Partial Differentiation

1 Introduction

In the first part of this course you have met the idea of a derivative. To recap what this means, recall that if you have a function, z say, then the slope of the curve of z at a point t is said to be the number,

$$z'(t) = \lim_{h \to 0} \frac{z(t+h) - z(t)}{h},$$

provided that this limit exists. If this limit does exist for every value of t then the function z is said to be *differentiable*, or *smooth*.

If z is given by a power series in t, let's say

$$z(t) = a_0 + tz_1 + z_2 t^2 + ...,$$

then z is differentiable and its derivative at $t = 0$ is equal to z_1. This is easy to see and we know from Taylor's theorem that a partial converse to this result exists. Namely, if a function is smooth and we can differentiate it several times, then we can approximate the function locally in terms of a polynomial. This idea is very important during this part of the course.

Let us introduce some terminology that is useful when dealing with derivatives.

Definition 1.1 (The absolute value or modulus function) *The absolute value function $x \mapsto |x|$, which is read "x maps to mod x", is given by*

$$|x| = \left\{ \begin{array}{ccc} x & ; & x \geq 0 \\ -x & ; & x \leq 0, \end{array} \right.$$

or $|x| = +\sqrt{x^2}$.

Definition 1.2 (Big O notation) *We use the object*

$$O(h^2)$$

*to mean **any function** of h that contains terms which are as small as h^2, and possibly smaller, when h is itself small. For instance, $2h^2 + h^{101}$ is $O(h^2)$, as is $99h^2 - h^9$; it's just a shorthand where we don't care too much what the*

higher terms are because they're very small if h is small. In fact, a function $f(h)$ is said to be $O(h)$ as $h \to 0$ precisely when the limit

$$\lim_{h \to 0} \frac{f(h)}{h}$$

exists.

Now, we can think of the derivative as a way of approximating the function z using a linear function. For instance, one can also define the derivative of $z(t)$ to be the function $z'(t)$ such that the statement

$$|z(t + h) - z(t) - z'(t)h| = O(h^2)$$

holds as $h \to 0$ for all values of t.

This simply means that if h is small, then for a fixed t the **function**

$$h \mapsto z(t + h),$$

looks (locally at least) a lot like the **linear** function of h

$$h \mapsto z(t) + hz'(t).$$

To clarify these ideas let us consider a familiar example.

Example 1 *Let us consider the function $f(x) = x^2$. Now, we know that $f'(x) = 2x$. But we also have*

$$|f(x + h) - f(x) - f'(x)h| = |(x + h)^2 - x^2 - 2xh| = h^2.$$

Let us consider the function $f(x) = e^x$. Now, we know that $f'(x) = e^x$, but for a fixed x we also have

$$|f(x + h) - f(x) - f'(x)h| = |e^{x+h} - e^x - e^x h| = |e^x(e^h - 1 - h)|.$$

Since $e^h = 1 + h + \frac{1}{2}h^2 + \frac{1}{6}h^3 + ...$, we can justifiably write $|e^x(e^h - 1 - h)| = O(h^2)$ as $h \to 0$ with x fixed.

So, to sum up this preamble to this part of the course, a derivative of a function $z(t)$ is computed using the familiar idea of a limiting process which provides a linear approximation to $z(t)$. In fact, one could say that the

function (of h) $L(h) = z(t) + z'(t)h$ is the best linear approximation to the function $z(t)$ at t.

We could also define the quadratic function of h,

$$Q(h) = z(t) + z'(t)h + \frac{1}{2}z''(t)h^2,$$

and then we would find that $|z(t) - Q(t)| = O(h^3)$. For small h, something that is $O(h^3)$ is smaller in size than something that is $O(h^2)$, and this is the mathematical way in which one writes that Q is a better approximation of z near t than L is.

Example 2 *Find the linear and quadratic approximations to the function $z(t) = \cos(t)$ at the point $t = 0$.*

In this case $z(0) = 1, z'(0) = 0$ and $z''(0) = -1$, from where $L(h) = z(0) + hz'(0) = 1$ and $Q(h) = z(0) + hz'(0) + \frac{1}{2}z''(0)h^2 = 1 - \frac{1}{2}h^2$.

2 Partial Derivatives

Often functions depend on more than one variable. For instance, the volume of a box is given by

$$V(x, y, z) = xyz,$$

where x, y and z are the lengths of the sides of the box. Since calculus is so useful when studying problems in one variable, such as when maximising, curve sketching, or deriving differential equations in physics, we would like to see if there is a calculus for functions which depend on two variables.

We shall write this as

$$(x, y) \mapsto f(x, y),$$

or just $f(x, y)$ for short. The symbol $\mapsto$ is read "maps to" and indicates that f is a black box, with (x, y) as input, and some value $f(x, y)$ as output.

Now, a function such as

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

can have derivatives too. For instance, we can forget about y for a second and just think about the limit

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

Or we could try to evaluate the limit

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

We can perform both of these operations, and you should verify that the first one gives us

$$2x$$

and the second

$$3y^2.$$

But the first answer is exactly what you get when you take f, hold y as a constant, and just differentiate the function, thinking of it as a function only of x. Indeed, we could write

$$\frac{dy}{dx}(x, y) = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}.$$

Indeed we shall perform this process to differentiate functions of two variables, but we shall use a slightly different notation instead to remind us of the fact that there are several variables in our function. We actually use a 'curly d' and write

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

and also

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

We do, however, read the symbol ∂ as a 'd'. A more compact notation is also used for partial derivative and the symbols

$$f_x(x, y) \text{ and } f_y(x, y)$$

are used in place of $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$.

Example 3 *Given $f(x, y) = xy$, find $f_x(x, y)$ and $f_y(x, y)$.*
    Using the definition, we find

$$f_x(x, y) = \lim_{h \to 0} \frac{(x + h)y - xy}{h} = y,$$

and

$$f_y(x, y) = \lim_{k \to 0} \frac{x(y + k) - xy}{k} = x.$$

4

Given the fact that differentiation of a function of one variable provides a way of constructing linear and quadratic approximations to a function. Does partial differentiation provide analogies of such approximations for functions of more than one variable? The answer to this question will be affirmative, but we must first examine the notation of higher-derivatives for functions of more than one variable.

2.1   Higher and Mixed Partial Derivatives

Given that we can differentiate a function $f(x, y)$ with respect to one of the variables, x or y, to form a partial derivative, can we evaluate derivatives of derivatives, that is, how do we obtain higher derivatives?

We shall write the 'second partial derivative of f with respect to x' as

$$\frac{\partial^2 f}{\partial x^2}(x, y) = f_{xx}(x, y) = \lim_{h \to 0} \frac{f_x(x + h, y) - f_x(x, y)}{h},$$

and then do the same for y:

$$\frac{\partial^2 f}{\partial y^2}(x, y) = f_{yy}(x, y) = \lim_{k \to 0} \frac{f_y(x, y + k) - f_y(x, y)}{k}.$$

We also have the 'mixed' second partial derivatives

$$\frac{\partial^2 f}{\partial x \partial y}(x, y) = f_{xy}(x, y) = \lim_{h \to 0} \frac{f_y(x + h, y) - f_y(x, y)}{h}$$

and

$$\frac{\partial^2 f}{\partial y \partial x}(x, y) = f_{yx}(x, y) = \lim_{k \to 0} \frac{f_x(x, y + k) - f_x(x, y)}{k}.$$

You may think that the order in which we perform each subsequent differentiation when finding partial derivatives depends on the order in which each is performed. Put another way: is f_{xy} different from f_{yx}?

We shall label the answer to this question as a theorem.

Theorem 2.1 *For smooth function $f(x, y)$ the mixed partials f_{xy} and f_{yx} are the same.*

Notice that a natural corollary of this theorem is that the higher derivative $f_{xxyxxyxyx}$ is the same as $f_{xxxxxxyyy}$ as the order of differentiation is of no importance!

Example 4 *Consider the function* $f(x, y) = x^2 y + ye^x$, *we then have the following derivatives:*

$$
\begin{array}{llll}
f_x = 2xy + ye^x & f_y = x^2 + e^x & & \\
f_{xx} = 2y + ye^x & f_{yy} = 0 & f_{xy} = 2x + e^x & f_{yx} = 2x + e^x \\
f_{xxx} = ye^x & f_{xxy} = 2 + e^x & f_{xyy} = 0 & f_{yyy} = 0.
\end{array}
$$

and so on!

2.2   2-D Taylor's Theorem

Definition 2.1 *Given a function* $f(x, y)$ *where* $f : \mathbb{R}^2 \to \mathbb{R}$, *let us define the two-dimensional vector*

$$
\nabla f(x, y) = (f_x(x, y), f_y(x, y)).
$$

The operation ∇ *which takes a function and gives us its partial derivatives is read 'grad'.*

Sometimes, the symbol ∇ is manipulated as an object in its own right, and you may see written

$$
\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right).
$$

This symbol can be manipulated as a two-dimensional vector, for instance we can take the scalar product to give a new operation:

$$
\nabla \bullet \nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right) \bullet \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right) = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.
$$

In other words, this means that

$$
(\nabla \bullet \nabla) f = f_{xx} + f_{yy}.
$$

The operation $\nabla \bullet \nabla$ arises in many different scientific contexts and is called the Laplacian operation.

Also, we may define the *Hessian* matrix of f by

$$
\nabla^2 f = \left[\begin{array}{cc} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{array} \right].
$$

Since $f_{xy} = f_{yx}$, the Hessian matrix of f satisfies the symmetry property

$$(\nabla^2 f)^T = \nabla^2 f.$$

Using this notation, we now proceed to derive the Taylor expansion for functions of two variables. So, consider

$$f(x + h, y + k),$$

where h and k are thought of as small and think of $y + k$ as fixed for the moment. Using what we know from the usual Taylor's theorem for functions of one variable, we can think of having the approximation

$$f(x + h, y + k) = f(x, y + k) + \frac{\partial f}{\partial x}(x, y + k)h + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x, y + k)h^2 + ...$$

But also we have the expanions

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

and

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

Adding these together we have

$$\begin{aligned} f(x + h, y + k) \;=\; & f(x, y) + \frac{\partial f}{\partial x}(x, y)h + \frac{\partial f}{\partial y}(x, y)k + \\ & \frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}(x, y)h^2 + 2\frac{\partial^2 f}{\partial x \partial y}(x, y)hk + \frac{\partial^2 f}{\partial y^2}(x, y)k^2\right) + ... \quad . \end{aligned}$$

While this may look a little cumbersome, we can use vector notation we defined above in order to spruce this expression up a litte. We can now write (1) as the rather elegant expression

$$f(\mathbf{x} + \mathbf{u}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \bullet \mathbf{u} + \frac{1}{2}\mathbf{u}^T \nabla^2 f(\mathbf{x})\mathbf{u} + ... \quad ,$$

where $\mathbf{x} = (x, y)$ and $\mathbf{u} = (h, k)$.

Definition 2.2 (Curly O notation) *We shall use the symbol $\mathcal{O}(n)$, where $n = 1, 2, 3, ...$ to denote the following types of functions of two variables.*

$$
\begin{array}{rcl}
\mathcal{O}(1) & : & x, y \\
\mathcal{O}(2) & : & x^2, xy, y^2 \\
\mathcal{O}(3) & : & x^3, x^2 y, yx^2, y^3 \\
\mathcal{O}(4) & : & x^4, x^3 y, x^2 y^2, yx^3, y^4
\end{array}
$$

In other words, $\mathcal{O}$ is used to denote the **order** *of the terms in the function. In general, if a function contains terms of mixed order, for instance*

$$ f(x, y) = x^2 y + x^2 y^2, $$

we say that f is $\mathcal{O}(3)$, as the lowest terms that appear in f are of order three.

Example 5 *The first few terms of the Taylor series for the function $f(x, y) = e^{x+y}$, where x and y are near 0, are found as follows. We have $f_x = f_y = f_{xx} = f_{xy} = f_{yy} = e^{x+y}$. Hence*

$$ e^{x+y} = 1 + x + y + \frac{1}{2}(x^2 + 2xy + y^2) + ... $$

Note that $e^{x+y} = 1 + x + y + \frac{1}{2}(x+y)^2 + ...$ from the definition of the exponential function, which of course agrees with our Taylor series.

Example 6 *The first few terms of the Taylor series for the function $f(x, y) = x\cos(y)$, where x and y are near 0, are found as follows. We have $f_x = \cos(y)$, $f_y = -x\sin(y)$, $f_{xx} = 0$, $f_{xy} = -\sin(y)$, $f_{yy} = -x\cos(y)$. Hence*
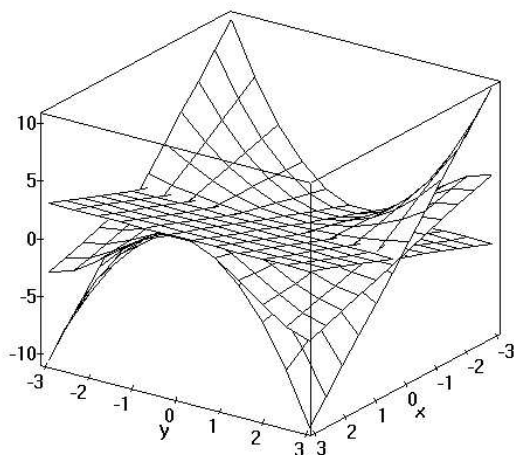
$$ x\cos(y) = x + \mathcal{O}(3). $$

This makes sense because we know from the one-dimensional Taylor theorem that

$$ x\cos(y) = x(1 - \frac{1}{2}y^2 + O(y^4)). $$

2.3   Tangent Plane Approximation

Given a function $f(x, y)$, which we may also write $f(\mathbf{x})$ using vector notation, the *tangent plane* or *linear* approximation of f at $\mathbf{x} = \mathbf{a}$ is defined as follows.

Figure 1: Tangent plane and quadratic approximation to $x\cos(y)$ near $(x, y) = (0, 0)$.

Definition 2.3 (Tangent plane approximation) *The tangent plane approximation of the function $f : \mathbb{R}^2 \to \mathbb{R}$ is the function $T(\mathbf{x})$, defined by*

$$T(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a}).$$

The idea behind the definition of T is that, just as we saw for functions of one variable,

$$\|f(\mathbf{x}) - T(\mathbf{x})\| = \|f(\mathbf{x}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a})\| = O(\|(\mathbf{x} - \mathbf{a})\|^2).$$

Hence $\|f(\mathbf{x}) - T(\mathbf{x})\|$ is a very small quantity when $\|\mathbf{x} - \mathbf{a}\|$ is small; that is, $\mathbf{x}$ is near to $\mathbf{a}$.
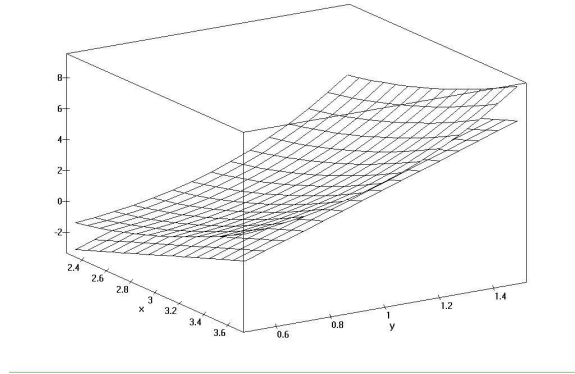
Let us also remark that we know from our work on vectors that the range of the function T is a certain plane. In terms of coordinates we may write the function T as

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

where $(a, b) = \mathbf{a}$. Let us also note that T has the properties that it not only agrees with f at the point $\mathbf{a}$,

$$T(\mathbf{a}) = f(\mathbf{a})$$

9

Figure 2: Tangent plane approximation to $f(x, y) = x^2 + 2y^3 - 3x$ near $(x, y) = (3, 1)$.

but the derivatives of T and f agree here too:

$$\nabla T(\mathbf{a}) = \nabla f(\mathbf{a}).$$

Example 7 *Find the tangent plane approximation to the function*

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

at $(x, y) = (3, 1)$.

Using the defining formulae, we find

$$T(x, y) = f(3, 1) + f_x(3, 1)(x - 3) + f_y(3, 1)(y - 1),$$

or

$$T(x, y) = 2 + 3(x - 3) + 6(y - 1) = -13 + 3x + 6y.$$

is the required function. See figure 2.

2.4 Quadratic approximation

Again, as in the one-variable case, we can approximate a function more closely by including higher order terms from the Taylor expanion. We shall call the following function, Q, defined by

$$Q(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a}) \bullet \nabla^2 f(\mathbf{a})\,(\mathbf{x} - \mathbf{a})$$

the quadratic approximation to f at $\mathbf{x} = \mathbf{a}$. This is really just a 'chopped' Taylor series, and in coordinates we can write

$$
\begin{aligned}
Q(x,y) \;=\; & f(a,b) + f_x(a,b)(x - a) + f_y(a,b)(y - b) \\
& + \frac{1}{2}\left(f_{xx}(a,b)(x - a)^2 + 2f_{xy}(a,b)(x - a)(y - b) + f_{yy}(a,b)(y - b)^2 \right).
\end{aligned}
$$

Let us also note that Q has the properties that

$$
Q(\mathbf{a}) = f(\mathbf{a}),\ \nabla Q(\mathbf{a}) = \nabla f(\mathbf{a}),
$$

and

$$
\nabla^2 Q(\mathbf{a}) = \nabla f^2(\mathbf{a}).
$$

In other words, T agrees with f near $\mathbf{a}$ up to the first-order derivatives, but Q agrees with f up to the second-order derivatives.

Example 8 *Find the quadratic approximation to the function $f(x,y) = x + 2y + xy + 9y^3$ about the point $(x,y) = (0,0)$.*

In this case, at an intuitive level we can simply 'throw away' those terms which are higher order than the quadratic terms x^2, xy and y^2. This leaves

$$
Q(x,y) = x + 2y + xy.
$$

*You can perform the details to show that this is indeed the required function.*

Example 9 *Find the quadratic approximation to the function $f(x,y) = x + 2y + xy + 9y^3$ about the point $(x,y) = (1,1)$.*

The function we are seeking is

$$
\begin{aligned}
Q(x,y) \;=\; & f(1,1) + f_x(1,1)(x - 1) + f_y(1,1)(y - 1) + \\
& \frac{1}{2}(f_{xx}(1,1)(x - 1)^2 + 2f_{xy}(1,1)(x - 1)(y - 1) + f_{yy}(1,1)(y - 1)^2) \\
\;=\; & 13 + 2(x - 1) + 30(y - 1) + \frac{1}{2}\left((x - 1)(y - 1) + 54(y - 1)^2 \right).
\end{aligned}
$$

See figure 3 which shows good agreement betweem f and Q locally.

Figure 3: Quadratic approximation to $f(x, y) = x + 2y + xy + 9y^3$ near $(x, y) = (1, 1)$.
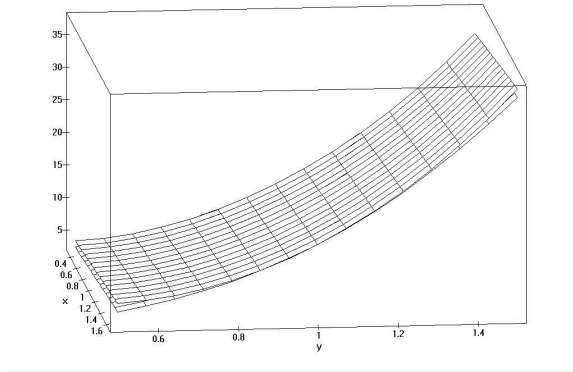
3   Stationary Points

Recall from one-dimensional calculus that a local extremum of the the function $f(x)$ occurs when

$$\frac{df}{dx}(x) = 0.$$

In two dimensions we have an analogy of this.

Definition 3.1 *We say that there is a* local extremum *or* stationary point *of $f(x, y)$ when the* **two equations**

$$\frac{\partial f}{\partial x}(x, y) = \frac{\partial f}{\partial y}(x, y) = 0$$

are satisfied at some point (x, y).

Example 10 *For instance, the function $f(x, y) = x^2 + y^2$ has a local extremum at $(x, y) = (0, 0)$. So too does $f(x, y) = 1 - x^2 - y^2$ and $f(x, y) = xy + y^2 x^2$.*

In calculus we recall classifying stationary points according to maxima, minima and points of inflection according to the derivatives of the function at that point. In two-dimensional calculus, the same question arises of how we

know when a particular local extremum can be classed as a local minimum, local maximum or neither of these.

Now, let a, b and c be parameters and consider the function

$$f(x, y) = \frac{1}{2}(ax^2 + 2bxy + cy^2).$$

It is simple to show that f has a local extremum at $x = y = 0$ and it is a natural question to ask what a plot of the surface

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

looks like for small x and y. This will depend on the parameters, but by completing the square we can see exactly what this dependence is.

So, suppose that

$$a \neq 0.$$

Then, noticing the presence of the 2 in front of f we have

$$
\begin{aligned}
2f(x, y) &= a(x^2 + \frac{2b}{a}xy) + cy^2 \\
&= a\left((x + \frac{b}{a}y)^2 - \frac{b^2}{a^2}y^2\right) + cy^2 \\
&= a(x + \frac{b}{a}y)^2 + \left(-\frac{b^2}{a} + c\right)y^2 \\
&= a(x + \frac{b}{a}y)^2 + \frac{1}{a}\left(-b^2 + ac\right)y^2.
\end{aligned}
$$

From this we can deduce the following information upon setting $\Delta = ac - b^2$.

- If $a > 0$ and $\Delta > 0$ then $f(x, y) > 0$ for all x and y near 0. Then $(0, 0)$ is called a local minimum of f.

- If $a < 0$ and $\Delta > 0$ then $f(x, y) < 0$ for all x and y near 0. Then $(0, 0)$ is called a local maximum of f.

- If $\Delta < 0$ then there are two lines in the $x - y$ plane such that $f(x, y) = 0$ if (x, y) lies on either of these lines near 0.

One can see this by writing $\Delta = -\mu^2 < 0$ and then $f(x, y) = 0$ when

$$a(x + \frac{b}{a}y)^2 + \frac{\Delta}{a}y^2 = 0$$

13

and this is equivalent to

$$a^2(x + \frac{b}{a}y)^2 = \mu^2 y^2,$$

which, by taking square roots, is the **pair** of equations

$$ax + by = \pm\mu y.$$

This can be solved for x as two different functions of y along which f evaluates to zero. In this situation $(x, y) = (0, 0)$ is said to be a saddle-point for f.

Example 11 *As an exercise, show that if $a = 0$ and $b \neq 0$, then there are also two lines on which $f = 0$ and therefore the point $(0, 0)$ is a saddle. In this case $\Delta = -b^2 < 0$ is automatically satisfied.*

The point of all this is that if $g(x, y)$ is some function such that

$$g_x(u, v) = g_y(u, v) = 0,$$

for some u and v, then we can write the local quadratic approximation of g near (u, v) in the form

$$
\begin{aligned}
Q(x, y) &= g(u, v) + g_x(u, v)(x - a) + g_y(u, v)(y - b) \\
&\quad + \frac{1}{2}\left(g_{xx}(u, v)(x - a)^2 + 2g_{xy}(u, v)(x - a)(y - b) + g_{yy}(u, v))(y - b)^2\right) \\
&= g(u, v) + \frac{1}{2}\left(g_{xx}(u, v)(x - a)^2 + 2g_{xy}(u, v)(x - a)(y - b) + g_{yy}(u, v)(y - b)^2\right).
\end{aligned}
$$

If we define

$$a := g_{xx}(u, v), b := g_{xy}(u, v) \text{ and } c := g_{yy}(u, v)$$

then

$$Q(x, y) = g(u, v) + f(x, y),$$

where $f(x, y) = \frac{1}{2}(ax^2 + 2bxy + cy^2)$, as above. Then, for (x, y) near to (u, v), the condition $f(x, y) > 0$ is equivalent to $g(x, y) > g(u, v)$, and we can distinguish various cases depending on the behaviour of a, b and c. For us to be able to make use of the above analysis we also define the discriminant

$$\Delta = ac - b^2 = g_{xx}(u, v)g_{yy}(u, v) - g_{xy}(u, v)^2,$$

14

and we also require that $a = g_{xx}(u, v) \neq 0$.

We can now determine the local behaviour of $g(x, y)$ from the knowledge of the local behaviour of its quadratic, or second-order, approximation that we denoted by f above. To do this, one simply ignores the terms that are higher than quadratic order and analyses the local nature of the quadratic terms. To clarify, let us consider some examples.

Example 12 *Classify the local extrema of the function*

$$g(x, y) = 3xy - 6x - 3y + 7.$$

Clearly $g_x = 3y - 6$ and $g_y = 3x - 3$. A local extremum therefore occurs where $(x, y) = (1, 2)$. Now $g_{xx} = 0$ and $g_{yy} = 0$ and the above analysis gives us no information concerning the nature of the local extremum.

Example 13 *Classify the local extrema of the function*

$$g(x, y) = 4xy - x^4 - y^4.$$

Clearly $g_x = 4y - 4x^3$ and $g_y = 4x - 4y^3$. A local extremum therefore occurs where

$$y = x^3 \text{ and } x = y^3.$$

Therefore $y = (y^3)^3 = y^9$ and therefore $y(y^8 - 1) = 0$. Hence

$$(x, y) = (0, 0), (1, 1) \text{ or } (-1, -1).$$

Now $g_{xx} = -12x^2, g_{xy} = 4$ and $g_{yy} = -12y^2$ and the above analysis gives us the following information.

1. *At $(x, y) = (0, 0)$ we have*

$$\Delta = g_{xx}g_{yy} - g_{xy}^2 = -16$$

   which is negative and $g_{xx} = 0$ here. This tells us that $(0, 0)$ is a saddle point.

2. *At $(x, y) = (1, 1)$ we have*

$$\Delta = g_{xx}g_{yy} - g_{xy}^2 = 128 > 0$$

   and $g_{xx} = -12 < 0$ so that $(1, 1)$ is a local maximum.

3. *At $(x, y) = (-1, -1)$ we have*

$$\Delta = g_{xx}g_{yy} - g_{xy}^2 = 128 > 0$$

   and $g_{xx} = -12 < 0$ too, so that $(1, 1)$ is also a local maximum.

3.1 Plotting Functions of the form $z = f(x, y)$

We consider two ways of getting information concerning the graph of a function $z = f(x, y)$. The first method is to choose various values, which we label c, and to plot solutions in the $x - y$ plane to the equation

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

The solution curves in the plane that you plot are called the *contours* of the function f and the value of c represents the height of the contour on the surface that we are trying to visualise.

Example 14 *The contours of the surface $z = x^2 + y^2$ are concentric circles and the surface is called a ''parabolic bowl''.*

Example 15 *Draw some contours of the function $f(x, y) = x^2 - 2y^2$. For instance, we could solve the five equations*

$$x^2 - 2y^2 = 0, \pm 1, \pm 2.$$

This solution curves are then obtained from

$$x = \pm \sqrt{2y^2 + c},$$

where c is our constant.

Another method for trying to visualise the surface $z = f(x, y)$ is to plot its sections. These are obtained by fixing $x = c$ and plotting the resulting function of y:

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

One could also take the section $y = c$ and plot the function of x:

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

Example 16 *The sections of the surface $z = x^2 + y^2$ are given by the shifted parabolas $z = x^2 + c^2$ and $z = c^2 + y^2$.*

Example 17 *Sections of the function $f(x, y) = x^2 - 2y^2$ are given by graphs of the function $z = x^2 - 2c^2$ and these are shifted parabolas too.*

It can be quite difficult to obtain a picture in your mind of the function you are given. For this reason, it is often useful to use the plotting routines in a package such as MAPLE or MATHEMATICA.

4   Chain Rule

Suppose that we have the equation of a surface

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

and we want to obtain a calculus for a particle moving along that surface. We can then think of (x, y) as a function of time, t, and form the height function

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

A natural question to ask is 'what is the derivative of z with respect to time?'. This derivative measures the rate at which the height is changing as time changes. We can find this expression by thinking about the difference

$$z(t + h) - z(t)$$

which is

$$f(x(t + h), y(t + h)) - f(x(t), y(t)).$$

With x and y smooth functions, we can write

$$x(t + h) = x(t) + hx'(t) + O(h^2),$$

and

$$y(t + h) = y(t) + hy'(t) + O(h^2).$$

But we may also use vector notation to write

$$f(\mathbf{x} + \mathbf{u}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \bullet \mathbf{u} + O(\|\mathbf{u}\|^2).$$

If we form the vector

$$\mathbf{u} = h(x'(t), y'(t)) + O(h^2)$$

which depends on h, then

$$
\begin{aligned}
z(t + h) - z(t) &= f(x(t + h), y(t + h)) - f(x(t), y(t)) \\
&= f((x(t), y(t)) + h(x'(t), y'(t)) + O(h^2)) - f(x(t), y(t)) \\
&= f(x(t), y(t)) + \nabla f(x(t), y(t)) \bullet (h(x'(t), y'(t))) + O(h^2) - f(x(t), y(t)), \\
&= h\nabla f(x(t), y(t)) \bullet (x'(t), y'(t)) + O(h^2).
\end{aligned}
$$

Therefore

$$\lim_{h\to 0}\frac{z(t+h)-z(t)}{h} = \nabla f(x(t),y(t)) \bullet (x'(t),y'(t))$$
$$= \frac{\partial f}{\partial x}(x(t),y(t))\frac{dx}{dt}(t) + \frac{\partial f}{\partial y}(x(t),y(t))\frac{dy}{dt}(t),$$

or if we want to be a little less pedantic then we can write

$$z'(t) = \lim_{h\to 0}\frac{z(t+h)-z(t)}{h} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Now, we can extend this idea by supposing that x and y are now functions of (t,s), and we now form

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

We can think of s being fixed and repeat the above process, replacing derivatives of x and y with partial derivatives. This would give us the result that

$$\frac{\partial z}{\partial t} = \lim_{h\to 0}\frac{z(t+h,s)-z(t,s)}{h} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t},$$

and, reasoning in an analogous fashion,

$$\frac{\partial z}{\partial s} = \lim_{k\to 0}\frac{z(t,s+k)-z(t,s)}{k} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}.$$

These expressions are known as the *chain rule* for functions of two variables. If we recall the chain rule from one-dimensional calculus, that

$$\frac{d}{dt}f(x(t)) = \frac{df}{dx}\frac{dx}{dt},$$

then we see that and extra term is required when we move to two dimensions. Explicitly,

$$\frac{d}{dt}f(x(t),y(t)) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

And again, just for clarity

$$\frac{\partial}{\partial t}f(x(t,s),y(t,s)) = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}.$$

18

Example 18 *Let $f(x, y) = \cos(xy)$ and let $x(t) = 2t$, $y(t) = t^2$, let us find $\frac{d}{dt}f(x(t), y(t))$ using the chain rule. Now*

$$\frac{d}{dt}f(x(t), y(t)) = f_x(x(t), y(t))x'(t) + f_y(x(t), y(t))y'(t)$$

which equals

$$-\sin(xy)yx'(t) - \sin(xy)xy'(t) = -\sin(2t^3)(2t^2 + 4t^2) = -\sin(2t^3)\ 6t^2.$$

Also $\frac{d}{dt}\cos(2t^3) = -\sin(2t^3)\ 6t^2$.

Example 19 *Let us use the chain rule to evalute the first partial derivatives of $h(x, y) = f(g(x, y))$ where*

$$f(x, y) = xy$$

and

$$g(x, y) = (x + y, x - y) =: (g_1(x, y), g_2(x, y)).$$

Let us remark that

$$h(x, y) = f(g(x, y)) = f(g_1(x, y), g_2(x, y)) = f(x+y, x-y) = (x+y)(x-y) = x^2 - y^2.$$

Hence we find that $h_x = 2x$ and $h_y = 2y$. Using the chain rule we have

$$h_x = f_x(g_1, g_2)g_{1_x} + f_y(g_1, g_2)g_{2_x} = g_2 \times 1 + g_1 \times 1 = g_2 + g_1 = 2x$$

and

$$h_y = f_x(g_1, g_2)g_{1_y} + f_y(g_1, g_2)g_{2_y} = g_2 \times 1 + g_y \times -1 = g_2 - g_1 = 2y.$$

19