendicular to the plane in which  $\mathbf{u}$  and  $\mathbf{v}$  lie. However, this does not fully define a direction, because we need to know which way to go along the perpendicular. For that the convention is to use the so-called right-hand rule: hold the fingers of your right hand so they curl round from  $\mathbf{u}$  to  $\mathbf{v}$  and then take the direction your thumb points (see Thomas figures 12.27 and 12.28). If you do DIY, you may find it helpful to remember that this is the direction a normal screw travels if you turn your screwdriver clockwise. Note that this definition only works in **three dimensions**: there is no well-defined vector product in  $n$  dimensions for  $n > 3$ .

The magnitude of  $\mathbf{v} \times \mathbf{w}$  is defined to be  $|\mathbf{v}||\mathbf{w}|\sin\theta$  ( $\theta$  as before). Geometrically this is the area of a parallelogram with sides  $\mathbf{v}$  and  $\mathbf{w}$ . Note that for perpendicular vectors this rule implies that the magnitude is  $|\mathbf{v}||\mathbf{w}|$ . These rules have the consequences that for any vectors  $\mathbf{u}, \mathbf{v}$  and  $\mathbf{w}$  and any scalar  $\lambda$ ,

$$\begin{aligned} \mathbf{v} \times \mathbf{w} &= -\mathbf{w} \times \mathbf{v}, \\ (\lambda \mathbf{v}) \times \mathbf{w} &= \lambda(\mathbf{v} \times \mathbf{w}) = \mathbf{v} \times (\lambda \mathbf{w}), \\ \mathbf{u} \times (\mathbf{v} + \mathbf{w}) &= (\mathbf{u} \times \mathbf{v}) + (\mathbf{u} \times \mathbf{w}), \\ (\mathbf{u} + \mathbf{v}) \times \mathbf{w} &= \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}, \end{aligned}$$

and  $\mathbf{v} \times \mathbf{w} = \mathbf{0}$  for non-zero  $\mathbf{v}, \mathbf{w}$  if and only if  $\mathbf{v}$  and  $\mathbf{w}$  are parallel or anti-parallel (in particular,  $\mathbf{v} \times \mathbf{v} = \mathbf{0}$  for any  $\mathbf{v}$ ).

**Note:** it is particularly important to note the sign-change property that  $\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$ . This looks “silly”, but is a consequence of the “handedness” of three-dimensional space, and which way round we choose to label our three coordinate axes.

From the notation used, the vector product is often called the cross product.<sup>7</sup>

---

<sup>6</sup>The main ones are that  $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$  and that for any vectors  $\mathbf{u}, \mathbf{v}$  and  $\mathbf{w}$  and any scalar  $\lambda$ ,

$$\begin{aligned} \mathbf{v} \cdot (\lambda \mathbf{w}) &= \lambda(\mathbf{v} \cdot \mathbf{w}) = (\lambda \mathbf{v}) \cdot \mathbf{w}, \\ \mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) &= (\mathbf{u} \cdot \mathbf{v}) + (\mathbf{u} \cdot \mathbf{w}), \\ (\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} &= (\mathbf{u} \cdot \mathbf{w}) + (\mathbf{v} \cdot \mathbf{w}), \\ \mathbf{v} \cdot \mathbf{v} &= |\mathbf{v}|^2 \geq 0, \\ \mathbf{v} \cdot \mathbf{v} &= 0 \Leftrightarrow \mathbf{v} = \mathbf{0}. \end{aligned}$$

<sup>7</sup>You will also find that in some texts it is denoted  $\mathbf{v} \wedge \mathbf{w}$ , but I strongly advise against using this notation as it leads to confusion in more general settings where  $\mathbf{v} \wedge \mathbf{w}$  is not a vector. The reason for this misuse is that  $\mathbf{v} \wedge \mathbf{w}$  is what’s called a two-form, and there is an operation called the Hodge dual, denoted by  $*$ , such that in three dimensions  $*(\mathbf{v} \wedge \mathbf{w}) = \mathbf{v} \times \mathbf{w}$ .

To get the expressions for the cross product in terms of components, we can start by noting that the unit vectors  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$  are perpendicular to one another (so the vector product of any two distinct ones among them has magnitude 1). This means that  $\mathbf{i} \times \mathbf{j}$  must have length 1 and be perpendicular to both of them, so it is either  $+\mathbf{k}$  or  $-\mathbf{k}$ . Since the usual  $x$ ,  $y$  and  $z$  axes, in that order, are a right-handed set, it will turn out that

$$\mathbf{i} \times \mathbf{j} = +\mathbf{k}, \quad \mathbf{j} \times \mathbf{k} = +\mathbf{i}, \quad \mathbf{k} \times \mathbf{i} = +\mathbf{j},$$

and therefore

$$\mathbf{j} \times \mathbf{i} = -\mathbf{k}, \quad \mathbf{k} \times \mathbf{j} = -\mathbf{i}, \quad \mathbf{i} \times \mathbf{k} = -\mathbf{j}.$$

(To remember these, think of the sequence  $\mathbf{ijkijk}...$ ; if the two vectors in the cross-product are in the same order as in that sequence, the RHS has a  $+$  sign, while if they are in reverse order there is a  $-$  sign.)

Also  $\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = \mathbf{0}$ . Using these we easily obtain

$$(v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}) \times (w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k}) = (v_2w_3 - v_3w_2)\mathbf{i} + (v_3w_1 - v_1w_3)\mathbf{j} + (v_1w_2 - v_2w_1)\mathbf{k}. \quad (1.37)$$

This can also be written as the formal determinant

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

One geometrical use of the cross product is in forming the volume of a parallelepiped with sides  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$ . Thinking of (say)  $\mathbf{u}$  and  $\mathbf{v}$  as the base, and  $\theta$  as the angle between  $\mathbf{u} \times \mathbf{v}$  and  $\mathbf{w}$ , so that the height is  $|\mathbf{w}| \cos \theta$ , we see that

$$\text{Volume of parallelepiped} = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} \quad (1.38)$$

(positive if  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$  are a right-handed set). This quantity is called the *scalar triple product* and it is easy to show that

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \mathbf{v} \cdot (\mathbf{w} \times \mathbf{u}) = \mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})$$

(but this is  $-\mathbf{v} \cdot (\mathbf{u} \times \mathbf{w})$  etc, remember). We can also show that swapping the dot and cross gives the same result, i.e.  $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}) = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$  from above, but note that the brackets also move, i.e. the cross product must be done first (inside the brackets) otherwise the result is nonsense. (Some textbooks may omit the brackets, but this is potentially confusing). Clearly swapping the two vectors inside the bracket changes the sign, and we can show that this is also true for swapping any two of the three vectors.

**Exercise 1.3.** Prove from the definitions that, for all  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$ ,

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}.$$

This quantity is called the **vector triple product**; note that the position of the brackets matters here.

□



## 1.8 Gradients and directional derivatives

(See Thomas 14.5 [and 16.2])

In Calculus II you met functions of more than one variable. Those that were discussed there were *scalar functions*, i.e. functions whose value at a particular point is a number. Such a scalar quantity (magnitude but no direction) that depends on position in space is called a **scalar field**. An example would be the temperature in a room – it has magnitude but not direction (so it is a scalar), and it is (in general) a function of position.

Suppose that  $V(x, y, z)$  or  $V(\mathbf{r})$  is a scalar field defined in some region. Then we can define a vector, the **gradient** of  $V$ , at each point, which we denote  $\nabla V$ , as follows:

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

So the  $x$ -,  $y$ - and  $z$ -components of the new vector are  $\partial V/\partial x$ ,  $\partial V/\partial y$  and  $\partial V/\partial z$ . See Thomas 14.5 if you need to revise this in more detail. Sometimes instead of  $\nabla V$  we write  $\text{grad } V$ : the two notations are interchangeable.

---

**Example 1.7.** If  $V(x, y, z) = x^2 \sin z$ , calculate  $\nabla V$ .

In this example,  $\partial V/\partial x = 2x \sin z$ ,  $\partial V/\partial y = 0$  and  $\partial V/\partial z = x^2 \cos z$ . Hence

$$\nabla V = 2x \sin z \mathbf{i} + x^2 \cos z \mathbf{k}.$$

---

**Exercise 1.4.** Evaluate the gradient  $\nabla f$  of the following scalar fields.

(a)  $f = x + y + z$ ,

(b)  $f = yx^2 + y^3 - y + 2x^2z$ ,

(c)  $f = \mathbf{a} \cdot \mathbf{r}$ , where  $\mathbf{a}$  is a constant vector.

□

Now  $\nabla V$  tells us how  $V$  changes if we move from one point to a nearby point. Suppose we start at a point  $\mathbf{r} = (x, y, z)$ , and then move a small distance  $d\mathbf{r} = (dx, dy, dz)$  to the new point  $\mathbf{r} + d\mathbf{r} = (x + dx, y + dy, z + dz)$ : we will get a small change in  $V$ , given by

$$\begin{aligned} dV &\equiv V(x + dx, y + dy, z + dz) - V(x, y, z) \\ &= \frac{\partial V}{\partial x} dx + \frac{\partial V}{\partial y} dy + \frac{\partial V}{\partial z} dz \end{aligned}$$

Here the second line uses the Taylor series in more than one variable, and discards terms in second and higher derivatives since  $d\mathbf{r}$  is small. But  $(dx, dy, dz) = d\mathbf{r}$ , and so the right-hand side is just  $\nabla V \cdot d\mathbf{r}$ . Hence for a small change  $d\mathbf{r}$ , the change in  $V$  is

$$dV = \nabla V \cdot d\mathbf{r} \tag{1.39}$$

**Note:** to use this, you **must** evaluate  $\nabla V$  at the point concerned.

In our original definition of  $\text{grad}$ , it was implicitly assumed that we were working in terms of some specified Cartesian coordinate system  $(x, y, z)$ . Equation (1.39) is important, because we can use it as a more

fundamental definition of grad, which will enable us to write down  $\nabla V$  in more general coordinate systems. We shall return to this point later, in Chapter 5

Next, consider a surface

$$V(\mathbf{r}) = \text{constant},$$

and suppose that the point  $\mathbf{r}_1$  is on that surface. If  $d\mathbf{r}$  is a displacement *on* this surface,  $V(\mathbf{r}_1 + d\mathbf{r}) = V(\mathbf{r}_1)$ . Thus  $dV = \nabla V|_{\mathbf{r}_1} \cdot d\mathbf{r} = 0$ . Since this applies for every small displacement  $d\mathbf{r}$  in the surface,  $\nabla V|_{\mathbf{r}_1}$  must be perpendicular to the surface at  $\mathbf{r}_1$ . This gives us a way of finding a normal to a surface when the surface is specified by a single equation: the unit normal  $\mathbf{n}$  will be  $\nabla V/|\nabla V|$  evaluated at the point concerned.

Suppose we now want a normal line to  $V = \text{constant}$ , or a tangent plane (see Thomas 14.6). As we know from (1.33) and (1.36), a line through point  $\mathbf{a}$  in direction  $\mathbf{q}$  can be written in parametric form as

$$\mathbf{r} = \mathbf{a} + t\mathbf{q}, \quad -\infty < t < \infty,$$

while the plane through  $\mathbf{a}$  perpendicular to  $\mathbf{q}$  is

$$(\mathbf{r} - \mathbf{a}) \cdot \mathbf{q} = 0.$$

so all we have to do is insert values of  $\mathbf{a}$  and  $\mathbf{q}$  in these formulae. For the tangent plane and normal line to a surface at a given point  $\mathbf{p}$ , this gives

$$(\mathbf{r} - \mathbf{p}) \cdot \nabla V|_{\mathbf{p}} = 0 \quad \text{and} \quad \mathbf{r} = \mathbf{p} + t\nabla V|_{\mathbf{p}}.$$

It is sometimes convenient to eliminate the parameter  $t$  for the normal line, which we can do by taking the cross product with  $\nabla V|_{\mathbf{p}}$ :

$$(\mathbf{r} - \mathbf{p}) \times \nabla V|_{\mathbf{p}} = \mathbf{0}.$$

Note that the forms  $(\mathbf{r} - \mathbf{p}) \cdot \mathbf{n} = 0$  and  $(\mathbf{r} - \mathbf{p}) \times \mathbf{n} = \mathbf{0}$ , using the unit normal  $\mathbf{n}$ , would give the same plane or line (though in  $\mathbf{r} = \mathbf{a} + t\mathbf{n}$  such a change alters the values of  $t$  for given points) so we need not calculate  $|\nabla V|$  to get the tangent plane or normal line.

**Repeated Note:** to use this, you **must** evaluate  $\nabla V$  at the point concerned.

**Exercise 1.5.** Find equations for the (i) tangent plane and (ii) normal line at the point  $P_0$  on each of the surfaces:

(a)  $x^2 + 3yz + 4xy = 27$ ,  $P_0 = (3, 1, 2)$ .

(b)  $y^2z + x^2y = 7$ ,  $P_0 = (2, 1, 3)$ .

[Answers: (a)  $10x + 18y + 3z = 54$ ,  $\mathbf{r} = (3 + 10t)\mathbf{i} + (1 + 18t)\mathbf{j} + (2 + 3t)\mathbf{k}$   
 (b)  $4x + 10y + z = 21$ ,  $\mathbf{r} = (2 + 4t)\mathbf{i} + (1 + 10t)\mathbf{j} + (3 + t)\mathbf{k}$  □

Suppose now that we want to calculate the rate of change of  $V(\mathbf{r})$  in a particular direction specified by the unit vector  $\mathbf{t}$ . Let  $s$  be the distance travelled in the direction of  $\mathbf{t}$ ; then  $d\mathbf{r} = \mathbf{t}ds$ . So  $dV = \nabla V \cdot \mathbf{t}ds$ . Hence we can conclude that the rate of change of  $V$  in the direction of  $\mathbf{t}$  is

$$\frac{dV}{ds} = \nabla V \cdot \mathbf{t} = \mathbf{t} \cdot \nabla V.$$

$\mathbf{t} \cdot \nabla V$  is called the *directional derivative*. Now

$$\nabla V \cdot \mathbf{t} = |\nabla V| |\mathbf{t}| \cos \theta = |\nabla V| \cos \theta,$$

where  $\theta$  is the angle between the vectors  $\nabla V$  and  $\mathbf{t}$ . This is maximized when  $\cos \theta = 1$ , i.e. when  $\theta = 0$ . Thus  $V$  changes most rapidly in the direction of  $\nabla V$ , and  $|\nabla V|$  is this most rapid rate of change. It is this property,

in the two-dimensional case, that gave rise to the name gradient, because  $|\nabla V|$  is the gradient of the surface given by  $z = f(x, y)$  in that case. Correspondingly the maximum decrease is when  $\mathbf{t}$  is opposite to  $\nabla V$ .

---

---

**Example 1.8.** Find the directions in which the function  $f(x, y, z) = (x/y) - yz$  increases and decreases most rapidly at the point  $P, (4, 1, 1)$ .

We can describe the directions in which  $f$  increases and decreases most rapidly by specifying the unit vectors in those directions. Now

$$\nabla f = \frac{1}{y}\mathbf{i} - \left(\frac{x}{y^2} + z\right)\mathbf{j} - y\mathbf{k} = (1, -5, -1) \text{ at } P.$$

The rate of change of  $f$  in the direction of unit vector  $\mathbf{t}$  is  $\nabla f \cdot \mathbf{t}$ . This has its maximum when  $\mathbf{t}$  is in the same direction as  $\nabla f$ ; so the direction  $\mathbf{t}$  in which  $f$  increases most rapidly is

$$\frac{\nabla f}{|\nabla f|} = \frac{1}{\sqrt{27}}(1, -5, -1).$$

and the actual rate is  $\sqrt{27}$ . The rate of *decrease* of  $f$  is greatest, at  $-\sqrt{27}$ , when  $\mathbf{t}$  is in the *opposite* direction, i.e.

$$\frac{-1}{\sqrt{27}}(1, -5, -1)$$

---

---

**Exercise 1.6.** Find the directional derivative of  $\Phi$  at the point  $(1, 2, 3)$  in the direction of the vector  $(1, 1, 1)$  where

$$\Phi = \frac{x^2}{3} + \frac{y^2}{9} + \frac{z^2}{27}.$$

□

We can write  $\nabla$  on its own as

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

and work with it like a vector field, although it is in fact not a vector field (since we cannot say what numerical value its components have at a particular point); strictly speaking  $\nabla$  is a **vector differential operator**. The name of the symbol  $\nabla$  is ‘nabla’ but often in speech we say ‘del’. It is easy to see how to take a two-dimensional version of  $\nabla$ . We will return to  $\nabla$  in Chapter 3, where we shall see how the  $\nabla$  operator can also be used to differentiate vectors.