Unit 7

Functions of several variables

Introduction

So far in this module, most of the mathematical models have concerned physical quantities that vary with respect to a single variable, which has usually been time. In this unit we look at models that depend on more than one variable.

As a simple example, let us consider the illumination from a street lamp, as shown in Figure 1. (That is, we are considering the illumination falling on, say, a flat paving slab at the point P, due to a street lamp placed at the point Q.) There are a number of variable quantities here. To simplify things a little, let us assume that the power rating of the light bulb is fixed at 1000 watts. Nevertheless, the height h of the bulb above the street, the distance x of P from the base of the lamp, the distance r from P to Q, and the angle θ (as shown) are all features that could be taken into consideration.

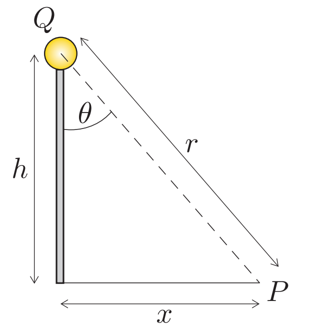


Figure 1 Street lamp placed at point Q, illuminating a paving slab at point P

It is possible to express h and x in terms of r and θ, or to express r and θ in terms of h and x. The quantities h and x are easy to measure, but the physics of the situation gives the illumination at P more readily in terms of the two variables r and θ, via the equation

$$I = \frac{1000\cos\theta}{r^2}, \qquad (1)$$

where I is the 'brightness' of the illumination per square metre falling on the slab at P.

The technical term for 'brightness' in this sense is *luminance*. The form of its dependence on r is due to the fact that the surface area of a sphere of radius r is proportional to r^2. The reason for its dependence on θ is that the light is falling obliquely on the slab.

In order to express I in terms of h and x rather than r and θ, we use the formulas $\cos\theta = h/\sqrt{h^2+x^2}$ and $r^2 = h^2 + x^2$, giving

$$I(h,x) = \frac{1000h}{(h^2+x^2)^{3/2}}. \qquad (2)$$

In equation (1) the two independent variables are r and θ, while in equation (2) the independent variables are h and x. Thus the initial choice of variables in a problem can affect the form of the function that we need to consider, and may well affect the difficulty of our calculations. In this case, it is quite reasonable to represent I as a function of the two variables h and x, so let us suppose that we have made that choice.

Often, we wish to gain some physical insight into a problem, but, as with equation (2), the practical implications of a particular formula are not obvious. However, we can gain some understanding of this situation if we keep the height h fixed, at say $h = 3$, and vary the distance x. The function $I = I(h,x)$ now becomes a function of the single variable x, so we can sketch its graph, as shown in Figure 2. We can now see that the point of brightest illumination is when P is immediately below the lamp (as intuition might lead us to expect).

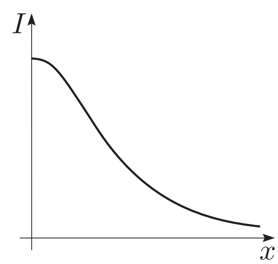


Figure 2 Plot of $I(3,x)$ against x

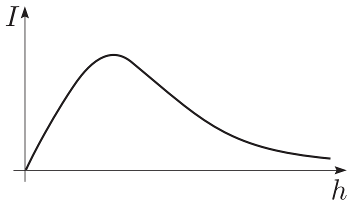


Figure 3 Plot of $I(h, 10)$ against h

Suppose now that we keep the point P fixed, at say $x = 10$, but vary the height of the lamp. Once again, I becomes a function of a single variable, in this case h, and again we can sketch a graph, as shown in Figure 3. This time the result is perhaps a little more unexpected, for our graph tells us that there is an optimum height for the lamp at which the illumination on the horizontal slab at P is greatest. (This is because as h increases, there is a trade-off between the increasing distance of the lamp from P, which tends to decrease the illumination, and the decreasing obliqueness of the angle, which tends to increase the illumination.)

You may well wonder if it is possible to sketch a graph that represents the variation of I when both variables h and x are allowed to change. Indeed it is, but we need more than two dimensions, and the resulting surface can be hard to interpret. (We will look at such surfaces in Section 1.)

A *surface* here is a set of points in space where coordinates satisfy a particular equation.

In order to deal sensibly with functions of two or more variables, we need mathematical tools similar to those that we have at our disposal for functions of a single variable. We should like to be able to differentiate these functions, to locate the points at which they take their greatest (or least) values, and to approximate them by polynomials. This may enable us to understand the behaviour of such functions and so give us an insight into the physical situations from which they arise. In fact, we need to develop a form of calculus that applies to such functions, and that is the purpose of this unit.

Section 1 concentrates on functions of two variables. We show how such functions may be used to define a surface, then extend the notion of derivative in order to investigate the slope of such a surface. Section 2 provides a brief discussion of Taylor polynomials for functions of one variable, then extends the discussion to Taylor polynomials for functions of several variables. Section 3 discusses the main topic of the unit: the classification of stationary points (where the first derivatives are zero) for functions of two or more variables.

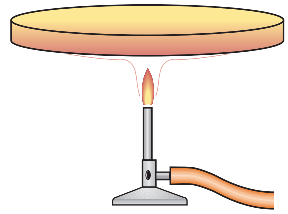
1 Functions of two variables

Our main objective in this section is to extend some of the ideas of the calculus of functions of one variable – most importantly, the concepts of *derivative*, *stationary point*, *local maximum* and *local minimum*. However, before we discuss the *calculus* of functions of two variables, we need to discuss the *concept* of a function of two variables. We introduce the concept by considering a physical situation.

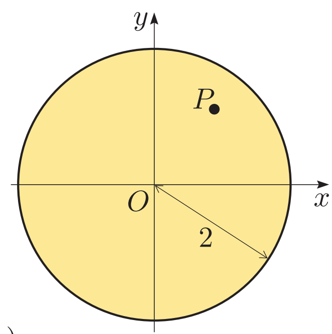
1.1 Introducing functions of two variables

Imagine an experiment in which a thin flat metal disc of radius 2 metres is heated (see Figure 4(a)). At a particular moment, we record the temperatures at various points on the upper surface of the disc.

We could specify the points on the surface of the disc by means of a Cartesian coordinate system, using the centre of the disc as the origin, as shown in Figure 4(b). The temperature may well vary over the surface of the disc, so that the temperature at O is higher than at P, for example. Nevertheless, at any given moment, each point (x, y) of the disc has a well-defined temperature (Θ, say, in degrees Celsius). We could denote this temperature by $\Theta(x, y)$ to remind ourselves that the value of Θ depends on our particular choice of x and y. We say that Θ is a function of the two variables x and y, and that the *dependent variable* Θ is a function of the two *independent variables* x and y.



(a)



(b)

Figure 4 (a) Bunsen burner heating the centre of a metal disc of radius 2 metres. (b) Surface of the disc depicted in the (x, y)-plane.

There is a natural restriction on the independent variables x and y arising from the physical situation: $\Theta(x, y)$ is defined only for points (x, y) lying on the disc. That is, the domain of the function Θ is the set of points (x, y) in the plane such that $x^2 + y^2 \leq 4$.

Suppose that a mathematical model of this situation predicts that

$$\Theta(x, y) = 10(10e^{-(x^2+y^2)} + 1) \quad (\text{for } x^2 + y^2 \leq 4).$$

Then it is a simple matter to calculate the predicted temperature at any point on the disc. For example, at the point $(1, 1.5)$, the predicted temperature (in degrees Celsius) is

$$\Theta(1, 1.5) = 10(10e^{-(1^2+1.5^2)} + 1) = 10(10e^{-3.25} + 1) \simeq 13.9.$$

Exercise 1

Given $f(x, y) = 3x^2 - 2y^3$, evaluate the following.

(a) $f(2, 3)$ (b) $f(3, 2)$ (c) $f(a, b)$ (d) $f(b, a)$

(e) $f(2a, b)$ (f) $f(a - b, 0)$ (g) $f(x, 2)$ (h) $f(y, x)$

Exercise 2

Figure 5 shows a double pendulum consisting of two light model rods OA and AB, each of length a, with a smooth joint at A. The rods move in a vertical plane, with O attached to a fixed point by means of a frictionless hinge. A particle of mass m is attached to B, and the angles θ and ϕ are as shown.

In this system the *potential energy* is given by mg multiplied by the vertical displacement from O (where g is the magnitude of the acceleration due to gravity). (You will learn more about potential energy in Unit 9.) Express the potential energy U of the system in terms of the independent variables θ and ϕ. What is the least possible value of U?

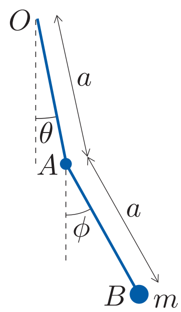


Figure 5 Double pendulum, fixed to point O, consisting of two light rods articulated at A and supporting a particle of mass m at B

1.2 Geometric interpretation

A function of two variables expresses a dependent variable in terms of two independent variables, so there are three varying quantities altogether. Thus a graph of such a function will require three dimensions.

Functions of two variables and surfaces

A **function of two variables** is a function f whose domain is $\mathbb{R}^2$ (or a subset of $\mathbb{R}^2$) and whose codomain is $\mathbb{R}$. Thus for each point (x, y) in the domain of f, there is a unique value z defined by

$$z = f(x, y).$$

The set of all points with coordinates $(x, y, z) = \big(x, y, f(x, y)\big)$, plotted in a three-dimensional Cartesian coordinate system, is the **surface** with equation $z = f(x, y)$.

Each variable is in the set of real numbers $\mathbb{R}$, hence the domain of a function of two variables is denoted by $\mathbb{R}^2$.

The definition of a function of two variables generalises in a straightforward way: a **function of *n* variables** is a function whose domain is $\mathbb{R}^n$ (or a subset of $\mathbb{R}^n$) and whose codomain is $\mathbb{R}$.

The simplest of all surfaces $z = f(x, y)$ for a function of two variables arises from choosing f to be the zero function. We then have the equation $z = 0$, and the surface is the plane that contains the x- and y-axes, known as the (x, y)-plane.

More generally, the surface corresponding to any *linear* function of two variables $z = f(x, y) = Ax + By + C$ (where A, B and C are constants) is a **plane**.

For example, the equation $z = f(x, y) = -\frac{2}{3}x - 2y + 2$ represents a plane passing through the three points $(3, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 2)$, and extending indefinitely. Part of this plane is illustrated in Figure 6.

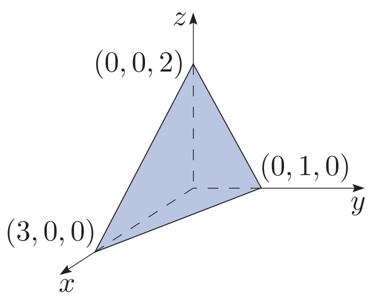


Figure 6 Part of the plane $z = -\frac{2}{3}x - 2y + 2$

A plane can be described geometrically by specifying a point through which it passes, say (a, b, c), and a unit vector that is normal (or, in other words, perpendicular) to the plane, denoted by $\widehat{\mathbf{n}}$, which is unique up to a sign.

Let $\mathbf{a} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$, and consider $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, the position vector of a general point on the plane. The vector $\mathbf{r} - \mathbf{a}$ is in the plane and therefore perpendicular to $\widehat{\mathbf{n}}$, hence the equation for the plane can be expressed in vector form as

$$(\mathbf{r} - \mathbf{a}) \cdot \widehat{\mathbf{n}} = 0. \qquad (3)$$

This construction is illustrated in Figure 7. Note that $|\mathbf{a} \cdot \widehat{\mathbf{n}}|$ is the perpendicular distance – i.e. the shortest distance – to the plane from the origin.

Recall from Unit 4 that the solution of simultaneous linear equations can be interpreted in terms of the intersection of planes.

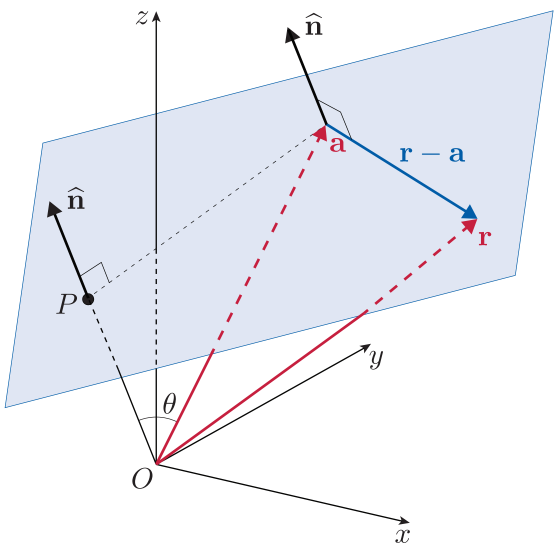


Figure 7 Plane passing through the point represented by vector $\mathbf{a}$, with unit vector $\widehat{\mathbf{n}}$ normal to the plane. P is the point on the plane closest to the origin, at distance $|\mathbf{a}|\cos\theta = \mathbf{a}\cdot\widehat{\mathbf{n}}$.

Example 1

Consider the plane $z = -\frac{2}{3}x - 2y + 2$ discussed above, and the general situation illustrated in Figure 7.

(a) Write down a vector $\mathbf{n}$ that is normal to the plane, and hence calculate $\widehat{\mathbf{n}}$.

(b) By expressing the equation of the plane in terms of $\widehat{\mathbf{n}}$ and the position vector $\mathbf{r}$ of a general point on the plane, show that the shortest distance from the plane to the origin is $6/7$.

Solution

(a) Note that the equation of the plane $z = -\frac{2}{3}x - 2y + 2$ can be written as

$$\tfrac{2}{3}x + 2y + z = 2,$$

or as $\mathbf{n}\cdot\mathbf{r} = 2$, where $\mathbf{r}$ is the position vector and

$$\mathbf{n} = \tfrac{2}{3}\mathbf{i} + 2\mathbf{j} + \mathbf{k},$$

which is normal to the plane, as can be deduced by comparison with equation (3). A unit normal $\widehat{\mathbf{n}}$ is given by $\widehat{\mathbf{n}} = \mathbf{n}/|\mathbf{n}|$, where

$$|\mathbf{n}|^2 = \tfrac{4}{9} + 4 + 1 = 49/9$$

so $|\mathbf{n}| = 7/3$. Thus

$$\widehat{\mathbf{n}} = \tfrac{2}{7}\mathbf{i} + \tfrac{6}{7}\mathbf{j} + \tfrac{3}{7}\mathbf{k}.$$

(b) Multiplying both sides of $\mathbf{n} \cdot \mathbf{r} = 2$ by $1/|\mathbf{n}|$ gives $\widehat{\mathbf{n}} \cdot \mathbf{r} = 6/7$, and comparison with equation (3) shows that $\widehat{\mathbf{n}} \cdot \mathbf{a}$, the shortest distance from the plane to the origin, is $6/7$, where $\mathbf{a}$ is the vector position of any point in the plane.

The surfaces corresponding to the functions $p(x, y) = x^2 + y^2$ (see Figure 8(a)) and $h(x, y) = y^2 - x^2$ (see Figure 8(b)) are not planes, but you will see later that their behaviour near the origin is of particular interest. The function $p(x, y)$ is a **paraboloid** (which can be obtained by plotting the parabola $z = x^2$ in the (x, z)-plane then rotating it about the z-axis). The function $h(x, y)$ is a **hyperboloid** (which cannot be obtained by rotating a curve about the z-axis or any other axis).

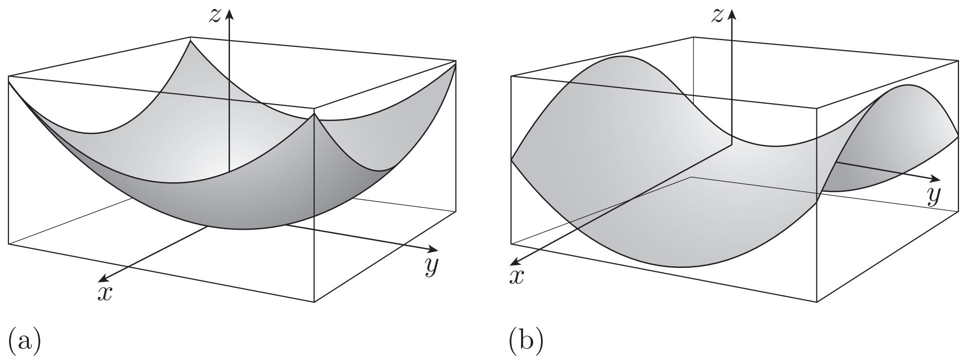


Figure 8 Surfaces of (a) the paraboloid $z = p(x, y) = x^2 + y^2$, and (b) the hyperboloid $z = h(x, y) = y^2 - x^2$

Section functions

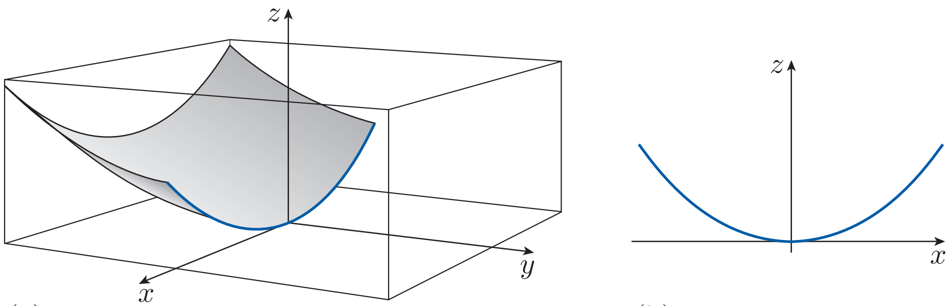
In general, the surface representing a function may be complicated and difficult to visualise. It is often helpful to consider the function obtained by fixing all the independent variables except one at specific values. In the case of the function $p(x, y) = x^2 + y^2$, for example, we might choose to fix the value of y at 2, in which case we are left with the function of a *single variable*

$$p(x, 2) = x^2 + 4.$$

This function is known as a **section function** of $p(x, y) = x^2 + y^2$, with y fixed at 2.

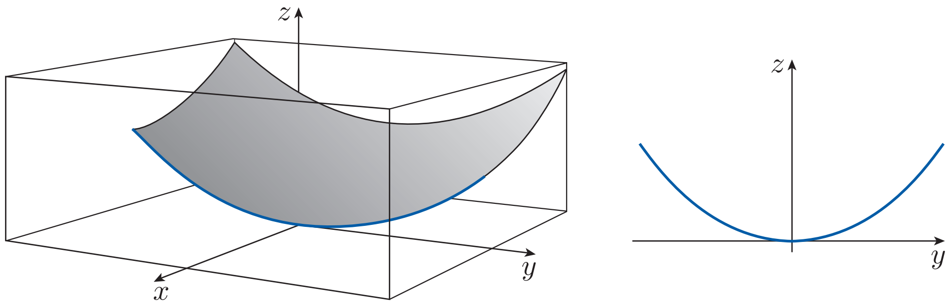
In the case of the surfaces shown in Figure 8, it is worth looking at their behaviour near the origin. The section functions $p(x, 0)$ and $p(0, y)$, and $h(x, 0)$ and $h(0, y)$, are quite illuminating. They show, very clearly, properties of the surfaces that you may have already observed.

The section functions $p(x, 0)$ and $p(0, y)$ are shown in Figures 9 and 10, respectively.



(a) (b)

Figure 9 (a) The surface $z = p(x,y) = x^2 + y^2$ cut along the plane $y = 0$. (b) The section function $z = p(x,0)$.

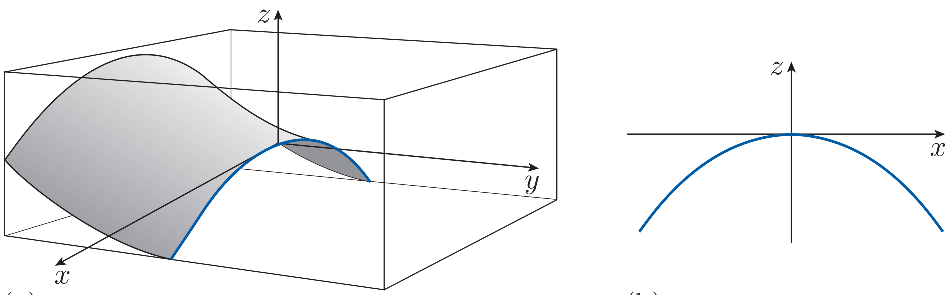


(a) (b)

Figure 10 (a) The surface $z = p(x,y) = x^2 + y^2$ cut along the plane $x = 0$. (b) The section function $z = p(0,y)$.

The surface $z = p(x,y) = x^2 + y^2$ has a local minimum at the origin; correspondingly, each of the section functions $p(x,0)$ and $p(0,y)$ has a local minimum there.

The section functions $h(x,0)$ and $h(0,y)$ are shown in Figures 11 and 12, respectively.



(a) (b)

Figure 11 (a) The surface $z = h(x,y) = y^2 - x^2$ cut along the plane $y = 0$. (b) The section function $z = h(x,0)$.

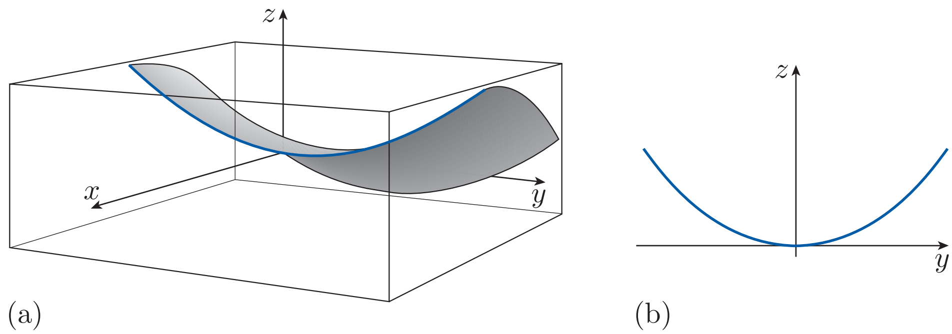


Figure 12 (a) The surface $z = h(x, y) = y^2 - x^2$, cut along the plane $x = 0$. (b) The section function $z = h(0, y)$.

The surface $z = h(x, y) = y^2 - x^2$ has neither a local maximum nor a local minimum at the origin, corresponding to the fact that the section function $h(x, 0)$ has a local maximum at the origin, while the section function $h(0, y)$ has a local minimum there.

Of course, we have not yet defined what we mean by a (local) maximum or minimum of a function of two variables (that comes in Section 3), but it is already clear that a surface can behave in a more complicated fashion than the graph of a function of one variable.

It is worth spending a little time at this point to recall how to use calculus to justify that the section functions $h(x, 0)$ and $h(0, y)$ have a local maximum and a local minimum, respectively, at the origin.

A standard method in this context is the *second derivative test*, which applies to 'sufficiently smooth' functions of one variable. (In this context, 'sufficiently smooth' means that the first and second derivatives exist.) If $f(x)$ is a function whose first and second derivatives exist, and at some point $x = a$ we have $f'(a) = 0$, then a is a **stationary point** of $f(x)$. Often (but not always) it will be a local maximum or a local minimum.

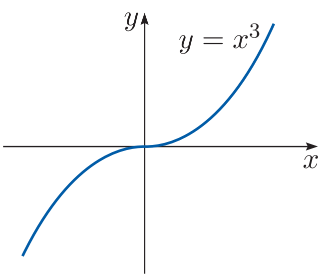


Figure 13 The curve $y = f(x) = x^3$, with a point of inflection at the origin

The **second derivative test** is based on the evaluation of $f''(x)$ at a. There are three possibilities:

- If $f''(a)$ is *negative*, then f has a *local maximum* at a.
- If $f''(a)$ is *positive*, then f has a *local minimum* at a.
- If $f''(a) = 0$, then the test is inconclusive. There may still be a local maximum or a local minimum at a, but another possibility is that there is a *point of inflection*, such as occurs at $x = 0$ for the function $f(x) = x^3$ (see Figure 13).

Example 2

For $h(x, y) = y^2 - x^2$, use the second derivative test to verify that the section function $h(x, 0)$ has a local maximum at $x = 0$ and that the section function $h(0, y)$ has a local minimum at $y = 0$.

Solution

We have $h(x,0) = -x^2$, and

$$\frac{d}{dx}(-x^2) = -2x,$$

which is zero when $x = 0$. Thus $h(x,0)$ has a stationary point at $x = 0$. Now

$$\frac{d^2}{dx^2}(-x^2) = -2,$$

which is negative for all values of x, so this stationary point is a local maximum.

We have $h(0,y) = y^2$, and

$$\frac{d}{dy}(y^2) = 2y,$$

which is zero when $y = 0$. Thus $h(0,y)$ has a stationary point at $y = 0$. Now

$$\frac{d^2}{dy^2}(y^2) = 2,$$

which is positive for all values of y, so this stationary point is a local minimum.

Exercise 3

Show that the section function of $F(x,y) = 100e^{-(x^2+y^2)}$ with y fixed at 0 has a local maximum at $x = 0$.

The concept of a section function can be extended easily to functions of more than two variables. For example, the section function of

$$w(x,y,t) = \frac{x}{y+2t}$$

The domain of w should be restricted so that $y + 2t$ can never be zero.

with x fixed at 3 and t fixed at 1 is

$$w(3,y,1) = \frac{3}{y+2},$$

which is a function of y only. (Section functions are always functions of *one* variable, so we have to fix the values of all the variables except one to obtain a section function.) One obvious advantage of considering section functions (rather than the original function of two or more variables) is that we can apply the familiar calculus techniques for functions of one variable.

Contour maps

Another way to visualise a function of two variables on a plane is to use contour maps. Consider a continuous function $f(x,y)$. A curve in the (x,y)-plane defined by the equation $f(x,y)=c$, for some constant c, is called a **contour curve** of $f(x,y)$. By plotting contour curves of $f(x,y)$ for different values of c chosen from a discrete set of constants, $c_1, c_2, \ldots$ say, we obtain a family of contour curves in the (x,y)-plane. Such a plot is called a **contour map**. It is reminiscent of the contours found on Ordnance Survey maps, which are curves joining points on land that are at a constant elevation above sea level.

As an example, consider contour maps for the paraboloid $p(x,y)=x^2+y^2$ (Figure 14(a)) and the hyperboloid $h(x,y)=y^2-x^2$ (Figure 14(b)), seen earlier.

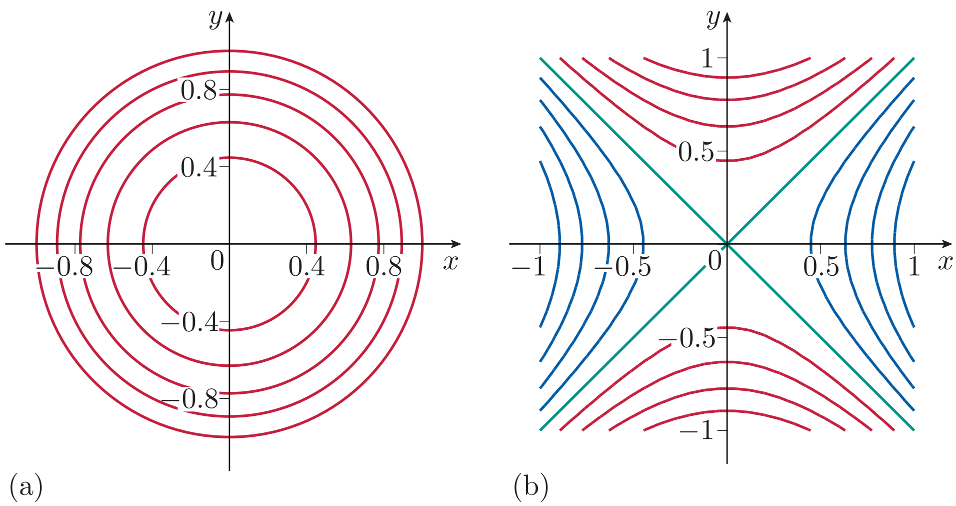


Figure 14 (a) Contours of the paraboloid $p(x,y)=x^2+y^2$, given by red circles with radii increasing with increasing values of $c=p(x,y)$, where $c>0$. If $c=0$, the contour is just a single point. Clearly, there are no contours with $c<0$. (b) Contours of the hyperboloid $h(x,y)=y^2-x^2$, given by $h(x,y)=c$; contours with $c>0$ are depicted in red, those with $c<0$ in blue, and those with $c=0$ in green.

Exercise 4

Sketch a contour map for the function $f(x,y)=x^2+y$, considering contours with positive, negative and zero values.

1.3 First-order partial derivatives

As a first step towards our goal of extending the ideas of calculus to functions of two (or more) variables, we begin by recalling the role played by the tangent to a graph in the calculus of functions of one variable. If we imagine a tangent line sliding along a graph, then each time the line is horizontal, we have a stationary point. We apply the same idea to functions of two variables, only this time we slide a *tangent plane* over the surface. Let us be a little more precise about this.

In many cases, the tangent line to a curve C at a point P on C is the straight line that *touches* C at P, but does not *cross* the curve at that point. Sometimes, however, even if C is a smooth curve as it goes through P, it is not possible to find a line with the 'non-crossing' property. Suppose, for example, that C is the graph of $y = x^3$ (see Figure 15) and P is the origin. Then every straight line through the origin 'crosses' C, as you may verify by placing a ruler at various angles through the origin on Figure 15. Nevertheless, the line that is the x-axis seems to be a good candidate for the 'tangent line' to C at $(0, 0)$; intuitively, it seems to pass through the curve at a 'zero angle'. Let us try to make this idea more mathematically robust.

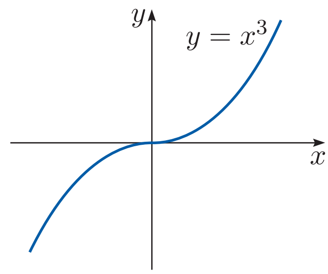


Figure 15 The curve $y = f(x) = x^3$

Consider again a general curve C and a point P on C at which we wish to find a tangent line (see Figure 16). If Q is a point on the curve close to P, then the **chord** through P and Q is the straight line through these points. If C is smooth enough to have a derivative at P, then as Q approaches P, the chord through P and Q approaches a well-defined line through P, which is defined to be the **tangent line** to C at P.

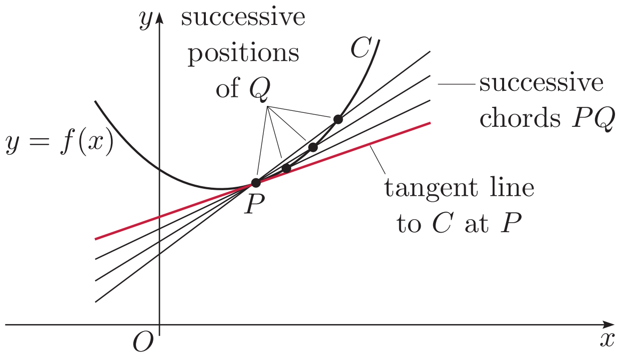


Figure 16 Chords PQ of a general curve C; as point Q approaches point P, the chord PQ tends to the tangent to C at P

The tangent line at P is the line that has the same *slope* as C at that point. To return to the case of the curve $y = f(x) = x^3$ (see Figure 15), since $f'(0) = 0$, the slope of C at $(0, 0)$ is 0, so the tangent line to C at $(0, 0)$ is the x-axis, that is, the line $y = 0$.

More generally, if the curve C is defined by the function $y = f(x)$, then at the point $P = (a, f(a)) = (a, b)$, the slope of C is $f'(a)$ and the equation of the tangent line is $y - b = f'(a)(x - a)$, this being the equation of a line through (a, b) with slope $f'(a)$.

In a similar way, a smooth surface without breaks or folds has a *tangent plane*. In the cases of the paraboloid and hyperboloid in Figure 8, the surface is horizontal at the point $(0,0,0)$, so for each of these surfaces, the tangent plane at $(0,0,0)$ is the (x,y)-plane, that is, $z=0$. In the case of the paraboloid, the (x,y)-plane touches the surface at that point, but does not cut through it. The hyperboloid, however, lies partly above and partly below the (x,y)-plane, so (as with the tangent line at $(0,0)$ to $y=x^3$) the (x,y)-plane cuts the hyperboloid – despite having the same 'slope in any direction' as the hyperboloid at $(0,0,0)$.

The tangent plane can be defined by a 'limiting' construction similar to that for the tangent line. We need three points to define a plane, so if P is a point on a smooth surface S, we must take two further points Q and R on the surface, and we must ensure that P, Q and R never lie on a straight line (when projected onto the (x,y)-plane). Then as Q and R separately approach P (from distinct directions), the plane through P, Q and R will approach a well-defined plane, the **tangent plane** at P (see Figure 17).

There are some technicalities concerning the choice of Q and R, but they are not important in this context.

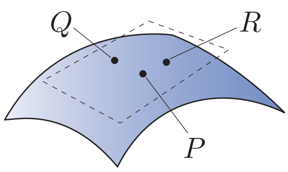


Figure 17 Plane passing through points P, Q and R on a curved surface; as Q and R approach P from different directions, the plane becomes tangent to the surface at P

Just as the slope of a curve C at a point P on C is equal to the slope of the tangent line at P, so the 'slope' of the surface S at the point P on S is equal to the 'slope' of the tangent plane at P. But what *is* this slope?

Imagine that the surface is a hillside and that you are walking across it (see Figure 18). Let us suppose that you want to measure the 'slope' of the hill at some particular point. You immediately encounter a problem: in which direction should you measure the slope? If you choose a direction pointing 'straight up the hill' you will get one value, and if you choose to move 'round the hill' you will get another. We will choose two specific directions in which to measure the slope: the x-direction and the y-direction. On a smooth hill, this will be sufficient to determine the slope in every direction (as you will see in Subsection 1.4). So we start by examining the rate at which a function of two variables changes when we keep one of the variables fixed.

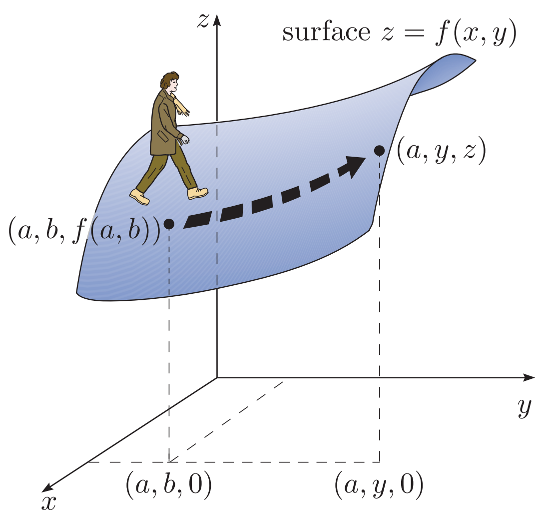


Figure 18 Path travelled on a surface defined by the equation $z=f(x,y)$

Suppose that you are walking over the surface defined by the equation $z = f(x, y)$ and that you are at the point $(a, b, f(a, b)) = (a, b, c)$. If $c > 0$, then directly below you is the point $(a, b, 0)$ lying in the (x, y)-plane (see Figure 18). Now you begin to move across the surface, taking care to ensure that the value of x stays fixed at a. As you move, the value of y changes, and your height z above the (x, y)-plane varies. We are going to investigate the rate of change of the height z with y, which you would recognise as the slope (at some particular point) of the path along which you are walking. We refer to this slope as the *slope of the surface in the y-direction* (at the point in question).

As a specific example, consider the function

$$f(x, y) = x^2 - y^3 + 2xy^2.$$

Suppose that we are trying to find the slope of this surface in the y-direction at the point $(2, 1, 7)$ on the surface. First, we construct the section function of $f(x, y)$ with x fixed at 2:

$$f(2, y) = 4 - y^3 + 4y^2.$$

Then we differentiate this function with respect to y, to obtain

$$\frac{df}{dy}(2, y) = -3y^2 + 8y.$$

Finally, we put $y = 1$ into this expression and obtain the value 5 for the slope of the surface in the y-direction at $(2, 1, 7)$.

This process would be the same for any fixed value of x thus can be generalised: fix x at some constant value, then differentiate $f(x, y)$ with respect to y, using all the standard rules of differentiation and *treating x as a constant*. The result is the expression $-3y^2 + 4xy$. This is the *partial derivative* of the function $f(x, y) = x^2 - y^3 + 2xy^2$ with respect to y, denoted by $\partial f/\partial y$. So we have

The use of the 'partial dee' symbol ∂ (rather than d) for partial derivatives distinguishes them from ordinary derivatives. It is important that there is a clear distinction between these symbols in written work.

$$\frac{\partial f}{\partial y}(x, y) = -3y^2 + 4xy.$$

If we put $x = 2$ and $y = 1$ in this expression, then we obtain

$$\frac{\partial f}{\partial y}(2, 1) = 5,$$

which is the same value as we found above.

The same method can be used to find the slope in the x-direction, but this time we keep y fixed and differentiate with respect to x. We obtain

$$\frac{\partial f}{\partial x}(x, y) = 2x + 2y^2.$$

Putting $x = 2$ and $y = 1$ into this expression, we obtain the value 6 for the slope of the surface in the x-direction at $(2, 1, 7)$.

More formally, the partial derivatives of a function $f(x,y)$ with respect to x and y are given by

The ordinary derivative of $f(t)$ with respect to t is formally defined as
$$\frac{df}{dt} = \lim_{h\to 0} \frac{f(t+h) - f(t)}{h}.$$

$$\frac{\partial f}{\partial x} = \lim_{\delta x\to 0} \frac{f(x+\delta x, y) - f(x,y)}{\delta x} \tag{4}$$

and

$$\frac{\partial f}{\partial y} = \lim_{\delta y\to 0} \frac{f(x, y+\delta y) - f(x,y)}{\delta y}, \tag{5}$$

where δx and δy denote (small) *increments* (i.e. changes) in the values of x and y, respectively.

Note the important difference between the symbols:

the 'delta' symbol	δ	represents an increment;
the 'partial dee' symbol	∂	represents partial differentiation.

We will need the more formal definitions later, but for now the following definitions will suffice.

The expression $\partial f/\partial x$ is read as 'partial dee f by dee x'.

First-order partial derivatives

Given a function f of two variables x and y, the **partial derivative** $\partial f/\partial x$ is obtained by differentiating $f(x,y)$ with respect to x while treating y as a constant. Similarly, the partial derivative $\partial f/\partial y$ is obtained by differentiating $f(x,y)$ with respect to y while treating x as a constant.

The partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ represent the slopes of the surface $z = f(x,y)$ at the point $(x, y, f(x,y))$ in the x- and y-directions, respectively. These partial derivatives are the **first-order partial derivatives** or the **first partial derivatives** of the function f.

We will define second-order and higher-order partial derivatives in Section 2.

We will also use the alternative notation f_x for $\partial f/\partial x$ and f_y for $\partial f/\partial y$.

Example 3

Calculate $\partial f/\partial x$ and $\partial f/\partial y$ for

$$f(x,y) = \sqrt{xy} + xy^2 \quad (x > 0,\ y > 0).$$

Find the slopes of the corresponding surface $z = \sqrt{xy} + xy^2$ in the x- and y-directions at the point $(4, 1, 6)$ on the surface.

Solution

First, we treat y as a constant and differentiate with respect to x (remembering that $\sqrt{xy} = \sqrt{x}\sqrt{y}$), to obtain

$$\frac{\partial f}{\partial x} = \frac{\sqrt{y}}{2\sqrt{x}} + y^2.$$

So the slope in the x-direction at $(4,1,6)$ is

$$\frac{\sqrt{1}}{2\sqrt{4}}+1=\tfrac{5}{4}.$$

Now we treat x as a constant and differentiate f with respect to y, to obtain

$$\frac{\partial f}{\partial y}=\frac{\sqrt{x}}{2\sqrt{y}}+2xy.$$

So the slope in the y-direction at $(4,1,6)$ is

$$\frac{\sqrt{4}}{2\sqrt{1}}+(2\times 4\times 1)=9.$$

In general, $\partial f/\partial x$ and $\partial f/\partial y$ are functions of the two variables x and y, so in Example 3 we could write

$$\frac{\partial f}{\partial x}(x,y)=\frac{\sqrt{y}}{2\sqrt{x}}+y^2,\quad \frac{\partial f}{\partial y}(x,y)=\frac{\sqrt{x}}{2\sqrt{y}}+2xy,$$

or alternatively,

$$f_x(x,y)=\frac{\sqrt{y}}{2\sqrt{x}}+y^2,\quad f_y(x,y)=\frac{\sqrt{x}}{2\sqrt{y}}+2xy,$$

if we wish to emphasise the fact that the partial derivatives are themselves functions. Thus we could write

$$f_x(4,1)=\tfrac{5}{4},\quad f_y(4,1)=9,$$

for the slopes in the x- and y-directions, respectively, at the point in question.

Exercise 5

Given

$$f(x,y)=x^3\cos y+y^2\sin x,$$

calculate $\partial f/\partial x$ and $\partial f/\partial y$.

Exercise 6

Given

$$f(x,y)=(x^2+y^3)\sin(xy),$$

calculate $\partial f/\partial x$ and $\partial f/\partial y$.

Of course, the two independent variables need not be denoted by x and y; any variable names will do, as the next example shows.

Example 4

Given $f(\alpha,t)=\alpha\sin(\alpha t)$, calculate $f_\alpha\left(\frac{\pi}{2},1\right)$ and $f_t\left(\frac{\pi}{2},1\right)$.

Solution

Differentiating partially with respect to α gives $f_\alpha = \sin(\alpha t) + \alpha t\cos(\alpha t)$, so

$$f_\alpha\left(\tfrac{\pi}{2}, 1\right) = \sin\tfrac{\pi}{2} + \tfrac{\pi}{2}\cos\tfrac{\pi}{2} = 1.$$

Differentiating partially with respect to t gives $f_t = \alpha^2\cos(\alpha t)$, so

$$f_t\left(\tfrac{\pi}{2}, 1\right) = \left(\tfrac{\pi}{2}\right)^2\cos\tfrac{\pi}{2} = 0.$$

Exercise 7

Given $u(\theta, \phi) = \sin\theta + \phi\tan\theta$, calculate u_θ and u_ϕ. Hence evaluate $u_\theta\left(\frac{\pi}{4}, \frac{\pi}{2}\right)$ and $u_\phi\left(\frac{\pi}{4}, \frac{\pi}{2}\right)$.

Often, a mathematical model will generate a relationship between variables. For example, we have the formula $V = \frac{1}{3}\pi r^2 h$ for the volume V of a cone in terms of the radius r of its circular base and its height h. We could introduce a function $f(r, h) = \frac{1}{3}\pi r^2 h$ and write the partial derivatives as $\partial f/\partial r$ and $\partial f/\partial h$. But it is often more convenient to let V denote both the variable and the function that defines it, enabling us to write $\partial V/\partial r$ in place of $\partial f/\partial r$, and $\partial V/\partial h$ in place of $\partial f/\partial h$, thus keeping the number of symbols to a minimum.

The notion of partial derivative can be extended to functions of more than two variables. For example, for a function $f(x, y, t)$ of three variables, to calculate $\partial f/\partial x$, we keep y and t fixed, and differentiate with respect to x (and similarly for the other partial derivatives).

Exercise 8

(a) Find all the first partial derivatives of the function

$$f(x, y, t) = x^2y^3t^4 + 2xy + 4t^2x^2 + y.$$

(b) Given $z = (1+x)^2 + (1+y)^3$, calculate $\partial z/\partial x$ and $\partial z/\partial y$. Sketch the section functions $z(x, 0)$ and $z(0, y)$. What is the relevance of the value of $(\partial z/\partial x)(0, 0)$ to the graph of $z(x, 0)$, and what is the relevance of the value of $(\partial z/\partial y)(0, 0)$ to the graph of $z(0, y)$?

1.4 Slope in an arbitrary direction

Suppose that the walker in Figure 18 is not walking in either the x-direction or the y-direction, but in some other direction. Can we use the partial derivatives $(\partial f/\partial x)(a, b)$ and $(\partial f/\partial y)(a, b)$ to find the slope at the point (a, b, c) in this other direction?

To answer this question, we start by looking at the formal definitions of partial derivatives in equations (4) and (5), and use them to investigate what happens when we move a small distance in a particular direction.

For a function $z = f(x, y)$, these equations are

$$\frac{\partial f}{\partial x} = \lim_{\delta x \to 0} \frac{f(x + \delta x, y) - f(x, y)}{\delta x},$$
$$\frac{\partial f}{\partial y} = \lim_{\delta y \to 0} \frac{f(x, y + \delta y) - f(x, y)}{\delta y}.$$

Small increments

If we write $\delta z_1 = f(x + \delta x, y) - f(x, y)$ and $\delta z_2 = f(x, y + \delta y) - f(x, y)$, then δz_1 is the change, or *increment*, in the value of z corresponding to a (small) increment δx in the value of x. Similarly, δz_2 is the increment in the value of z corresponding to a (small) increment δy in the value of y. So we have

$$\frac{\partial f}{\partial x} = \lim_{\delta x \to 0} \frac{\delta z_1}{\delta x}, \quad \frac{\partial f}{\partial y} = \lim_{\delta y \to 0} \frac{\delta z_2}{\delta y}. \tag{6}$$

The expressions of the form $\dfrac{\delta *}{\delta t}$ here are quotients, *not* derivatives. Do not confuse the symbol δ with ∂ or d.

We know from Subsection 1.3 that $\partial f/\partial x$ and $\partial f/\partial y$ are the slopes of $z = f(x, y)$ in the x- and y-directions, respectively. To find the slope of $z = f(x, y)$ in an arbitrary direction, we need to find the increment δz in the value of f when we move a short distance in that direction. Such a movement can be achieved by small increments δx and δy, and equations (6) relate these approximately to the corresponding increments δz_1 and δz_2, as shown below.

Figure 19 shows a small part of the surface $z = f(x, y)$ and the point (a, b) in the (x, y)-plane. We are interested in the effect on z when x and y are increased by the small amounts δx and δy, respectively. The increment δz is produced by moving first from (a, b) to $(a + \delta x, b)$, and then from $(a + \delta x, b)$ to $(a + \delta x, b + \delta y)$. We specify δz_1 to be the increment in z on moving from (a, b) to $(a + \delta x, b)$, and δz_2 to be the increment in z on moving from $(a + \delta x, b)$ to $(a + \delta x, b + \delta y)$.

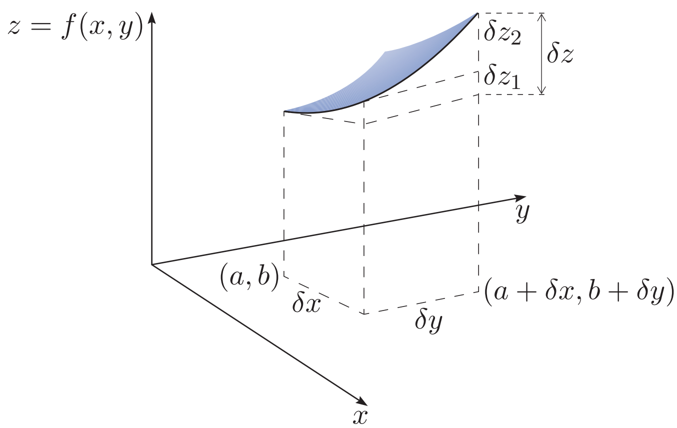


Figure 19 Small part of the surface $z = f(x, y)$ above the point $(x, y) = (a, b)$

We consider δz_1 first. This is $f(a+\delta x, b) - f(a,b)$, the difference between the function values at (a,b) and $(a+\delta x, b)$. Since x has moved from a to $a+\delta x$ while y has remained constant, the increment in z is approximately equal to the increment in x multiplied by the partial derivative with respect to x, that is,

The approximation used here holds because for small δx, we can take the slope $\partial f/\partial x$ in the x-direction to be almost constant on the interval between $x=a$ and $x=a+\delta x$. Similarly, for small δy, the slope $\partial f/\partial y$ in the y-direction can be taken to be almost constant on the interval between $y=b$ and $y=b+\delta y$.

$$\delta z_1 \simeq \frac{\partial f}{\partial x}(a,b)\,\delta x. \tag{7}$$

This follows from the first of equations (6).

The increment δz_2 is obtained by holding x constant at the value $a+\delta x$ while incrementing y from b to $b+\delta y$. It is thus approximately equal to the increment in y multiplied by the partial derivative with respect to y, that is,

$$\delta z_2 \simeq \frac{\partial f}{\partial y}(a+\delta x, b)\,\delta y. \tag{8}$$

At this point in the argument, it is necessary to assume that f is a sufficiently smooth function so that the methods of calculus that we require can be applied. We assume, in particular, that the partial derivatives are continuous. So for small δx we have

$$\frac{\partial f}{\partial y}(a+\delta x, b) \simeq \frac{\partial f}{\partial y}(a,b). \tag{9}$$

Putting equations (7), (8) and (9) together, we now have our expression linking small increments in z with small increments in x and y:

$$\delta z = \delta z_1 + \delta z_2 \simeq \frac{\partial f}{\partial x}(a,b)\,\delta x + \frac{\partial f}{\partial y}(a,b)\,\delta y, \tag{10}$$

which may be easier to remember when written in the form

$$\delta z \simeq \frac{\partial z}{\partial x}\,\delta x + \frac{\partial z}{\partial y}\,\delta y. \tag{11}$$

The approximation (11) has an important application to error analysis, as the next example shows.

Example 5

The volume V of a cone of height h with base radius r is given by

$$V = \tfrac{1}{3}\pi r^2 h.$$

(a) Determine the approximate change in the volume if the radius increases from 2 to $2+\delta r$ and the height increases from 5 to $5+\delta h$.

(b) If the radius and height measurements are each subject to an error of magnitude up to 0.01, how accurate is the estimate

$$V \simeq \tfrac{1}{3} \times \pi \times 2^2 \times 5 = \tfrac{20}{3}\pi \ (= 20.94,\ \text{to 4 s.f.})$$

of the volume?

Solution

(a) Calculating the partial derivatives of $V = V(r,h)$ with respect to r and h, we have

$$\frac{\partial V}{\partial r} = \tfrac{2}{3}\pi r h, \quad \frac{\partial V}{\partial h} = \tfrac{1}{3}\pi r^2.$$

Setting $r = 2$ and $h = 5$, we calculate

$$\frac{\partial V}{\partial r}(2,5) = \tfrac{2}{3}\pi \times 2 \times 5 = \tfrac{20}{3}\pi, \quad \frac{\partial V}{\partial h}(2,5) = \tfrac{1}{3}\pi \times 2^2 = \tfrac{4}{3}\pi,$$

thus, from approximation (11), the approximate change in volume is

$$\delta V \simeq \tfrac{20}{3}\pi\,\delta r + \tfrac{4}{3}\pi\,\delta h.$$

(b) If the maximum possible magnitudes of δr and δh are 0.01, then the maximum possible magnitude of δV is approximately

$$(\tfrac{20}{3} + \tfrac{4}{3})\pi \times 0.01 = 8\pi \times 0.01 \simeq 0.25.$$

Thus the volume estimate of 20.94 may compare with an actual value as high as 21.19 or as low as 20.69. The estimate is accurate to only two significant figures (and should be given as $V = 21$, correct to two significant figures).

Exercise 9

Given $z = (1+x)^2 + (1+y)^3$, find the approximate increment δz in z when x is incremented from 0 to δx, and y is incremented from 2 to $2 + \delta y$.

(*Hint*: You may wish to use your solution to Exercise 8(b).)

Rate of change along a curve

Now we focus on the rate of change of $z = f(x,y)$ if (x,y) is constrained to move along a curve in the (x,y)-plane. Let us suppose that x and y are themselves functions of a parameter t, so that as t varies, the point $(x(t), y(t))$ moves along a curve in the (x,y)-plane, passing through (a,b) when $t = t_0$ and through $(a+\delta x, b+\delta y)$ when $t = t_0 + \delta t$. From equation (11) we have

You may recall from previous studies that a curve described in terms of a parameter in this way is called a *parametrised curve*.

$$\frac{\delta z}{\delta t} \simeq \frac{\partial z}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial z}{\partial y}\frac{\delta y}{\delta t}. \tag{12}$$

Having introduced the parameter t, we can think of x, y and z as functions of the single variable t. Thinking of them in this way, we have

$$\delta x = x(t_0 + \delta t) - x(t_0),$$
$$\delta y = y(t_0 + \delta t) - y(t_0),$$
$$\delta z = z(t_0 + \delta t) - z(t_0).$$

Thus

$$\frac{\delta x}{\delta t} = \frac{x(t_0 + \delta t) - x(t_0)}{\delta t},$$
$$\frac{\delta y}{\delta t} = \frac{y(t_0 + \delta t) - y(t_0)}{\delta t},$$
$$\frac{\delta z}{\delta t} = \frac{z(t_0 + \delta t) - z(t_0)}{\delta t}.$$

So from the definition of an ordinary derivative, as $\delta t \to 0$, these become the derivatives (not fractions) dx/dt, dy/dt and dz/dt, respectively. Hence as $\delta t \to 0$, approximation (12) becomes

On the curve, z, x and y are functions of t only, so it is consistent to write, for example, dz/dt rather than $\partial z/\partial t$.

$$\frac{dz}{dt}(t_0) = \frac{\partial z}{\partial x}(a,b)\,\frac{dx}{dt}(t_0) + \frac{\partial z}{\partial y}(a,b)\,\frac{dy}{dt}(t_0).$$

This may remind you of the chain rule for the derivative of z with respect to t if z is a function of the single variable x, and x is a function of t; that is, if $z = z(x(t))$, then

$$\frac{dz}{dt} = \frac{dz}{dx}\,\frac{dx}{dt}.$$

Indeed, the formula for the rate of change along a parametrised curve is the two-dimensional analogue of the chain rule of ordinary differentiation.

Chain rule

Let $z = f(x,y)$ be a function whose first partial derivatives exist and are continuous. Then the rate of change of z with respect to t along a curve parametrised by $(x(t), y(t))$ is given by

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\,\frac{dx}{dt} + \frac{\partial z}{\partial y}\,\frac{dy}{dt}. \tag{13}$$

Example 6

Given $z = \sin x - 3\cos y$, find the rate of change of z along the curve $(x(t), y(t))$, where $x(t) = t^2$ and $y(t) = 2t$.

Solution

We have

$$\frac{\partial z}{\partial x} = \cos x, \quad \frac{\partial z}{\partial y} = 3\sin y, \quad \frac{dx}{dt} = 2t, \quad \frac{dy}{dt} = 2.$$

Thus

$$\begin{aligned}\frac{dz}{dt} = \frac{\partial z}{\partial x}\,\frac{dx}{dt} + \frac{\partial z}{\partial y}\,\frac{dy}{dt} &= 2t\cos x + 6\sin y \\ &= 2t\cos(t^2) + 6\sin(2t).\end{aligned}$$

(Note that the final answer is expressed in terms of t.)

Exercise 10

Given $z = y \sin x$, find the rate of change of z along the curve $(x(t), y(t))$, where $x = e^t$ and $y = t^2$. Evaluate this rate of change at $t = 0$.

The above form of the chain rule is easy to remember, but if there is a reason to emphasise that the partial derivatives are evaluated at a particular point, then it may be convenient to write it as

$$\frac{dz}{dt}(t_0) = \frac{\partial f}{\partial x}(a,b) \times \frac{dx}{dt}(t_0) + \frac{\partial f}{\partial y}(a,b) \times \frac{dy}{dt}(t_0), \tag{14}$$

where $x(t_0) = a$ and $y(t_0) = b$.

Writing $z(t)$ is a slight abuse of notation since it suggests that z is a function of just one variable, t (which is true only along the parametrised curve), whereas in general z depends on x and y through $z = f(x,y)$. Rather, $z(t)$ is shorthand for $z(x(t), y(t))$.

The chain rule is of fundamental importance, and we will refer back to it both in this unit and in the remainder of the module. For the moment, we use it to continue our discussion of the slope of a surface.

Slope in an arbitrary direction

We want to find the slope of the surface $z = f(x,y)$ at a point $(a,b,c) = (a,b,f(a,b))$ on the surface in an arbitrary given direction. This given direction is specified by a straight line through (a,b) in the (x,y)-plane. Suppose that on this line x and y are functions of the parameter t, so that $x = x(t)$ and $y = y(t)$. Let us also take $x(0) = a$ and $y(0) = b$. Then our straight line can be parametrised as

$$x(t) = a + t\cos\alpha, \quad y(t) = b + t\sin\alpha,$$

The parametric form of a straight line is discussed in Subsection 2.1 of Unit 2.

where α is the anticlockwise angle that the line makes with the positive x-axis, and t measures the distance along this line, as illustrated in Figure 20. As (x,y) moves along this line, the point $(x,y,f(x,y))$ moves along a curve in the surface $z = f(x,y)$ (see Figure 21).

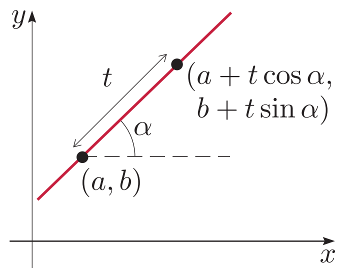


Figure 20 Parametrisation of a straight line passing through the point (a,b) with angle α relative to the x-axis

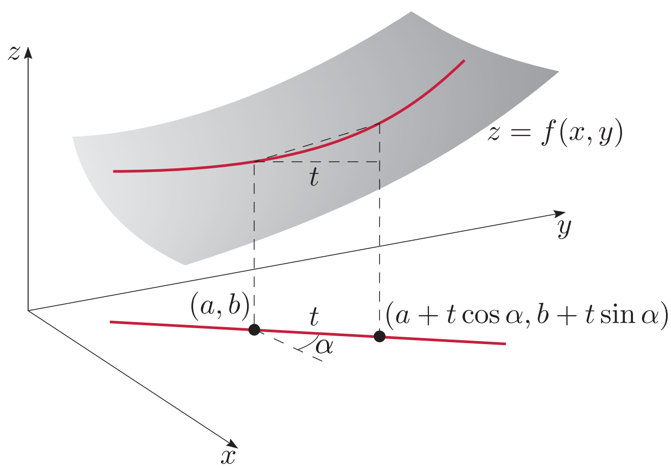


Figure 21 The line in the (x,y)-plane, as plotted in Figure 20, is projected onto the surface $z = f(x,y)$

The advantage of defining the line in terms of the parameter t is that the derivative of z with respect to t is the quantity that we are looking for – the slope of the surface $z = f(x, y)$ at the point (a, b, c), in the direction that makes an angle α with the direction of the x-axis (see Figure 21).

Moreover, we can now use the chain rule given in equation (13). Since $dx/dt = \cos\alpha$ and $dy/dt = \sin\alpha$, we obtain

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\cos\alpha + \frac{\partial z}{\partial y}\sin\alpha. \tag{15}$$

That is, we have shown that the slope of the surface $z = f(x, y)$ at the point $(a, b, f(a, b))$ in the direction making an anticlockwise angle α with the positive x-axis is

$$\frac{\partial f}{\partial x}(a, b)\cos\alpha + \frac{\partial f}{\partial y}(a, b)\sin\alpha$$

or

$$f_x(a, b)\cos\alpha + f_y(a, b)\sin\alpha.$$

This defines the slope of a surface *in a particular direction*. But what is *the* slope of the surface at $(a, b, f(a, b))$? The answer is that it can be thought of as a vector, with a component $f_x(a, b)$ in the x-direction and a component $f_y(a, b)$ in the y-direction, that is, as the vector $f_x(a, b)\mathbf{i} + f_y(a, b)\mathbf{j}$. The advantage of this formulation is that the slope in the direction of an arbitrary unit vector $\widehat{\mathbf{d}} = (\cos\alpha)\mathbf{i} + (\sin\alpha)\mathbf{j}$ in the (x, y)-plane is the dot product of the vectors $f_x(a, b)\mathbf{i} + f_y(a, b)\mathbf{j}$ and $\widehat{\mathbf{d}}$. This idea is important enough to deserve some terminology of its own.

$|\widehat{\mathbf{d}}|^2 = \cos^2\alpha + \sin^2\alpha = 1.$

Gradient

The vector

$$\mathbf{grad}\, f(a, b) = f_x(a, b)\mathbf{i} + f_y(a, b)\mathbf{j} \tag{16}$$

is called the **gradient** of the function $f(x, y)$ at the point (a, b), and is alternatively denoted by $\boldsymbol{\nabla} f(a, b)$.

Thus $\boldsymbol{\nabla} f(x, y)$ is a (vector) function of two variables, called the **gradient function**.

The symbol $\boldsymbol{\nabla}$ is called 'del', or sometimes 'nabla' (after a type of Phoenician harp said to be shaped like an inverted Delta). Note that $\boldsymbol{\nabla} f$ is a vector function, so in written work, the symbol must be underlined (note how $\boldsymbol{\nabla}$ has been typeset in bold).

The result concerning the slope of a surface can now be written as follows.

Slope of a surface

Let $z = f(x, y)$ be a surface described by a function whose first partial derivatives exist and are continuous. Then the slope of the surface at the point $(a, b, f(a, b))$ in the direction of the unit vector $\widehat{\mathbf{d}} = (\cos\alpha)\mathbf{i} + (\sin\alpha)\mathbf{j}$ in the (x, y)-plane is the dot product

$$(\boldsymbol{\nabla} f(a, b)) \cdot \widehat{\mathbf{d}} = f_x(a, b)\cos\alpha + f_y(a, b)\sin\alpha. \tag{17}$$

Exercise 11

Consider the surface defined by $z = f(x, y) = 2x^2y + 3xy^3$.

(a) Find the gradient function $\boldsymbol{\nabla} f(x, y) = f_x\mathbf{i} + f_y\mathbf{j}$.

(b) Find the slope of the surface at the point $(2, 1, 14)$ in the direction of the unit vector $\widehat{\mathbf{d}} = \frac{3}{5}\mathbf{i} + \frac{4}{5}\mathbf{j}$.

Exercise 12

This exercise is quite difficult.

By varying the angle α (measured anticlockwise from the positive x-axis), we can examine the slope of the surface at the fixed point (a, b, c) in any direction we wish.

Calculate the greatest slope of the surface $z = f(x, y) = \frac{1}{2}x^2 + \sqrt{3}y^2$ at the point $(2, 1, 2 + \sqrt{3})$ on the surface. Show that this greatest slope is in the direction of $\boldsymbol{\nabla} f(2, 1)$.

In general, as the result from Exercise 12 illustrates, the direction of the gradient function $\boldsymbol{\nabla} f(x, y)$ at a point (a, b) corresponds to the direction of greatest slope of the surface $z = f(x, y)$ at the point $(a, b, f(a, b))$. To see this, we observe that

$$\boldsymbol{\nabla} f \cdot \widehat{\mathbf{d}} = |\boldsymbol{\nabla} f|\,|\widehat{\mathbf{d}}| \cos\theta = |\boldsymbol{\nabla} f| \cos\theta, \tag{18}$$

where θ is the angle between $\boldsymbol{\nabla} f$ and $\widehat{\mathbf{d}}$. Then $\theta = 0$ gives the maximum value.

We can also see from equation (18) that $\boldsymbol{\nabla} f(x, y)$ at (a, b) points in a direction normal to the contour curve passing through (a, b). To see this, recall that $f(x, y)$ is constant along the contour curve, therefore the slope of the surface $z = f(x, y)$ in directions tangential to a contour curve must be zero (since z does not vary along contour curves). Thus if $\widehat{\mathbf{d}}$ is tangential to a contour curve, then $\boldsymbol{\nabla} f \cdot \widehat{\mathbf{d}} = 0$, implying that $\boldsymbol{\nabla} f$ is perpendicular to $\widehat{\mathbf{d}}$. This is illustrated in Figure 22.

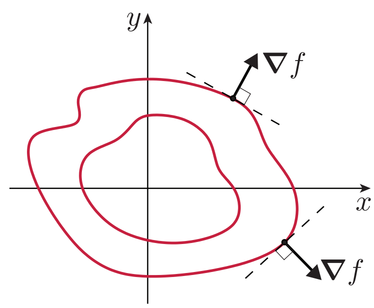


Figure 22 The vector $\boldsymbol{\nabla} f$ points in directions normal to contours of the function $f(x, y)$

Exercise 13

Consider the surface defined by $z = f(x, y) = (x + 2y)^3 - (2x - y)^2$.

(a) Find the gradient function $\boldsymbol{\nabla} f(x, y) = f_x\mathbf{i} + f_y\mathbf{j}$.

(b) Find the slope of the surface at the point $(1, 0, -3)$ in the direction of the vector $\mathbf{i} + \mathbf{j}$.

(c) Find the greatest slope of the surface at the point $(1, 0, -3)$, and the direction of this maximum slope expressed as a unit vector.

(d) Find the direction of a unit vector that is tangential to the contour curve at the point $(1, 0, -3)$.

2 Taylor polynomials

The aim of this unit is to extend the techniques of calculus to functions of two (or more) variables, in order to be able to tackle a wider range of problems in applied mathematics. However, before continuing, it is necessary to review Taylor polynomials and Taylor approximations as they apply to functions of one variable. These are revised in Subsection 2.1. Subsection 2.2 introduces higher-order partial derivatives of functions of two variables (which are conceptually very like their counterparts for functions of one variable). In Subsection 2.3 we generalise Taylor polynomials and Taylor approximations to functions of two variables.

2.1 Functions of one variable

Many useful functions (e.g. trigonometric and exponential functions) cannot generally be evaluated exactly, so the best that we can do is approximate them. Polynomial functions are often used as approximations because they are easy to evaluate and manipulate. In many cases, a good approximation to a function f can be obtained near some point a in the domain of f by finding the polynomial of a chosen degree that agrees with f at a, and also agrees with the first few derivatives of f evaluated at a.

p_1 has been given the subscript 1 because it is a *first*-order polynomial.

For example, the function $f(x) = \sin x$ can be approximated reasonably well near $x = 0$ by the first-order polynomial $p_1(x) = x$ (see Figure 23(a)). So $\sin x \simeq x$ is a good approximation for small x (say $|x| \leq 0.1$). This is because the functions $y = x$ and $y = \sin x$ at $x = 0$ agree in value (they are both 0), in their first derivatives (they are both 1), and in their second derivatives (they are both 0).

Similarly, the function $g(x) = \cos x$ can be approximated quite well near $x = 0$ by the quadratic (second-order) polynomial $p_2(x) = 1 - \frac{1}{2}x^2$ (see Figure 23(b)). This is because $p_2(x)$ and $\cos x$ agree at $x = 0$ in their values, and also in the values of their first, second and third derivatives.

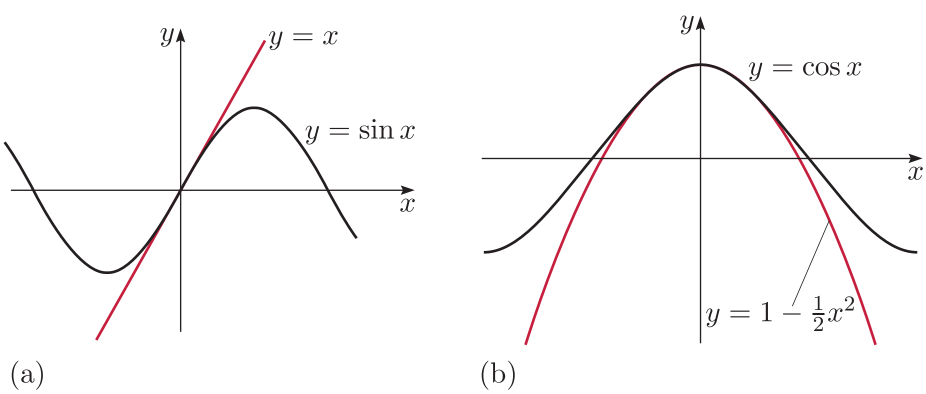


Figure 23 (a) First-order polynomial approximation $y = p_1(x) = x$ to $y = f(x) = \sin x$. (b) Second-order polynomial approximation $y = p_2(x) = 1 - \frac{1}{2}x^2$ to $y = g(x) = \cos x$.

Exercise 14

Verify the above statements concerning $p_1(x)$ and $p_2(x)$. That is, check that at $x = 0$, $p_1(x)$ and $f(x) = \sin x$ agree in value and in their first and second derivatives, and that $p_2(x)$ and $g(x) = \cos x$ agree in value and in their first, second and third derivatives.

We call $p_1(x) = x$ the *tangent approximation* to the function $f(x) = \sin x$ near $x = 0$, or the *Taylor polynomial of degree 1* for $f(x) = \sin x$ about $x = 0$. Similarly, $p_2(x) = 1 - \frac{1}{2}x^2$ is the *quadratic approximation* to the function $g(x) = \cos x$ near $x = 0$, or the *Taylor polynomial of degree 2* for $g(x) = \cos x$ about $x = 0$.

In fact, $p_1(x)$ is also the quadratic approximation to $\sin x$, and $p_2(x)$ is also the cubic approximation to $\cos x$.

Taylor polynomials

For a function $f(x)$ that has n continuous derivatives near $x = a$, the **Taylor polynomial of degree n about** $x = a$, or the **nth-order Taylor polynomial about** $x = a$, is

$$p_n(x) = f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \frac{1}{3!}f'''(a)(x-a)^3 + \cdots + \frac{1}{n!}f^{(n)}(a)(x-a)^n. \tag{19}$$

We use the phrases 'Taylor polynomial of degree n' and 'nth-order Taylor polynomial' synonymously.

Note that at $x = a$, $p_n(x)$ agrees with $f(x)$ in value, and they also agree in the values of their first n derivatives.

If $a = 0$, equation (19) becomes the simpler expression

$$p_n(x) = f(0) + f'(0)x + \frac{1}{2!}f''(0)x^2 + \frac{1}{3!}f'''(0)x^3 + \cdots + \frac{1}{n!}f^{(n)}(0)x^n. \tag{20}$$

This is known as a *Maclaurin series*.

Example 7

For the function $f(x) = e^{2x}$, calculate $f(0)$, $f'(0)$, $f''(0)$ and $f'''(0)$. Write down the third-order Taylor polynomial $p_3(x)$ for $f(x)$ about $x = 0$.

Solution

Since $f(x) = e^{2x}$, it follows that

$$f'(x) = 2e^{2x}, \quad f''(x) = 4e^{2x}, \quad f'''(x) = 8e^{2x}.$$

Hence

$$f(0) = 1, \quad f'(0) = 2, \quad f''(0) = 4, \quad f'''(0) = 8.$$

The third-order Taylor polynomial for e^{2x} about $x = 0$ is therefore

$$p_3(x) = 1 + 2x + \frac{4}{2!}x^2 + \frac{8}{3!}x^3 = 1 + 2x + 2x^2 + \tfrac{4}{3}x^3.$$

For the functions that you will meet in the remainder of this module, successive higher-order Taylor polynomials will give successively better approximations, at least for values of x that are reasonably close to a.

Exercise 15

For $f(x) = e^{2x}$ (as in Example 7), write down the Taylor polynomials $p_0(x)$, $p_1(x)$ and $p_2(x)$ about $x = 0$. Evaluate $p_0(0.1)$, $p_1(0.1)$ and $p_2(0.1)$, and compare these values with the value of $f(0.1)$ obtained on a calculator.

The change in independent variable from x to t has no particular significance.

One important application of Taylor polynomials is in examining the local behaviour of a function. Suppose that the function $f(t)$ has a stationary point at $t = a$ (so $f'(a) = 0$). It can be shown that close to $t = a$, the behaviour of $f(t)$ will be the same as the behaviour of the second-order Taylor polynomial for $f(t)$ about $t = a$. (This is a consequence of Taylor's theorem.) So we can use the second-order Taylor polynomial about $t = a$ to determine the nature of this stationary point. Close to $t = a$, we have

$$f(t) \simeq p_2(t) = f(a) + \frac{1}{2!}f''(a)(t-a)^2.$$

The behaviour of the function $f(t)$ near $t = a$ is determined by the sign of $f''(a)$, assuming that $f''(a) \neq 0$ (see Figure 24). In fact,

$$f(t) - f(a) \simeq \frac{1}{2!}f''(a)(t-a)^2;$$

thus if $f''(a) \neq 0$, the right-hand side of this equation does not change sign near $t = a$. So either $f(t) \geq f(a)$ near $t = a$ (if $f''(a) > 0$), or $f(t) \leq f(a)$ near $t = a$ (if $f''(a) < 0$).

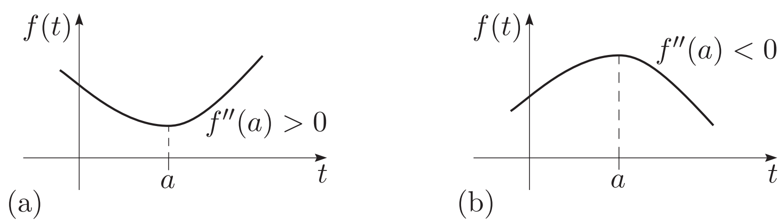


Figure 24 Shape of function $f(t)$ near its stationary point at $t = a$: (a) a minimum if $f''(a) > 0$; (b) a maximum if $f''(a) < 0$

So if $f''(a) > 0$, then $f(t)$ has a local minimum at $t = a$, and if $f''(a) < 0$, then $f(t)$ has a local maximum at $t = a$. However, if $f''(a) = 0$, then the polynomial $p_2(t)$ can tell us nothing about the nature of the stationary point of $f(t)$ at $t = a$. This is the result known as the *second derivative test* that you saw in Section 1. To illustrate the usefulness of Taylor polynomials, let us examine the case when $f''(a) = 0$ a little further. (In this case, the second derivative test is of no help.)

Example 8

Suppose that you are told that $f'(a) = f''(a) = 0$, but $f'''(a) \neq 0$, for a function $f(t)$. If $f(t)$ is approximated by its third-order Taylor polynomial $p_3(t)$, what does $p_3(t)$ tell you about the stationary point of $f(t)$ at $t = a$?

When a Taylor polynomial is used to approximate a function, we refer to the polynomial as a *Taylor approximation* to the function.

Solution

We have

$$f(t) \simeq p_3(t) = f(a) + \frac{1}{3!}f'''(a)(t-a)^3,$$

so

$$f(t) - f(a) \simeq \frac{1}{3!}f'''(a)(t-a)^3.$$

This will change sign as t passes through a (whatever the sign of $f'''(a)$). It follows that the stationary point is neither a local maximum nor a local minimum; such a point is a point of inflection.

You should take particular notice of the reasoning in Example 8, because we will apply a similar process in discussing functions of two variables. It is the sign of $f(t) - f(a)$ near $t = a$ that determines the nature of the stationary point at $t = a$.

If we are concerned with the local behaviour of $f(t)$ near $t = a$, then we are interested in the behaviour of $f(t) - f(a)$ when $t - a$ is small (and certainly less than 1). If $t - a$ is small, then $(t-a)^2$ is even smaller, while higher powers of $t - a$ are smaller still. When, for example, we use the second-order Taylor polynomial as an approximation to $f(t)$, we are effectively ignoring terms involving $(t-a)^3$, $(t-a)^4$, ... This is another idea that extends to functions of two variables.

2.2 Higher-order partial derivatives

In the next subsection we will extend the concept of Taylor polynomials to functions of two variables, but first we must develop the concept of partial derivatives a little further.

You saw in Section 1 that we can differentiate a function of two variables $f(x, y)$ partially with respect to x and partially with respect to y, to obtain the partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$. For example, if $f(x, y) = \sin(xy)$, then $\partial f/\partial x = y\cos(xy)$ and $\partial f/\partial y = x\cos(xy)$. The partial derivatives are themselves functions of two variables, so it is possible to calculate *their* partial derivatives. In the case of the function $f(x, y) = \sin(xy)$, we may differentiate $\partial f/\partial x$ partially with respect to x and obtain

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}\big(y\cos(xy)\big) = -y^2\sin(xy). \tag{21}$$

Each of $\partial f/\partial x$ and $\partial f/\partial y$ can be partially differentiated with respect to either variable, so for this particular function we have, in addition to equation (21),

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial y}\big(y\cos(xy)\big) = \cos(xy) - xy\sin(xy),$$
$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial x}\big(x\cos(xy)\big) = \cos(xy) - xy\sin(xy),$$
$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial}{\partial y}\big(x\cos(xy)\big) = -x^2\sin(xy).$$

Second-order partial derivatives

The **second-order partial derivatives** (or **second partial derivatives**) of a function $f(x,y)$ are

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \quad \frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right),$$
$$\frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right), \quad \frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right).$$

They are often abbreviated as f_{xx}, f_{yy}, f_{xy} and f_{yx}, respectively.

Both notations are in common use, and we will use them interchangeably. f_{xy} represents the partial derivative with respect to x of the partial derivative of f with respect to y, that is, differentiate first with respect to y, then differentiate the result with respect to x.

We can extend the ideas here to obtain higher-order partial derivatives of any order. For example, we can obtain third-order partial derivatives by partially differentiating the second-order partial derivatives.

Example 9

Determine the second-order partial derivatives of the function

$$f(x,y) = e^x\cos y + x^2 - y + 1.$$

Solution

We have

$$\frac{\partial f}{\partial x} = e^x\cos y + 2x, \quad \frac{\partial f}{\partial y} = -e^x\sin y - 1,$$

so

$$\frac{\partial^2 f}{\partial x^2} = e^x\cos y + 2, \quad \frac{\partial^2 f}{\partial x\,\partial y} = -e^x\sin y,$$
$$\frac{\partial^2 f}{\partial y\,\partial x} = -e^x\sin y, \quad \frac{\partial^2 f}{\partial y^2} = -e^x\cos y.$$

Exercise 16

Given $f(x,y) = x\sin(xy)$, calculate f_{xx}, f_{yy}, f_{xy} and f_{yx}.

In both Example 9 and Exercise 16 (as well as in the work on $f(x,y) = \sin(xy)$ at the beginning of this subsection), you can see that

$$\frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial^2 f}{\partial y\,\partial x}, \tag{22}$$

that is, $f_{xy} = f_{yx}$. This is no accident; this result is always true, provided that the function $f(x,y)$ is sufficiently smooth for the second-order partial derivatives to exist and to be continuous.

Theorem 1 Mixed derivative theorem

For any function $f(x,y)$ that is sufficiently smooth for the second-order partial derivatives to exist and to be continuous,

$$\frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial^2 f}{\partial y\,\partial x}.$$

This can be written as $f_{xy} = f_{yx}$.

We assume throughout the remainder of this unit that the functions that we consider are smooth enough for the mixed derivative theorem to apply.

Exercise 17

Given $f(x,y) = e^{2x+3y}$, calculate $f_x(0,0)$, $f_y(0,0)$, $f_{xx}(0,0)$, $f_{yy}(0,0)$, $f_{xy}(0,0)$ and $f_{yx}(0,0)$.

Exercise 18

Determine the second partial derivatives of $f(x,y) = (x^2 + 2y^2 - 3xy)^3$. Evaluate these partial derivatives at $(1,-1)$.

The ideas in this subsection can be extended to functions of more than two variables, though we do not do so here.

2.3 Functions of two variables

In the case of a function f of one variable, say x, you saw in Subsection 2.1 that the nth-order Taylor polynomial agrees with f in value and in the values of the first n derivatives at the chosen point $x = a$. This property of agreement in function value and values of the derivatives is crucial in the definition of Taylor polynomials, and is the property that we generalise to more than one variable.

Consider, for example, the function

$$f(x, y) = e^{x+2y}.$$

If we wish to find a first-order polynomial $p(x, y)$ that approximates $f(x, y)$ near $(0, 0)$, then it seems that we must ensure that p agrees with f in value and in the values of the first partial derivatives at $(0, 0)$. That is, we need to find a first-order polynomial

A first-order (linear) polynomial in one variable has the form $f(x) = c + mx$, where c and m are constants. A first-order polynomial in two variables has linear terms in both variables, so is of the form $f(x, y) = c + mx + ny$, where c, m and n are constants.

$$p(x, y) = \alpha + \beta x + \gamma y, \qquad (23)$$

where α, β and γ are constants, with the properties that

$$p(0, 0) = f(0, 0), \quad p_x(0, 0) = f_x(0, 0), \quad p_y(0, 0) = f_y(0, 0).$$

In our example, $f_x(x, y) = e^{x+2y}$ and $f_y(x, y) = 2e^{x+2y}$, so

$$f(0, 0) = 1, \quad f_x(0, 0) = 1, \quad f_y(0, 0) = 2.$$

Thus, in defining p, we need to choose the coefficients α, β and γ so that

$$p(0, 0) = 1, \quad p_x(0, 0) = 1, \quad p_y(0, 0) = 2.$$

From equation (23), $p(0, 0) = \alpha$, so we need $\alpha = 1$. Also, differentiating equation (23), we obtain

$$p_x(x, y) = \beta, \quad p_y(x, y) = \gamma.$$

In particular,

$$p_x(0, 0) = \beta, \quad p_y(0, 0) = \gamma.$$

So we need $\beta = 1$ and $\gamma = 2$. The required Taylor polynomial is therefore

$$p(x, y) = 1 + x + 2y.$$

More generally, whatever the function $f(x, y)$ may be (provided that we can find its first partial derivatives), we can make $p(x, y) = \alpha + \beta x + \gamma y$ agree with $f(x, y)$ at $(0, 0)$ by setting $\alpha = f(0, 0)$. Also, we can make the first partial derivatives agree at $(0, 0)$ by setting $\beta = f_x(0, 0)$ and $\gamma = f_y(0, 0)$.

First-order Taylor approximation

For a function $f(x, y)$ that is sufficiently smooth near $(x, y) = (0, 0)$, the **first-order Taylor approximation** (or **tangent approximation**) **to** $f(x, y)$ **near** $(0, 0)$ is

$$p_1(x, y) = f(0, 0) + f_x(0, 0)x + f_y(0, 0)y.$$

Exercise 19

Given $f(x, y) = e^{3x-y}$, find the tangent approximation to $f(x, y)$ near $(0, 0)$.

In order to proceed further, we need the notion of higher-order polynomials in two variables.

Nth-order polynomial in two variables

An **Nth-order polynomial** (or **polynomial of degree N**) in two variables, x and y, is a sum of terms $a_{m,n}x^m y^n$ with $m+n \leq N$, where m, n are non-negative integers and $a_{m,n}$ are constants.

As in the case of functions of one variable, we obtain a more accurate approximation than the tangent approximation if we use a second-order polynomial that agrees with the function not only in the above respects, but also in the values of the second partial derivatives at $(0,0)$.

For the moment, we continue to consider approximations near $(0,0)$, though we will generalise shortly.

A general second-order polynomial in x and y takes the form

$$q(x,y) = \alpha + \beta x + \gamma y + Ax^2 + Bxy + Cy^2. \tag{24}$$

In order to fit the value of $q(x,y)$ and its first and second partial derivatives at $(0,0)$ to those of a function $f(x,y)$, it is necessary to determine the partial derivatives of q. You are asked to do this in the next exercise.

In terms of the constants $a_{m,n}$ defined above, we have $a_{0,0} = \alpha$, $a_{1,0} = \beta$, $a_{0,1} = \gamma$, $a_{2,0} = A$, $a_{1,1} = B$ and $a_{0,2} = C$.

Exercise 20

Consider $q(x,y)$ described in equation (24).

(a) Find the functions q_x, q_y, q_{xx}, q_{xy} and q_{yy}.

(b) Evaluate $q(x,y)$ and its first and second partial derivatives at $(0,0)$.

From the results of Exercise 20, it follows that in order to approximate $f(x,y)$ near $(0,0)$ by $q(x,y)$ in equation (24), we must set

$$\alpha = f(0,0), \quad \beta = f_x(0,0), \quad \gamma = f_y(0,0),$$
$$A = \tfrac{1}{2}f_{xx}(0,0), \quad B = f_{xy}(0,0), \quad C = \tfrac{1}{2}f_{yy}(0,0).$$

Second-order Taylor approximation

For a function $f(x,y)$ that is sufficiently smooth near $(x,y) = (0,0)$, the **second-order Taylor approximation** (or **quadratic approximation**) **to $f(x,y)$ near $(0,0)$** is

$$\begin{aligned} p_2(x,y) = {} & f(0,0) + f_x(0,0)x + f_y(0,0)y \\ & + \tfrac{1}{2}\left(f_{xx}(0,0)x^2 + 2f_{xy}(0,0)xy + f_{yy}(0,0)y^2\right). \end{aligned} \tag{25}$$

If we take only the *linear* terms in equation (25), then we obtain the first-order Taylor polynomial. This is the equation of the *tangent plane* at $(0,0)$. The reason why it is called the tangent plane should become clear from the following exercise.

Exercise 21

Verify that

$$z = g(x,y) = f(0,0) + f_x(0,0)x + f_y(0,0)y$$

is the equation of a plane through $(0,0,f(0,0))$.

Show that the plane has the same gradient as the surface $z = f(x,y)$ at $(0,0,f(0,0))$; that is, show that $\boldsymbol{\nabla} g(0,0) = \boldsymbol{\nabla} f(0,0)$.

It is also possible to obtain similar approximations for $f(x,y)$ near an arbitrary point (a,b). In this case, x is replaced by $x-a$, y is replaced by $y-b$, and the function f and its partial derivatives are evaluated at (a,b).

Taylor polynomials about arbitrary points

For a function $f(x,y)$ that is sufficiently smooth near $(x,y) = (a,b)$, the **first-order Taylor polynomial for $f(x,y)$ about (a,b)** (or **tangent approximation to $f(x,y)$ near (a,b)**) is

$$p_1(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b).$$

The **tangent plane** to the surface $z = f(x,y)$ at $(a,b,f(a,b))$ is given by $z = p_1(x,y)$.

For a function $f(x,y)$ that is sufficiently smooth near $(x,y) = (a,b)$, the **second-order Taylor polynomial for $f(x,y)$ about (a,b)** (or **quadratic approximation to $f(x,y)$ near (a,b)**) is

$$\begin{aligned} p_2(x,y) = f(a,b) &+ f_x(a,b)(x-a) + f_y(a,b)(y-b) \\ &+ \tfrac{1}{2}\big(f_{xx}(a,b)(x-a)^2 + 2f_{xy}(a,b)(x-a)(y-b) \\ &\qquad + f_{yy}(a,b)(y-b)^2\big). \end{aligned} \tag{26}$$

Example 10

Determine the Taylor polynomials of degrees 1 and 2 about $(2,1)$ for the function $f(x,y) = x^3 + xy - 2y^2$.

Solution

Differentiating the function partially with respect to x and partially with respect to y gives

$$f_x = 3x^2 + y, \quad f_y = x - 4y.$$

Differentiating partially again gives

$$f_{xx} = 6x, \quad f_{xy} = 1, \quad f_{yy} = -4.$$

It follows that

$$f(2,1) = 8, \quad f_x(2,1) = 13, \quad f_y(2,1) = -2,$$
$$f_{xx}(2,1) = 12, \quad f_{xy}(2,1) = 1, \quad f_{yy}(2,1) = -4,$$

therefore

$$\begin{aligned} p_1(x,y) &= 8 + 13(x-2) - 2(y-1), \\ p_2(x,y) &= 8 + 13(x-2) - 2(y-1) \\ &\qquad + 6(x-2)^2 + (x-2)(y-1) - 2(y-1)^2. \end{aligned}$$

Exercise 22

Use the results of Exercise 17 to write down the second-order Taylor polynomial for $f(x,y) = e^{2x+3y}$ about $(0,0)$.

Exercise 23

In Exercise 2 we discussed the potential energy U of the mechanical system shown in Figure 5, and you saw that $U(\theta,\phi) = -mga(\cos\theta + \cos\phi)$. Find the second-order Taylor polynomial for $U(\theta,\phi)$ about $(0,0)$.

Exercise 24

Determine the second-order Taylor polynomial about $(0,0)$ for the function $f(x,y) = e^{xy} + (x+y)^2$.

Exercise 25

(a) Determine the Taylor polynomials of degrees 1 and 2 about $(0,0)$ for the function $f(x,y) = (2+x+2y)^3$. Compare your answers to the expression obtained by expanding $(2+x+2y)^3$.

(b) Determine the Taylor polynomials of degrees 1 and 2 about $(1,-1)$ for the function $f(x,y) = (2+x+2y)^3$.

(c) Putting $X = x-1$ and $Y = y+1$ (so that $X = 0$ when $x = 1$, and $Y = 0$ when $y = -1$), we have

$$\begin{aligned} f(x,y) = (2+x+2y)^3 &= \big(2 + (X+1) + 2(Y-1)\big)^3 \\ &= (1+X+2Y)^3. \end{aligned}$$

Write down the Taylor polynomials of degrees 1 and 2 about $(0,0)$ for the function $F(X,Y) = (1+X+2Y)^3$.

We could extend the arguments of this subsection to define Taylor polynomials of order higher than two, or even for functions of more than two variables, but we do not do so here.

3 Classification of stationary points

The main purpose of this section is to extend to functions of two variables the second derivative test that was discussed in Section 1. If $f(x, y)$ is sufficiently smooth so that the first and second partial derivatives are defined, and $f_x(a, b) = f_y(a, b) = 0$ at some point (a, b), then there is a distinct possibility that $f(x, y)$ has a local maximum or a local minimum at (a, b).

In the case of a function of one variable, the sign of the second derivative at a stationary point is enough to distinguish whether that point is a local maximum or a local minimum (provided that the second derivative is not zero at that point). The situation is more complicated for functions of two variables (for a start, there are three second partial derivatives to consider), but similarly, a knowledge of the values of these derivatives at a stationary point will often tell us whether it is a local maximum, a local minimum, or neither.

3.1 Extrema

In searching for the local maxima and local minima of a function of one variable, the first step is to locate the stationary points, that is, the points where the derivative is zero. The same is true in the case of functions of two or more variables.

The definition of a stationary point can be extended to functions of three or more variables.

Stationary points

A **stationary point** of a function $f(x, y)$ is a point (a, b) in the domain of $f(x, y)$ at which $f_x(a, b) = f_y(a, b) = 0$.

The corresponding point $(a, b, f(a, b))$ on the surface S defined by $z = f(x, y)$ is a **stationary point** on S.

Example 11

Locate the stationary points of the function $f(x, y) = 5 + (x - 1)^3 + y^2$.

Solution

Partially differentiating gives $f_x = 3(x - 1)^2$, which is zero when $x = 1$, and $f_y = 2y$, which is zero when $y = 0$. So $(1, 0)$ is the only stationary point (corresponding to the point $(1, 0, 5)$ on the surface).

Generally, to find the stationary points, we need to solve a pair of simultaneous equations, as the following example shows.

Example 12

Locate the stationary points of the function

$$f(x,y) = x^2 + y^2 + (x-1)(y+2).$$

Solution

Partially differentiating gives $f_x = 2x + y + 2$ and $f_y = 2y + x - 1$. To find the stationary points, we need to solve the pair of simultaneous equations

$$\begin{aligned} 2x + y &= -2, \\ x + 2y &= 1. \end{aligned}$$

We find that $x = -\frac{5}{3}$ and $y = \frac{4}{3}$. It follows that $\left(-\frac{5}{3}, \frac{4}{3}\right)$ is the only stationary point.

Exercise 26

Locate the stationary points of the function

$$f(x,y) = 3x^2 - 4xy + 2y^2 + 4x - 8y.$$

A word of warning! The simultaneous equations that must be solved in order to find stationary points are in general non-linear. For example, if

$$f(x,y) = e^{x+2y} + x^4 y + xy^3,$$

then $f_x = e^{x+2y} + 4x^3 y + y^3$ and $f_y = 2e^{x+2y} + x^4 + 3xy^2$, so we need to solve the pair of simultaneous equations

$$\begin{aligned} &e^{x+2y} + 4x^3 y + y^3 = 0, \\ &2e^{x+2y} + x^4 + 3xy^2 = 0. \end{aligned}$$

We will not ask you to tackle problems as difficult as this by hand, but the next exercise involves a pair of non-linear equations that can be solved by factorisation.

Exercise 27

Locate the stationary points of the function $f(x,y) = xy(x+y-3)$.

If a function f of n variables is smooth enough to have first partial derivatives everywhere on $\mathbb{R}^n$, then any local maxima or local minima that exist will occur at stationary points. For a function of one variable, we could apply the second derivative test to the stationary points. We now generalise this to a method of classifying the stationary points of functions of several variables, using second partial derivatives.

The purpose of this subsection is to describe a method of doing this. But before going any further, we need definitions of 'local maximum' and 'local minimum' for functions of more than one variable. We will state such definitions for functions of two variables; it is not difficult to see how to generalise these to several variables.

Extrema

A function $f(x, y)$, defined on a domain D, has a **local minimum** at (a, b) in D if for all (x, y) in D sufficiently close to (a, b), we have $f(x, y) \geq f(a, b)$.

A function $f(x, y)$, defined on a domain D, has a **local maximum** at (a, b) in D if for all (x, y) in D sufficiently close to (a, b), we have $f(x, y) \leq f(a, b)$.

A point that is either a local maximum or a local minimum is an **extremum** (and vice versa).

If the function $f(x, y)$ has an extremum at (a, b), then the section function $z = f(x, b)$ (a function of x only) must also have an extremum at $x = a$, so assuming that $f(x, b)$ is sufficiently smooth, $(\partial f/\partial x)(a, b) = 0$. Similarly, $(\partial f/\partial y)(a, b) = 0$. It follows that every extremum of $f(x, y)$, assuming that the function is sufficiently smooth, is a stationary point of $f(x, y)$. However, not every stationary point is necessarily an extremum.

Saddle points

A stationary point of a function of two variables that is not an extremum is a **saddle point**.

The term saddle point originates from the shape of surfaces near some such points for functions of two variables. An example is provided by the shape of the hyperboloid shown in Figure 8(b) near a stationary point (in this case, the origin) – it looks like a rider's saddle. (A saddle drops *down* on either side of the rider, but rises *up* in front and behind.) For completeness, it is worth noting that there are more complicated possibilities for the shapes of surfaces near stationary points that are not extrema; all such points are referred to as saddle points, however.

Exercise 28

In Exercise 2 you found that the potential energy U of the mechanical system shown in Figure 5 can be written as $U(\theta, \phi) = -mga(\cos\theta + \cos\phi)$. Show that $U(\theta, \phi)$ has a stationary point at $(0, 0)$, and that this point is a local minimum.

If we are looking for the extrema of a given function, then we know that we should look among the stationary points. However, not all cases are as straightforward as Exercise 28, and we need some general means of classifying them (much as we have for functions of one variable). You will see next that we can use the second-order Taylor polynomial to construct a useful test that will often distinguish between local maxima, local minima and saddle points.

Let $f(x,y)$ have a stationary point at (a,b). To ensure that $f(x,y)$ has a local minimum at (a,b), it is not enough to stipulate that each of the section functions $f(x,b)$ and $f(a,y)$ through (a,b) has a local minimum at that point. There may still be directions through (a,b) along which the value of $f(x,y)$ decreases as we move away from (a,b).

For example, consider the function $f(x,y) = x^2 + 6xy + 7y^2$, which possesses a stationary point at $(0,0)$. The section functions through $(0,0)$ are $f(x,0) = x^2$ and $f(0,y) = 7y^2$, each of which has a local minimum at $(0,0)$. But let us move along the parametrised curve (actually a straight line) given by $x(t) = 2t$, $y(t) = -t$. Then

$f_x = 2x + 6y$, so $f_x(0,0) = 0$.
$f_y = 6x + 14y$, so $f_y(0,0) = 0$.

$$f(x(t), y(t)) = (2t)^2 + 6(2t)(-t) + 7(-t)^2 = -t^2.$$

As we move along this line from $(0,0)$, the value of $f(x(t), y(t))$ becomes negative. As $f(0,0) = 0$, it follows that $(0,0)$ cannot be a local minimum of f.

One way to understand the reason for this behaviour is to express the function as a difference of two squares:

$$f(x,y) = x^2 + 6xy + 7y^2 = 2(x+2y)^2 - (x+y)^2.$$

Strictly, as a difference of two squares this is
$$(\sqrt{2}x + 2\sqrt{2}y)^2 - (x+y)^2.$$

Then no matter how close to the origin we look, we see some points where the function is positive and some where it is negative. In particular, in the direction along which $x + 2y = 0$, the expression on the right-hand side reduces to $-(x+y)^2$, which is negative except when $(x,y) = (0,0)$. However, in the direction $x + y = 0$, the expression is $2(x+2y)^2$, which is positive except when $(x,y) = (0,0)$. Therefore $(0,0)$ is a saddle point of $f(x,y)$.

In general, $f(x,y)$ may not be a function that can be manipulated as easily as the polynomial above. Nevertheless, the quadratic approximation to $f(x,y)$ about (a,b) *is* a polynomial, and you will see in the next subsection that its second-order terms can usually be manipulated in such a way as to enable us to classify the stationary point at (a,b). The test that we will derive, based on expressing the quadratic approximation in terms of a matrix, is similar to the second derivative test for functions of one variable.

Before going on to derive this test, it is worth recalling the logic behind the second derivative test. The quadratic approximation to a function $f(x)$ of one variable about $x = a$ is given by

$$f(x) \simeq f(a) + f'(a)(x-a) + \tfrac{1}{2}f''(a)(x-a)^2.$$

If a is a stationary point, then $f'(a) = 0$ so $f(x) - f(a) \simeq \frac{1}{2}f''(a)(x-a)^2$.

Therefore (provided that $f''(a) \neq 0$) the quadratic approximation will have a minimum or a maximum, depending on the sign of $f''(a)$. Close enough to a, the approximation will behave like the function itself, thus allowing us to conclude that the function has a local minimum or maximum. Thus the second derivative test uses the second-order terms in the quadratic approximation to classify the stationary point.

3.2 Classifying stationary points using eigenvalues: functions of two variables

Consider a stationary point of $f(x,y)$ at (a,b) so that we have $f_x(a,b) = f_y(a,b) = 0$. It will prove convenient to introduce the variables p and q defined as displacements from the stationary point as $p = x - a$ and $q = y - b$. It will also be useful to write

$$A = f_{xx}(a,b), \quad B = f_{xy}(a,b)\,\big(= f_{yx}(a,b)\big), \quad C = f_{yy}(a,b). \tag{27}$$

As for functions of one variable, *close* to a point (a,b), the second-order Taylor polynomial for $f(x,y)$ about (a,b) behaves in the same way as $f(x,y)$ itself. This is a consequence of a theorem (similar to Taylor's theorem) that we do not include in this module.

The quadratic approximation to $f(x,y)$ is the second-order Taylor polynomial. Since we are assuming that the first derivatives are zero at (a,b), equation (26) gives

$$f(x,y) \simeq f(a,b) + \tfrac{1}{2}Q(x-a, y-b),$$

where

$$Q(p,q) = Ap^2 + 2Bpq + Cq^2.$$

Thus we will be able to classify the stationary point if we can determine the sign of $Q(p,q)$.

The first step in doing this is to define the real symmetric matrix

$$\mathbf{H} = \begin{pmatrix} A & B \\ B & C \end{pmatrix}, \tag{28}$$

The polynomial $Q(p,q)$ is an example of a *quadratic form*, and it is always possible to express a quadratic form in terms of a symmetric matrix as in equation (29).

which enables us to write

$$Q(p,q) = \begin{pmatrix} p & q \end{pmatrix} \mathbf{H} \begin{pmatrix} p \\ q \end{pmatrix}. \tag{29}$$

Exercise 29

Verify equation (29).

Let λ_1 and λ_2 be the eigenvalues of $\mathbf{H}$, with corresponding eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$. Note that $\mathbf{H}$ is real and symmetric. We learned in Unit 5 that the eigenvalues of a real symmetric matrix are *real*, and that its eigenvectors are *real* and can be made *orthogonal*, that is, $\mathbf{v}_1^T\mathbf{v}_2 = 0$. (More specifically, for real symmetric matrices, if the eigenvalues are distinct, then the corresponding eigenvectors are orthogonal; otherwise, any two linearly independent vectors are eigenvectors and we are free to select any two that are mutually orthogonal.)

Furthermore, recall that an eigenvector is specified up to a constant scalar multiple. Thus, for later convenience, we are free to select the scalar multiple so that the eigenvectors have magnitude 1, that is, they are chosen to be *unit* vectors, so that $|\mathbf{v}_1|^2 = \mathbf{v}_1^T\mathbf{v}_1 = 1$ and similarly for $\mathbf{v}_2$. Eigenvectors with this property are said to be *normalised*. So, to summarise, we choose $\mathbf{v}_1$ and $\mathbf{v}_2$ to have the following properties:

$$\mathbf{v}_1^T\mathbf{v}_1 = \mathbf{v}_2^T\mathbf{v}_2 = 1, \quad \mathbf{v}_1^T\mathbf{v}_2 = \mathbf{v}_2^T\mathbf{v}_1 = 0. \tag{30}$$

Now, the vector $(p \quad q)^T$ can be expressed as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$:

$$\begin{pmatrix} p \\ q \end{pmatrix} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2,$$

where c_1 and c_2 are real and depend on x and y, which are kept close to the stationary-point values. In fact, the values of c_1 and c_2 indicate the direction of (x, y) relative to the stationary point (a, b). Hence

For example, if $c_1 \neq 0$ and $c_2 = 0$, then $(p \quad q)^T$ points in the direction of $\mathbf{v}_1$, whereas it points in the direction of $\mathbf{v}_2$ if $c_1 = 0$ and $c_2 \neq 0$.

$$\begin{aligned} \mathbf{H}\begin{pmatrix} p \\ q \end{pmatrix} &= c_1\mathbf{H}\mathbf{v}_1 + c_2\mathbf{H}\mathbf{v}_2 \\ &= c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2, \end{aligned}$$

and therefore

$$\begin{aligned} Q(p,q) &= \begin{pmatrix} p & q \end{pmatrix}\mathbf{H}\begin{pmatrix} p \\ q \end{pmatrix} \\ &= (c_1\mathbf{v}_1^T + c_2\mathbf{v}_2^T)(c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2) \\ &= \lambda_1c_1^2\mathbf{v}_1^T\mathbf{v}_1 + \lambda_2c_1c_2\mathbf{v}_1^T\mathbf{v}_2 + \lambda_1c_2c_1\mathbf{v}_2^T\mathbf{v}_1 + \lambda_2c_2^2\mathbf{v}_2^T\mathbf{v}_2. \end{aligned}$$

Finally, applying the normalisation and orthogonality properties of the eigenvectors, as given in equations (30), leads to the expression

$$Q(p,q) = \lambda_1c_1^2 + \lambda_2c_2^2. \tag{31}$$

Since c_1 and c_2 are real, the sign of $Q(p, q)$ is wholly dependent on the signs of λ_1 and λ_2, the eigenvalues of $\mathbf{H}$. If both eigenvalues are positive, then $Q(p, q)$ is always positive, so the stationary point must be a local minimum. If both eigenvalues are negative, then $Q(p, q)$ is always negative, so the stationary point is a local maximum. However, if one eigenvalue is positive and the other is negative, then the sign of $Q(p, q)$ will depend on the direction of travel, which indicates that the stationary point is a saddle point. Finally, if one or both of the eigenvalues is zero, then the nature of the stationary point cannot be determined using this method – we need to consider higher-order terms in the Taylor expansion. For, clearly, if $\lambda_1 = \lambda_2 = 0$, then $Q(p, q) = 0$ along all directions, so the test is inconclusive. And if just one of the eigenvalues is zero, then $Q(p, q) = 0$ along the direction corresponding to the eigenvector with zero eigenvalue, so the test is still inconclusive.

Thus this so-called *eigenvalue test* can be summarised as follows.

Eigenvalue test for classifying a stationary point

Given that the (sufficiently smooth) function $f(x, y)$ has a stationary point at (a, b), let

$$A = f_{xx}(a, b), \quad B = f_{xy}(a, b) \left(= f_{yx}(a, b)\right), \quad C = f_{yy}(a, b)$$

and

$$\mathbf{H} = \begin{pmatrix} A & B \\ B & C \end{pmatrix},$$

and let λ_1 and λ_2 (which must be real) be the eigenvalues of $\mathbf{H}$.

- If λ_1 and λ_2 are both positive, then there is a local minimum at (a, b).
- If λ_1 and λ_2 are both negative, then there is a local maximum at (a, b).
- If λ_1 and λ_2 are non-zero and opposite in sign, then there is a saddle point at (a, b).
- If either, or both, of λ_1 or λ_2 is zero, then the test is inconclusive.

Example 13

Locate and classify the stationary point of the function

$$f(x, y) = 2x^2 - xy - 3y^2 - 3x + 7y.$$

Solution

Start by determining the first partial derivatives:

$$f_x = 4x - y - 3, \quad f_y = -x - 6y + 7.$$

The stationary point is determined by $f_x = f_y = 0$. From $f_x = 0$, $y = 4x - 3$, and substituting this into $f_y = 0$ gives $7 = x + 6y = 25x - 18$, so there is a stationary point at $(1, 1)$. The second partial derivatives are

$$f_{xx} = 4, \quad f_{xy} = f_{yx} = -1, \quad f_{yy} = -6.$$

It follows that $\mathbf{H} = \begin{pmatrix} 4 & -1 \\ -1 & -6 \end{pmatrix}$, and the characteristic equation is $\lambda^2 + 2\lambda - 25 = 0$. The eigenvalues are therefore $\sqrt{26} - 1$ and $-\sqrt{26} - 1$. The eigenvalues are non-zero and of opposite sign, so the stationary point is a saddle point.

3.3 $AC - B^2$ criterion

The eigenvalue test can be extended to functions of more than two variables, as will be explained in the next subsection. In this subsection we continue to confine ourselves to functions of just two variables and show that it is possible to gain knowledge of the signs of the eigenvalues of $\mathbf{H}$ without explicitly evaluating them, in such a way that the nature of the stationary point can be deduced. It will turn out that this can be done solely from the sign of $AC - B^2$ and the sign of A if the former is positive, where A, B and C are given by equations (27). This leads to the somewhat simpler $AC - B^2$ test for classifying a fixed point.

To arrive at this test, recall from Unit 5 the following results for the sum and product of the eigenvalues λ_1 and λ_2 of the matrix $\mathbf{H} = \begin{pmatrix} A & B \\ B & C \end{pmatrix}$ in equation (28):

We showed in Unit 5 that the sum and product of the eigenvalues of any matrix $\mathbf{M}$ are $\operatorname{tr}\mathbf{M}$ and $\det\mathbf{M}$, respectively.

$$\lambda_1 + \lambda_2 = \operatorname{tr}\mathbf{H} = A + C, \tag{32}$$

$$\lambda_1\lambda_2 = \det\mathbf{H} = AC - B^2. \tag{33}$$

We now consider three cases for $AC - B^2$.

- $AC - B^2 > 0$: It follows from equation (33) that $\lambda_1\lambda_2 > 0$, so λ_1 and λ_2 have the same sign. Moreover, $AC > B^2 \geq 0$, so A and C have the same sign. Thus from equation (32), the sign of λ_1 and λ_2 is given by the sign of A. Hence if $A > 0$, λ_1 and λ_2 are both positive, so the stationary point is a minimum. Conversely, if $A < 0$, λ_1 and λ_2 are both negative, so the stationary point is a maximum.
- $AC - B^2 < 0$: It follows from equation (33) that $\lambda_1\lambda_2 < 0$, so λ_1 and λ_2 are non-zero but have opposite signs. Hence the stationary point is a saddle point.
- $AC - B^2 = 0$: It follows from equation (33) that $\lambda_1\lambda_2 = 0$, so either one or both of the eigenvalues is zero. Therefore the nature of the stationary point cannot be determined using this method.

Hence the resulting test can be summarised as follows.

$AC - B^2$ test for classifying a stationary point

Given that the (sufficiently smooth) function $f(x, y)$ has a stationary point at (a, b), let

$$A = f_{xx}(a,b), \quad B = f_{xy}(a,b)\ \big(= f_{yx}(a,b)\big), \quad C = f_{yy}(a,b).$$

- If $AC - B^2 > 0$, there is:
  - a local minimum at (a, b) if $A > 0$
  - a local maximum at (a, b) if $A < 0$.
- If $AC - B^2 < 0$, there is a saddle point at (a, b).
- If $AC - B^2 = 0$, the test is unable to classify the stationary point.

Note that this test is significantly simpler than the previous eigenvalue test for classifying a stationary point, in that it just requires the calculation of A, B and C, and the evaluation of $AC - B^2$, without having to explicitly solve the characteristic equation to determine the eigenvalues.

Example 14

Locate and classify the stationary point of the function

$$f(x,y) = e^{-(x^2+y^2)}.$$

Solution

Partially differentiating gives $f_x = -2xe^{-(x^2+y^2)}$ and $f_y = -2ye^{-(x^2+y^2)}$. Since $f_x = 0$ only when $x = 0$, and $f_y = 0$ only when $y = 0$, the stationary point is at $(0,0)$.

Since $f_{xx} = -2e^{-(x^2+y^2)} + 4x^2e^{-(x^2+y^2)}$, we have $A = f_{xx}(0,0) = -2$. Also, $f_{yy} = -2e^{-(x^2+y^2)} + 4y^2e^{-(x^2+y^2)}$, therefore $C = f_{yy}(0,0) = -2$. Finally, $f_{xy} = 4xye^{-(x^2+y^2)}$, therefore $B = f_{xy}(0,0) = 0$.

So we see that $AC - B^2 = 4 > 0$, and since $A = -2 < 0$, the stationary point is a local maximum.

Use the $AC - B^2$ test to do the following four exercises.

Exercise 30

Locate and classify the stationary point of the function

$$f(x,y) = 2x^2 - xy - 3y^2 - 3x + 7y.$$

Exercise 31

Locate and classify the four stationary points of the function

$$f(x,y) = x^3 - 12x - y^3 + 3y.$$

Exercise 32

Locate and classify the stationary points of the following functions.

(a) $f(x,y) = \sqrt{1 - x^2 + y^2}$

(b) $T(x,y) = \cos x + \cos y$

Exercise 33

Locate and classify the four stationary points of the function

$$f(x,y) = 3x^2y^3 - 9x^2y + x^3 - 15x - 50.$$

(The function has two extrema and two saddle points.)

3.4 Classifying stationary points using eigenvalues: functions of n variables

The eigenvalue test can be extended to functions of three or more variables. If f is a sufficiently smooth function of n variables, then a stationary point of f is a point $(a_1, a_2, \ldots, a_n)$ at which all the n first partial derivatives are zero. There will be n^2 second partial derivatives, and their values at $(a_1, a_2, \ldots, a_n)$ can be written in the form of an $n \times n$ matrix $\mathbf{H}$, the entry in the ith row and jth column being the value $\dfrac{\partial^2 f}{\partial x_i\, \partial x_j}(a_1, \ldots, a_n)$. This is called the **Hessian matrix** of f at the point $(a_1, \ldots, a_n)$.

Note that $\mathbf{H}$ in equation (28) is a two-dimensional version of the Hessian matrix.

It turns out that, as with the two-variable case, in the vicinity of the stationary point, f is well approximated by the quadratic approximation

$$f(x_1, \ldots, x_n) \simeq f(a_1, \ldots, a_n) + \tfrac{1}{2}\mathbf{p}^T\mathbf{H}\mathbf{p},$$

where

$$\mathbf{p}^T = (x_1 - a_1, \ldots, x_n - a_n).$$

Recall that $\frac{\partial f}{\partial x_i}(a_1, \ldots, a_n) = 0$, $i = 1, \ldots, n$, for a stationary point.

Our assumption that f is sufficiently smooth implies that $\mathbf{H}$ will be a *symmetric matrix*, since the mixed derivative theorem extends from the two-variable case to the several-variable case, so that $\mathbf{p}^T\mathbf{H}\mathbf{p}$ is a quadratic form whose sign determines the nature of the stationary point. Moreover, since $\mathbf{H}$ is a real symmetric matrix, its eigenvalues $\lambda_1, \ldots, \lambda_n$ are real, and its eigenvectors are real and can be made orthogonal. Hence, as with the two-variable case, we have

$$\mathbf{p}^T\mathbf{H}\mathbf{p} = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2, \tag{34}$$

where $c_1, \ldots, c_n$ are real. Equation (34) is just the n-variable generalisation of equation (31), and its derivation follows similarly, where now $\mathbf{H}$ has n orthogonal normalised eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and we write $\mathbf{p} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Equation (34) now leads to the following classification.

We have

$$\mathbf{v}_i^T\mathbf{v}_j = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases}$$

- If all the eigenvalues are positive, then there is a local minimum at $(a_1, \ldots, a_n)$.
- If all the eigenvalues are negative, then there is a local maximum at $(a_1, \ldots, a_n)$.
- If all the eigenvalues are non-zero but they are not all of the same sign, then there is a saddle point at $(a_1, \ldots, a_n)$.
- If one or more of the eigenvalues is zero, then the test is inconclusive.

In the case of more than two variables, the analogy of a saddle is not very helpful in visualising a saddle point. However, it remains true that there are sections through a saddle point giving a section function with a local minimum, and others giving a section function with a local maximum.

Example 15

Locate and classify the stationary point of the function

$$w(x, y, z) = 3y^2 + 3z^2 - 4xy - 2yz - 4zx.$$

Solution

We find the stationary point by solving the simultaneous equations

$$w_x = -4y - 4z = 0,$$
$$w_y = 6y - 4x - 2z = 0,$$
$$w_z = 6z - 2y - 4x = 0.$$

The only solution is $x = y = z = 0$.

The second partial derivatives of $w(x, y, z)$ are constants:

$$w_{xx} = 0, \quad w_{yy} = 6, \quad w_{zz} = 6,$$
$$w_{xy} = w_{yx} = -4, \quad w_{yz} = w_{zy} = -2, \quad w_{zx} = w_{xz} = -4.$$

Thus the required 3×3 Hessian matrix is

$$\mathbf{H} = \begin{pmatrix} w_{xx} & w_{xy} & w_{xz} \\ w_{yx} & w_{yy} & w_{yz} \\ w_{zx} & w_{zy} & w_{zz} \end{pmatrix} = \begin{pmatrix} 0 & -4 & -4 \\ -4 & 6 & -2 \\ -4 & -2 & 6 \end{pmatrix}.$$

To find the eigenvalues of $\mathbf{H}$, we find the values of λ that satisfy the characteristic equation $\det(\mathbf{H} - \lambda\mathbf{I}) = 0$. The solutions of the equation $\det(\mathbf{H} - \lambda\mathbf{I}) = -(\lambda - 8)^2(\lambda + 4) = 0$ are $\lambda_1 = -4$, $\lambda_2 = 8$ and $\lambda_3 = 8$. Our test then tells us immediately that there is a saddle point at the origin.

Exercise 34

Locate and classify the stationary point of the function

$$w(x, y, z) = 3x^2 + 3y^2 + 4z^2 - 2xy - 2yz - 2xz,$$

given that the characteristic equation of the matrix

$$\begin{pmatrix} 6 & -2 & -2 \\ -2 & 6 & -2 \\ -2 & -2 & 8 \end{pmatrix}$$

is $(8 - \lambda)\left((6 - \lambda)^2 - 12\right) = 0$.

Exercise 35

Classify the stationary point at the origin of the function

$$w(x, y, z) = x^2 + 2y^2 + z^2 + 2\sqrt{3}xz.$$

Learning outcomes

After studying this unit, you should be able to:

- calculate first and second partial derivatives of a function of several variables
- understand the use of a surface to represent a function of two variables
- construct the equation of the tangent plane at a given point on a surface
- calculate the Taylor polynomials of degree n for a function of one variable
- calculate the first-order and second-order Taylor polynomials for a function of two variables
- locate the stationary points of a function of two (or more) variables by solving a system of two (or more) simultaneous equations
- classify the stationary points of a function of two variables by using the $AC - B^2$ test
- classify the stationary points of a function of two (or more) variables by examining the signs of the eigenvalues of an appropriate matrix.

Solutions to exercises

Solution to Exercise 1

(a) $f(2,3) = 12 - 54 = -42$.

(b) $f(3,2) = 27 - 16 = 11$.

(c) $f(a,b) = 3a^2 - 2b^3$.

(d) $f(b,a) = 3b^2 - 2a^3$.

(e) $f(2a,b) = 3(2a)^2 - 2b^3 = 12a^2 - 2b^3$.

(f) $f(a-b,0) = 3(a-b)^2$.

(g) $f(x,2) = 3x^2 - 16$.

(h) $f(y,x) = 3y^2 - 2x^3$.

Solution to Exercise 2

The potential energy U of a particle of mass m placed at height h (relative to O) is given by $U(h) = mgh$. We can see that A is $a\cos\theta$ below O, and B is $a\cos\phi$ below A; so B is $a(\cos\theta + \cos\phi)$ below O. Thus in this case we have

$$U(\theta,\phi) = -mga(\cos\theta + \cos\phi).$$

The least possible value of U occurs when $\cos\theta$ and $\cos\phi$ take their greatest values, and this happens when $\theta = \phi = 0$. So the least value of U is $U(0,0) = -2mga$, and it occurs when the system is hanging vertically.

Solution to Exercise 3

The section function of $F(x,y) = 100e^{-(x^2+y^2)}$ with y fixed at 0 is $F(x,0) = 100e^{-x^2}$. So

$$\frac{d}{dx}(F(x,0)) = -200xe^{-x^2}.$$

This derivative is zero when $x = 0$, so $F(x,0)$ has a stationary point at $x = 0$.

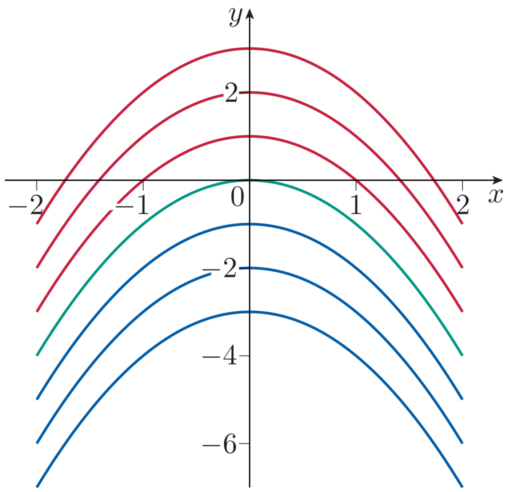
Differentiating again with respect to x, we obtain

$$\frac{d^2}{dx^2}\big(F(x,0)\big) = -200\,\frac{d}{dx}\big(xe^{-x^2}\big) = -200(1-2x^2)e^{-x^2}.$$

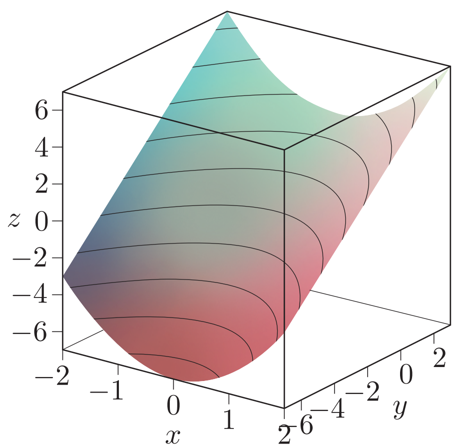
This derivative equals -200 when $x = 0$. It follows that the section function has a local maximum at $x = 0$.

Solution to Exercise 4

The contours are given by $f(x,y) = x^2 + y = c$ or $y = c - x^2$ for various constant values of c. These are shown in the following figure, with red contours for $c > 0$, blue for $c < 0$, and green for $c = 0$.



Some additional insight can be gained from viewing the contours superimposed on the surface $z = f(x, y) = x^2 + y$ in a three-dimensional plot (although this was not requested), as in the following figure.



Solution to Exercise 5

$$\frac{\partial f}{\partial x} = 3x^2 \cos y + y^2 \cos x,$$

$$\frac{\partial f}{\partial y} = -x^3 \sin y + 2y \sin x.$$

Solution to Exercise 6

$$\frac{\partial f}{\partial x} = 2x \sin(xy) + (x^2 + y^3) y \cos(xy),$$

$$\frac{\partial f}{\partial y} = 3y^2 \sin(xy) + (x^2 + y^3) x \cos(xy).$$

Solution to Exercise 7

$u_\theta = \cos\theta + \phi\sec^2\theta$ and $u_\phi = \tan\theta$. Hence, on recalling that $\cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}$, $\sec\frac{\pi}{4} = \sqrt{2}$ and $\tan\frac{\pi}{4} = 1$, we have

$$u_\theta\left(\tfrac{\pi}{4},\tfrac{\pi}{2}\right) = \tfrac{1}{\sqrt{2}} + \tfrac{\pi}{2}\times 2 = \tfrac{1}{\sqrt{2}} + \pi$$

and

$$u_\phi\left(\tfrac{\pi}{4},\tfrac{\pi}{2}\right) = 1.$$

Solution to Exercise 8

(a) Treating y and t as constants, and differentiating with respect to x,

$$f_x = 2xy^3t^4 + 2y + 8t^2x.$$

Treating x and t as constants, and differentiating with respect to y,

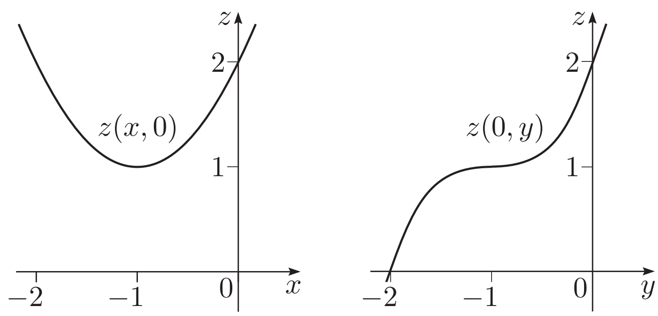
$$f_y = 3x^2y^2t^4 + 2x + 1.$$

Treating x and y as constants, and differentiating with respect to t,

$$f_t = 4x^2y^3t^3 + 8tx^2.$$

(b) We have $\dfrac{\partial z}{\partial x} = 2(1+x)$ and $\dfrac{\partial z}{\partial y} = 3(1+y)^2$.

Sketches of the section functions $z(x,0) = (1+x)^2 + 1$ and $z(0,y) = 1 + (1+y)^3$ are shown below.



We obtain $(\partial z/\partial x)(0,0) = 2$, which represents the slope of the left-hand graph where $x = 0$. Also, we obtain $(\partial z/\partial y)(0,0) = 3$, which represents the slope of the right-hand graph where $y = 0$.

Solution to Exercise 9

From Exercise 8(b),

$$\frac{\partial z}{\partial x} = 2(1+x), \quad \frac{\partial z}{\partial y} = 3(1+y)^2.$$

So we have

$$\frac{\partial z}{\partial x}(0,2) = 2, \quad \frac{\partial z}{\partial y}(0,2) = 27.$$

Hence

$$\delta z \simeq 2\,\delta x + 27\,\delta y.$$

Solution to Exercise 10

We have

$$\frac{\partial z}{\partial x} = y\cos x, \quad \frac{\partial z}{\partial y} = \sin x, \quad \frac{dx}{dt} = e^t, \quad \frac{dy}{dt} = 2t,$$

so

$$\begin{aligned}\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} &= (y\cos x)e^t + (\sin x)2t \\ &= t^2\cos(e^t)e^t + \sin(e^t)2t \\ &= t\left(te^t\cos(e^t) + 2\sin(e^t)\right).\end{aligned}$$

Thus

$$\frac{dz}{dt}(0) = 0.$$

Solution to Exercise 11

(a) We have

$$\nabla f(x,y) = (4xy + 3y^3)\mathbf{i} + (2x^2 + 9xy^2)\mathbf{j}.$$

(b) $\nabla f(2,1) = 11\mathbf{i} + 26\mathbf{j}$, so the slope at $(2,1,14)$ in the direction $\frac{3}{5}\mathbf{i} + \frac{4}{5}\mathbf{j}$ is

$$(11\mathbf{i} + 26\mathbf{j}) \cdot \left(\tfrac{3}{5}\mathbf{i} + \tfrac{4}{5}\mathbf{j}\right) = \tfrac{137}{5}.$$

Solution to Exercise 12

We need to evaluate the partial derivatives at $x = 2$ and $y = 1$. Since $\partial z/\partial x = x$ and $\partial z/\partial y = 2\sqrt{3}y$, equation (15) becomes $dz/dt = 2\cos\alpha + 2\sqrt{3}\sin\alpha$. We see that the slope dz/dt is a function of α only, but to make the argument clear, we replace the slope dz/dt by $s(\alpha)$, so

$$s(\alpha) = 2\cos\alpha + 2\sqrt{3}\sin\alpha.$$

We are required to find the maximum value of $s(\alpha)$ for varying values of α.

The stationary points of $s(\alpha)$ occur when

$$\frac{ds}{d\alpha} = -2\sin\alpha + 2\sqrt{3}\cos\alpha = 0,$$

which gives $\tan\alpha = \sqrt{3}$. This equation has two solutions in the interval $0 \le \alpha < 2\pi$, at $\alpha = \frac{\pi}{3}$ and $\alpha = \frac{4\pi}{3}$. Using the second derivative test, we see that $\alpha = \frac{\pi}{3}$ corresponds to a maximum value of $s(\alpha)$ (while $\alpha = \frac{4\pi}{3}$ gives a minimum value). It follows that the greatest slope at the point $(2, 1, 2+\sqrt{3})$ is

$$2\cos\tfrac{\pi}{3} + 2\sqrt{3}\sin\tfrac{\pi}{3} = 1 + (2\sqrt{3})\tfrac{\sqrt{3}}{2} = 4.$$

The direction of the greatest slope of $z = f(x,y)$ at $(2,1,2+\sqrt{3})$ is at an angle of $\frac{\pi}{3}$ measured anticlockwise from the positive x-axis.

Now $\nabla f(2,1) = 2\mathbf{i} + 2\sqrt{3}\mathbf{j}$, so the angle θ between $\nabla f(2,1)$ and the positive x-axis (measured anticlockwise) is also given by $\tan\theta = \sqrt{3}$. So since both components of $\nabla f(2,1)$ are positive, $\theta = \frac{\pi}{3}$. Thus the direction of greatest slope is the same as the direction of $\nabla f(2,1)$.

Solution to Exercise 13

(a) We have

$$f_x = 3(x+2y)^2 - 4(2x-y) \quad \text{and} \quad f_y = 6(x+2y)^2 + 2(2x-y),$$

so

$$\nabla f(x,y) = \big(3(x+2y)^2 - 4(2x-y)\big)\mathbf{i} + \big(6(x+2y)^2 + 2(2x-y)\big)\mathbf{j}.$$

(b) Note that

$$\nabla f(1,0) = -5\mathbf{i} + 10\mathbf{j}.$$

The vector $\mathbf{i}+\mathbf{j}$ has length $\sqrt{2}$, so $\widehat{\mathbf{d}} = \frac{1}{\sqrt{2}}(\mathbf{i}+\mathbf{j})$ is a unit vector in the required direction. At the point $(1,0,-3)$, we have

$$\nabla f \cdot \widehat{\mathbf{d}} = (-5\mathbf{i}+10\mathbf{j}) \cdot \tfrac{1}{\sqrt{2}}(\mathbf{i}+\mathbf{j}) = \tfrac{5}{\sqrt{2}}.$$

(c) Note that from equation (18), the greatest slope of the surface at $(1,0,-3)$ is given by $|\nabla f(1,0)|$, and that the vector $\nabla f(1,0)$ points along the direction of greatest slope. Hence the greatest slope is given by

$$|\nabla f(1,0)| = |-5\mathbf{i}+10\mathbf{j}| = \sqrt{5^2+10^2} = 5\sqrt{5},$$

and the unit vector $\widehat{\mathbf{n}}$ in the direction of $\nabla f(1,0)$ (direction of greatest slope) is given by

$$\widehat{\mathbf{n}} = \frac{\nabla f(1,0)}{|\nabla f(1,0)|} = \frac{1}{5\sqrt{5}}(-5\mathbf{i}+10\mathbf{j}) = \frac{1}{\sqrt{5}}(-\mathbf{i}+2\mathbf{j}).$$

(d) We know from the previous discussion that $\nabla f(1,0)$ points in the direction normal to the contour curve through the point $(1,0,-3)$ and is therefore perpendicular to a vector tangential to the contour at this point. If $\widehat{\mathbf{t}}$ denotes a tangential unit vector and $\widehat{\mathbf{n}}$ the unit vector along $\nabla f(1,0)$ as found in part (c), then we must have that $\widehat{\mathbf{n}} \cdot \widehat{\mathbf{t}} = 0$. Writing $\widehat{\mathbf{t}} = t_1\mathbf{i} + t_2\mathbf{j}$, the condition $\widehat{\mathbf{n}} \cdot \widehat{\mathbf{t}} = 0$ becomes

$$\frac{1}{\sqrt{5}}(-\mathbf{i}+2\mathbf{j}) \cdot (t_1\mathbf{i}+t_2\mathbf{j}) = 0,$$

thus $-t_1 + 2t_2 = 0$, so $t_1 = 2t_2$, therefore $\widehat{\mathbf{t}} = t_2(2\mathbf{i}+\mathbf{j})$. But $|\widehat{\mathbf{t}}| = 1$, therefore $t_2\sqrt{1+2^2} = 1$, so finally we have

$$\widehat{\mathbf{t}} = \frac{1}{\sqrt{5}}(2\mathbf{i}+\mathbf{j}).$$

Note that $-\widehat{\mathbf{t}}$ is also tangential to the contour through $(1,0,-3)$, so in this sense the tangential vector is not unique.

Solution to Exercise 14

$p_1(0) = 0$ and $f(0) = \sin 0 = 0$, thus $p_1(0) = f(0)$.

$p_1'(x) = 1$, so $p_1'(0) = 1$, and $f'(x) = \cos x$, so $f'(0) = 1$.
Thus $p_1'(0) = f'(0)$.

$p_1''(x) = 0$, so $p_1''(0) = 0$, and $f''(x) = -\sin x$, so $f''(0) = 0$.
Thus $p_1''(0) = f''(0)$.

$p_2(0) = 1 - 0 = 1$ and $g(0) = \cos 0 = 1$, thus $p_2(0) = g(0)$.

$p_2'(x) = -x$, so $p_2'(0) = 0$, and $g'(x) = -\sin x$, so $g'(0) = 0$.
Thus $p_2'(0) = g'(0)$.

$p_2''(x) = -1$, so $p_2''(0) = -1$, and $g''(x) = -\cos x$, so $g''(0) = -1$.
Thus $p_2''(0) = g''(0)$.

$p_2'''(x) = 0$, so $p_2'''(0) = 0$, and $g'''(x) = \sin x$, so $g'''(0) = 0$.
Thus $p_2'''(0) = g'''(0)$.

Solution to Exercise 15

From Example 7, we have

$$p_0(x) = 1, \quad p_1(x) = 1 + 2x, \quad p_2(x) = 1 + 2x + 2x^2.$$

Thus

$$p_0(0.1) = 1, \quad p_1(0.1) = 1.2, \quad p_2(0.1) = 1.22.$$

Also, $f(0.1) = e^{0.2} = 1.221\,40$, to five decimal places.

Thus

$p_0(0.1) = f(0.1)$ to the nearest integer,
$p_1(0.1) = f(0.1)$ to 1 d.p.,
$p_2(0.1) = f(0.1)$ to 2 d.p..

Solution to Exercise 16

$f_x = \sin(xy) + xy\cos(xy)$ and $f_y = x^2\cos(xy)$, so

$$\begin{aligned}
f_{xx} &= y\cos(xy) + y\cos(xy) - xy^2\sin(xy) = 2y\cos(xy) - xy^2\sin(xy),\\
f_{yy} &= -x^3\sin(xy),\\
f_{xy} &= 2x\cos(xy) - x^2y\sin(xy),\\
f_{yx} &= x\cos(xy) + x\cos(xy) - x^2y\sin(xy) = 2x\cos(xy) - x^2y\sin(xy).
\end{aligned}$$

Solution to Exercise 17

$f_x = 2e^{2x+3y}$ and $f_y = 3e^{2x+3y}$, so

$$f_{xx} = 4e^{2x+3y}, \quad f_{xy} = f_{yx} = 6e^{2x+3y}, \quad f_{yy} = 9e^{2x+3y}.$$

Then

$$\begin{aligned}
&f_x(0,0) = 2, \quad f_y(0,0) = 3,\\
&f_{xx}(0,0) = 4, \quad f_{yy}(0,0) = 9, \quad f_{xy}(0,0) = f_{yx}(0,0) = 6.
\end{aligned}$$

Solution to Exercise 18

$f_x(x,y) = 3(x^2+2y^2-3xy)^2(2x-3y)$ and
$f_y(x,y) = 3(x^2+2y^2-3xy)^2(4y-3x)$, so

$$\begin{aligned}
f_{xx}(x,y) &= 6(x^2+2y^2-3xy)(2x-3y)^2 + 6(x^2+2y^2-3xy)^2,\\
f_{xy}(x,y) &= 6(x^2+2y^2-3xy)(2x-3y)(4y-3x) - 9(x^2+2y^2-3xy)^2,\\
f_{yy}(x,y) &= 6(x^2+2y^2-3xy)(4y-3x)^2 + 12(x^2+2y^2-3xy)^2.
\end{aligned}$$

Then

$$\begin{aligned}
f_{xx}(1,-1) &= 6(1+2+3)(2+3)^2 + 6(1+2+3)^2 = 1116,\\
f_{xy}(1,-1) &= 6(1+2+3)(2+3)(-4-3) - 9(1+2+3)^2 = -1584,\\
f_{yy}(1,-1) &= 6(1+2+3)(-4-3)^2 + 12(1+2+3)^2 = 2196.
\end{aligned}$$

Solution to Exercise 19

$f_x = 3e^{3x-y}$ and $f_y = -e^{3x-y}$, so

$$f(0,0) = 1, \quad f_x(0,0) = 3, \quad f_y(0,0) = -1.$$

Thus the tangent approximation near $(0,0)$ is

$$\begin{aligned}
p_1(x,y) &= f(0,0) + f_x(0,0)x + f_y(0,0)y\\
&= 1 + 3x - y.
\end{aligned}$$

Solution to Exercise 20

(a) $q_x = \beta + 2Ax + By, \quad q_y = \gamma + Bx + 2Cy,$

$q_{xx} = 2A, \quad q_{xy} = B, \quad q_{yy} = 2C.$

(b) $q(0,0) = \alpha, \quad q_x(0,0) = \beta, \quad q_y(0,0) = \gamma,$

$q_{xx}(0,0) = 2A, \quad q_{xy}(0,0) = B, \quad q_{yy}(0,0) = 2C.$

Solution to Exercise 21

$f(0,0)$, $f_x(0,0)$ and $f_y(0,0)$ are constants, so the equation is that of a plane. Since $g(0,0) = f(0,0)$, the plane goes through $(0,0,f(0,0))$.

Now $g_x(x,y) = f_x(0,0)$ and $g_y(x,y) = f_y(0,0)$, so $g_x(0,0) = f_x(0,0)$ and $g_y(0,0) = f_y(0,0)$.

Thus at $(0,0)$, the plane has gradient

$$\begin{aligned}
\nabla g(0,0) &= g_x(0,0)\,\mathbf{i} + g_y(0,0)\,\mathbf{j}\\
&= f_x(0,0)\,\mathbf{i} + f_y(0,0)\,\mathbf{j}\\
&= \nabla f(0,0),
\end{aligned}$$

which is the gradient of $z = f(x,y)$ at $(0,0)$.

Solution to Exercise 22

We have

$$f(0,0) = 1, \quad f_x(0,0) = 2, \quad f_y(0,0) = 3,$$
$$f_{xx}(0,0) = 4, \quad f_{yy}(0,0) = 9, \quad f_{xy}(0,0) = 6.$$

Substituting these values into equation (25), we obtain

$$\begin{aligned} p_2(x,y) &= 1 + 2x + 3y + \tfrac{1}{2}(4x^2 + 12xy + 9y^2) \\ &= 1 + 2x + 3y + 2x^2 + 6xy + \tfrac{9}{2}y^2. \end{aligned}$$

Solution to Exercise 23

We have $U(0,0) = -2mga$, and

$$\frac{\partial U}{\partial \theta} = mga \sin\theta, \quad \frac{\partial U}{\partial \phi} = mga \sin\phi,$$

so

$$\frac{\partial U}{\partial \theta}(0,0) = \frac{\partial U}{\partial \phi}(0,0) = 0.$$

We also see that

$$\frac{\partial^2 U}{\partial \theta^2} = mga \cos\theta, \quad \frac{\partial^2 U}{\partial \phi^2} = mga \cos\phi, \quad \frac{\partial^2 U}{\partial \theta\, \partial \phi} = 0,$$

so

$$\frac{\partial^2 U}{\partial \theta^2}(0,0) = \frac{\partial^2 U}{\partial \phi^2}(0,0) = mga, \quad \frac{\partial^2 U}{\partial \theta\, \partial \phi}(0,0) = 0.$$

It follows from equation (25) that the second-order Taylor polynomial about $(0,0)$ is

$$p_2(\theta,\phi) = -2mga + \tfrac{1}{2}mga(\theta^2 + \phi^2).$$

Solution to Exercise 24

$f_x = ye^{xy} + 2(x+y)$ and $f_y = xe^{xy} + 2(x+y)$, so

$$f_{xx} = y^2e^{xy} + 2, \quad f_{xy} = e^{xy} + xye^{xy} + 2, \quad f_{yy} = x^2e^{xy} + 2.$$

Therefore

$$f(0,0) = 1, \quad f_x(0,0) = f_y(0,0) = 0,$$
$$f_{xx}(0,0) = 2, \quad f_{xy}(0,0) = 3, \quad f_{yy}(0,0) = 2.$$

Thus, from equation (25), the second-order Taylor polynomial about $(0,0)$ is

$$\begin{aligned} p_2(x,y) &= 1 + \tfrac{1}{2}\big(2x^2 + 2(3xy) + 2y^2\big) \\ &= 1 + x^2 + 3xy + y^2. \end{aligned}$$

Solution to Exercise 25

(a) $f_x = 3(2+x+2y)^2$ and $f_y = 6(2+x+2y)^2$, so

$$\begin{aligned} &f_{xx} = 6(2+x+2y), \quad f_{xy} = 12(2+x+2y),\\ &f_{yy} = 24(2+x+2y). \end{aligned}$$

It follows that

$$\begin{aligned} &f(0,0) = 8, \quad f_x(0,0) = 12, \quad f_y(0,0) = 24,\\ &f_{xx}(0,0) = 12, \quad f_{xy}(0,0) = 24, \quad f_{yy}(0,0) = 48. \end{aligned}$$

Therefore

$$\begin{aligned} p_1(x,y) &= f(0,0) + f_x(0,0)x + f_y(0,0)y\\ &= 8 + 12x + 24y \end{aligned}$$

and

$$\begin{aligned} p_2(x,y) &= f(0,0) + f_x(0,0)x + f_y(0,0)y\\ &\quad + \tfrac{1}{2}\big(f_{xx}(0,0)x^2 + 2f_{xy}(0,0)xy + f_{yy}(0,0)y^2\big)\\ &= 8 + 12x + 24y + 6x^2 + 24xy + 24y^2. \end{aligned}$$

Expanding the bracket in the expression for f, we obtain

$$\begin{aligned} f(x,y) = 8 &+ 12x + 24y + 6x^2 + 24xy + 24y^2 + x^3\\ &+ 6x^2y + 12xy^2 + 8y^3. \end{aligned}$$

Notice that the function $f(x,y)$ is itself a polynomial of degree 3, and that the Taylor polynomial of degree 2 about $(0,0)$ consists of the terms of f of degree less than 3.

(b) $f(1,-1) = 1, \quad f_x(1,-1) = 3, \quad f_y(1,-1) = 6,$

$f_{xx}(1,-1) = 6, \quad f_{xy}(1,-1) = 12, \quad f_{yy}(1,-1) = 24.$

It follows that

$$\begin{aligned} p_1(x,y) &= f(1,-1) + f_x(1,-1)(x-1) + f_y(1,-1)(y+1)\\ &= 1 + 3(x-1) + 6(y+1) \end{aligned}$$

and

$$\begin{aligned} p_2(x,y) &= f(1,-1) + f_x(1,-1)(x-1) + f_y(1,-1)(y+1)\\ &\quad + \tfrac{1}{2}\big(f_{xx}(1,-1)(x-1)^2 + 2f_{xy}(1,-1)(x-1)(y+1)\\ &\qquad + f_{yy}(1,-1)(y+1)^2\big)\\ &= 1 + 3(x-1) + 6(y+1) + 3(x-1)^2\\ &\quad + 12(x-1)(y+1) + 12(y+1)^2. \end{aligned}$$

(c) $F_X = 3(1+X+2Y)^2$ and $F_Y = 6(1+X+2Y)^2$, so

$$\begin{aligned} &F_{XX} = 6(1+X+2Y), \quad F_{XY} = 12(1+X+2Y),\\ &F_{YY} = 24(1+X+2Y). \end{aligned}$$

It follows that

$$\begin{aligned} &F(0,0) = 1, \quad F_X(0,0) = 3, \quad F_Y(0,0) = 6,\\ &F_{XX}(0,0) = 6, \quad F_{XY}(0,0) = 12, \quad F_{YY}(0,0) = 24. \end{aligned}$$

Therefore

$$p_1(X,Y) = 1 + 3X + 6Y$$

and

$$p_2(X,Y) = 1 + 3X + 6Y + 3X^2 + 12XY + 12Y^2.$$

(By substituting $X = x - 1$ and $Y = y + 1$, these polynomials are the same as those that you obtained in part (b). Finding the Taylor polynomials for f near $(1,-1)$ is equivalent to making a suitable change of variables and then calculating the Taylor polynomials near $(0,0)$. This may lead to simpler arithmetic, because evaluations of the partial derivatives at $(0,0)$ are often particularly easy. You can then change variables back again to obtain the polynomial in terms of the original variables.)

Solution to Exercise 26

$f_x(x,y) = 6x - 4y + 4$ and $f_y(x,y) = -4x + 4y - 8$, so there is one stationary point, at the solution of the simultaneous equations

$$\begin{aligned} 6x - 4y &= -4, \\ -4x + 4y &= 8, \end{aligned}$$

which is $x = 2$, $y = 4$. Thus the only stationary point is $(2,4)$.

Solution to Exercise 27

$f_x = 2xy + y^2 - 3y = y(2x + y - 3)$ and
$f_y = 2xy + x^2 - 3x = x(2y + x - 3)$, so to find the stationary points, we have to solve the simultaneous equations

$$\begin{aligned} y(2x + y - 3) &= 0, \\ x(2y + x - 3) &= 0. \end{aligned}$$

From the first equation, we have two cases.

If $y = 0$, then the second equation becomes $x(x - 3) = 0$; thus $x = 0$ or $x = 3$, so we have found the two solutions $(0,0)$ and $(3,0)$.

If $2x + y - 3 = 0$, that is, $y = 3 - 2x$, then the second equation becomes $x(3 - 3x) = 0$, and we have $x = 0$ or $x = 1$. Substituting these values for x into $y = 3 - 2x$ gives $y = 3$ when $x = 0$ and $y = 1$ when $x = 1$. Thus we have two more solutions, $(0,3)$ and $(1,1)$.

So we have four stationary points: $(0,0)$, $(3,0)$, $(0,3)$ and $(1,1)$.

Solution to Exercise 28

We have $U_\theta = mga\sin\theta$ and $U_\phi = mga\sin\phi$, so $U_\theta(0,0) = U_\phi(0,0) = 0$, which shows that $(0,0)$ is a stationary point. Also, $U(0,0) = -2mga$, and near $(0,0)$, $0 < \cos\theta \leq 1$ and $0 < \cos\phi \leq 1$, so $U(\theta,\phi) \geq -2mga$, and it follows that $(0,0)$ is a local minimum.

Solution to Exercise 29

We have

$$\mathbf{H}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} A & B \\ B & C \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} Ap + Bq \\ Bp + Cq \end{pmatrix},$$

so

$$\begin{aligned}(p \quad q)\,\mathbf{H}\begin{pmatrix} p \\ q \end{pmatrix} = (p \quad q)\begin{pmatrix} Ap + Bq \\ Bp + Cq \end{pmatrix} &= p(Ap + Bq) + q(Bp + Cq) \\ &= Ap^2 + 2Bpq + Cq^2 \\ &= Q(p, q).\end{aligned}$$

Solution to Exercise 30

$f_x = 4x - y - 3$ and $f_y = -x - 6y + 7$, so to find the stationary point, we need to solve the simultaneous equations

$$\begin{aligned} 4x - y - 3 &= 0, \\ -x - 6y + 7 &= 0. \end{aligned}$$

The solution is $x = 1$, $y = 1$, so the stationary point is at $(1, 1)$.

Also, $f_{xx} = 4$, $f_{xy} = -1$ and $f_{yy} = -6$, so $AC - B^2 = -25$; therefore the stationary point is a saddle point.

Note that this exercise is the same as Example 13, which was solved using the eigenvalue test.

Solution to Exercise 31

$f_x = 3x^2 - 12$ and $f_y = -3y^2 + 3$, so the stationary points are at the points (x, y) where $x^2 = 4$ and $y^2 = 1$, namely $(2, 1)$, $(2, -1)$, $(-2, 1)$, $(-2, -1)$.

Also, $f_{xx} = 6x$, $f_{xy} = 0$ and $f_{yy} = -6y$. Thus at $(2, 1)$ and $(-2, -1)$, we have $AC - B^2 = -36xy = -72$, and these are saddle points. At $(2, -1)$ and $(-2, 1)$, we have $AC - B^2 = -36xy = 72$. Since $A > 0$ at $(2, -1)$, this is a local minimum; since $A < 0$ at $(-2, 1)$, this is a local maximum.

Solution to Exercise 32

(a) Partially differentiating,

$$f_x = -x(1 - x^2 + y^2)^{-1/2}, \quad f_y = y(1 - x^2 + y^2)^{-1/2},$$

so the only stationary point is at $(0, 0)$. We also have

$$f_{xx} = -(1 - x^2 + y^2)^{-1/2} - x^2(1 - x^2 + y^2)^{-3/2},$$

so $A = f_{xx}(0, 0) = -1$. Since

$$f_{xy} = xy(1 - x^2 + y^2)^{-3/2},$$

we have $B = f_{xy}(0, 0) = 0$. Also,

$$f_{yy} = (1 - x^2 + y^2)^{-1/2} - y^2(1 - x^2 + y^2)^{-3/2},$$

so $C = f_{yy}(0, 0) = 1$.

So $AC - B^2 = -1 < 0$, and there is a saddle point at the origin.

(b) $T_x = -\sin x$ and $T_y = -\sin y$, so the stationary points occur when $\sin x = 0$ and $\sin y = 0$, that is, at the points $(n\pi, m\pi)$ where n and m are integers.

We also see that $T_{xx} = -\cos x$, $T_{xy} = 0$ and $T_{yy} = -\cos y$, so at the stationary point $(n\pi, m\pi)$ we have $A = -\cos n\pi$, $B = 0$ and $C = -\cos m\pi$. There are three cases to consider.

If m and n are both even, then $AC - B^2 = 1 > 0$ and $A = -1 < 0$, so there is a local maximum at $(n\pi, m\pi)$.

If m and n are both odd, then $AC - B^2 = 1 > 0$ and $A = 1 > 0$, so there is a local minimum at $(n\pi, m\pi)$.

Otherwise, $AC - B^2 = -1 < 0$ and there is a saddle point at $(n\pi, m\pi)$.

Solution to Exercise 33

Partially differentiating,

$$f_x = 6xy^3 - 18xy + 3x^2 - 15, \quad f_y = 9x^2y^2 - 9x^2 = 9x^2(y^2 - 1),$$

The stationary points must satisfy $f_y = 0$, which has solutions at $x = 0$ and $y = \pm 1$. They must also satisfy $f_x = 0$, but note that $x = 0$ can never solve $f_x = 0$. Substituting $y = 1$ into the equation $f_x = 0$ gives

$$3x^2 - 12x - 15 = 3(x^2 - 4x - 5) = 3(x + 1)(x - 5) = 0,$$

which has solutions $x = -1$ and $x = 5$. So there are stationary points at $(-1, 1)$ and $(5, 1)$. Substituting $y = -1$ into the equation $f_x = 0$ gives

$$3x^2 + 12x - 15 = 3(x^2 + 4x - 5) = 3(x - 1)(x + 5) = 0,$$

which has solutions $x = 1$ and $x = -5$. So there are two more stationary points at $(1, -1)$ and $(-5, -1)$.

To classify these stationary points, we need

$$\begin{aligned} f_{xx} &= 6y^3 - 18y + 6x = 6y(y^2 - 3) + 6x, \\ f_{xy} &= 18x(y^2 - 1), \\ f_{yy} &= 18x^2y. \end{aligned}$$

So for $(-1, 1)$ we have $A = f_{xx}(-1, 1) = -18$, $B = f_{xy}(-1, 1) = 0$ and $C = f_{yy}(-1, 1) = 18$. Since $AC - B^2 = -18^2 < 0$, there is a saddle point at $(-1, 1)$.

For $(5, 1)$ we have $A = f_{xx}(5, 1) = 18$, $B = f_{xy}(5, 1) = 0$ and $C = f_{yy}(5, 1) = 450$. Since $AC - B^2 = 18 \times 450 > 0$ and $A = 18 > 0$, there is a local minimum at $(5, 1)$.

For $(1, -1)$ we have $A = f_{xx}(1, -1) = 18$, $B = f_{xy}(1, -1) = 0$ and $C = f_{yy}(1, -1) = -18$. Since $AC - B^2 = -18^2 < 0$, there is a saddle point at $(1, -1)$.

Finally, for $(-5, -1)$ we have $A = f_{xx}(-5, -1) = -18$, $B = f_{xy}(-5, -1) = 0$ and $C = f_{yy}(-5, -1) = -450$. Since $AC - B^2 = 18 \times 450 > 0$ and $A = -18 < 0$, there is a local maximum at $(-5, -1)$.

Solution to Exercise 34

First we find the stationary point by solving the simultaneous equations

$$\begin{aligned} w_x &= 6x - 2y - 2z = 0, \\ w_y &= 6y - 2x - 2z = 0, \\ w_z &= 8z - 2y - 2x = 0. \end{aligned}$$

The only solution is $x = y = z = 0$, so the stationary point is at $(0, 0, 0)$.

Now $w_{xx} = 6$, $w_{yy} = 6$, $w_{zz} = 8$, $w_{xy} = -2$, $w_{xz} = -2$ and $w_{yz} = -2$, so the Hessian matrix is

$$\begin{pmatrix} w_{xx} & w_{xy} & w_{xz} \\ w_{yx} & w_{yy} & w_{yz} \\ w_{zx} & w_{zy} & w_{zz} \end{pmatrix} = \begin{pmatrix} 6 & -2 & -2 \\ -2 & 6 & -2 \\ -2 & -2 & 8 \end{pmatrix}.$$

The given characteristic equation of this matrix is

$$\begin{aligned} 0 &= (8 - \lambda)\left((6 - \lambda)^2 - 12\right) \\ &= (8 - \lambda)(6 - \lambda + 2\sqrt{3})(6 - \lambda - 2\sqrt{3}), \end{aligned}$$

so the eigenvalues are 8, $6 - 2\sqrt{3}$ and $6 + 2\sqrt{3}$. These are all positive, thus there is a local minimum at the origin.

Solution to Exercise 35

We have

$$\begin{aligned} &w_x = 2x + 2\sqrt{3}z, \quad w_y = 4y, \quad w_z = 2z + 2\sqrt{3}x, \\ &w_{xx} = 2, \quad w_{yy} = 4, \quad w_{zz} = 2, \\ &w_{xy} = 0, \quad w_{xz} = 2\sqrt{3}, \quad w_{yz} = 0. \end{aligned}$$

So the Hessian matrix is

$$\begin{pmatrix} 2 & 0 & 2\sqrt{3} \\ 0 & 4 & 0 \\ 2\sqrt{3} & 0 & 2 \end{pmatrix}.$$

The characteristic equation is $(4 - \lambda)(\lambda^2 - 4\lambda - 8) = 0$, obtained by expanding the relevant determinant by the middle row, giving the eigenvalues 4, $2 - 2\sqrt{3}$ and $2 + 2\sqrt{3}$. Two of these are positive and one is negative, thus the stationary point is a saddle point.