ADVANCED CALCULUS

MA1132

HILARY TERM 2017/2018

DR. ANTHONY BROWN

Contents

2. Partial Derivatives

In the first chapter, we generalized some of the results of the calculus you would have been familiar with before the course by studying function which had a vector space (usually $\mathbb{R}^2$ or $\mathbb{R}^3$), rather than $\mathbb{R}$ as their codomain.

In this chapter we will have the codomain as $\mathbb{R}$ but this time change the domain to $\mathbb{R}^2$ or $\mathbb{R}^3$, and a lot of this work can easily be generalized to $\mathbb{R}^n$.

2.1. Functions of Two or More Variables.

Let us start with a definition.

Definition 2.1.1. A *function of two variables* is a function which has as its domain a subset of $\mathbb{R}^2$.

Remark 2.1.2.

(a) We can similarly define functions of three variables, or indeed, functions of n variables. Sometimes we refer to all such functions as functions of several variables, or functions of many variables.
(b) A common way to denote a function of two variables is $f(x, y)$, where (x, y) is the point in $\mathbb{R}^2$ that the function is acting on. Similarly, we denote functions of three variables by $f(x, y, z)$, functions of four variables by $f(x, y, z, w)$ and functions of n variables by $f(x_1, x_2, \ldots, x_n)$.
(c) As was the the case with vector-valued functions, if the domain of the function of two or more variables is not given, then it will be understood to be the largest one that 'makes sense'. Since we now have to deal with more than one variable, it can be trickier to find the domain and get a handle on what it looks like.

Warning 2.1.3. It is not the case that functions of several variables have more than one domain. The domain of such functions is a set of points in $\mathbb{R}^n$, so a single element of the domain is a point $(x_1, x_2, \ldots, x_n)$. In particular the numbers $x_1, x_2, \ldots, x_n$ are not elements of the domain.

When we were studying vector-valued functions, we noted that the 'graph' we studied there was not really a graph. For functions of two variables, we return to the accepted definition of a graph. For convenience, I will repeat it here.

Definition 2.1.4. Given a function

$$\begin{aligned} f\colon A &\to B \\ x &\mapsto f(x), \end{aligned}$$

then its *graph* is defined to be the set $\{(x, f(x))\colon x \in A\}$.

However, for functions of two variables, the domain lies in a subset of $\mathbb{R}^2$, so that instead of x in Definition 2.1.4, we have (x, y). Thus in this case the graph is $\{((x, y), f(x, y))\colon x \in A\}$. Note that $((x, y), f(x, y))$ may also be regarded as $(x, y, f(x, y))$, a point in $\mathbb{R}^3$. Thus the graph of a function of two variables will be a set of points in $\mathbb{R}^3$, and for the sort of functions we will be studying in this course, it will usually be a two dimensional surface lying in $\mathbb{R}^3$. Unfortunately, this approach is not much use for functions of three or more variables, since when we add the dimension for the codomain, we end up with 'surfaces' in four or more dimensions.

There is another method that is commonly used to get an idea about what the graphs of functions of two and three variables 'look like'. The idea is similar to the idea of contour lines on maps.

Definition 2.1.5. A *level set* of a function of n variables with domain A is a set of the form $\{(x_1, x_2, \ldots, x_n) \in A\colon f(x_1, x_2, \ldots, x_n) = c\}$, where c is a constant.

Remark 2.1.6.

(a) If we are dealing with a function of two variables, then its level sets will be one dimensional lines lying in $\mathbb{R}^2$, and are often called level curves. If we are dealing with a function of three variables then its level sets will be two dimensional surfaces lying in $\mathbb{R}^3$, and are often called level surfaces. Thus level sets give us an additional way of visualizing functions of two variables and a way of visualizing functions of three variables. They are no use for visualizing functions of four and higher variables, but still may be useful algebraically in such cases.
(b) When dealing with functions of two variables, then a useful technique can be to plot several level curves for various values of the constant c. This allows us to visualize the graph of the function in the same way that we can visualize the shape of a mountain from contour lines on a map.

2.2. **Limits and Continuity.**

As was the case when we were dealing with vector-valued functions, before we can study the continuity and differentiability of functions of several variables, then we first need to study limits, since continuity and differentiability are defined in terms of these.

This is somewhat more complicated for functions of several variables than it was for vector-valued functions. With vector-valued functions, it turned out that we could reduce the problem of taking a limit to the problem of taking several limits of functions that map $\mathbb{R}$ to $\mathbb{R}$, something we already knew how to do. Unfortunately this is not the case with functions of several variables, so we need to develop some new techniques.

The essential problem is as follows: If we are dealing with a function with domain a subset of $\mathbb{R}$ and investigating a limit as x tends to a, then there are only two directions that x can approach a, from the left and from the right. On the other hand, if the domain is a subset of $\mathbb{R}^n$ with $n > 1$ then there are infinitely many directions in which $(x_1, x_2, \ldots, x_n)$ can approach $(a_1, a_2, \ldots, a_n)$. This is not an analysis course, so our focus will be on finding limits if they exist. However, we still will give the formal definition of a limit.

Definition 2.2.1. Suppose that a function f of n variables is defined on a punctured open ball

$$\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \colon \|(x_1, x_2, \ldots, x_n) - (a_1, a_2, \ldots, a_n)\| < r\} \backslash \{(a_1, a_2, \ldots, a_n)\},$$

then we say that

$$\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f(x_1, x_2, \ldots, x_n) = L,$$

if for each $\epsilon > 0$, we can find a $\delta > 0$ such that $|f(x_1, x_2, \ldots, x_n) - L| < \epsilon$ for all $(x_1, x_2, \ldots, x_n)$ that satisfy $0 < \|(x_1, x_2, \ldots, x_n) - (a_1, a_2, \ldots, a_n)\| < \delta$.

Remark 2.2.2.

(a) We don't insist that the function be defined at $(a_1, a_2, \ldots, a_n)$, and this is why we need the condition $0 < \|(x_1, x_2, \ldots, x_n) - (a_1, a_2, \ldots, a_n)\|$.
(b) We could also write the definition more compactly by replacing $(x_1, x_2, \ldots, x_n)$ with $\mathbf{x}$ and $(a_1, a_2, \ldots, a_n)$ with $\mathbf{a}$.
(c) The definition is rather similar to the definition you will be familiar with for functions that map subsets of $\mathbb{R}$ to $\mathbb{R}$. We have just replaced x with $(x_1, x_2, \ldots, x_n)$, a with $(a_1, a_2, \ldots, a_n)$, and have used $\| \cdot \|$ rather than $| \cdot |$.
(d) It is also possible to define limits at boundary points of the domain of a function of several variables, but we won't do this in MA1132.

As with the real case, Definition 2.2.1 is not what we will use to determine if limits exist, or to find them if they do. The following theorem can be useful, especially if we want to show a limit doesn't exist.

Theorem 2.2.3. *The limit* $\displaystyle\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f(x_1, x_2, \ldots, x_n)$ *exists and equals* L *if and only if* $f(x_1, x_2, \ldots, x_n) \to L$ *as* $(a_1, a_2, \ldots, a_n) \to (x_1, x_2, \ldots, x_n)$ *along every smooth curve.*

Remark 2.2.4.

(a) We use 'smooth curve' in the sense of Chapter 1.
(b) In practice, we don't use this theorem to show that a limit exists, since it is very hard in general to show that the limit is the same along every smooth curve approaching a point.
(c) It is very useful though, to show a limit doesn't exist. The usual way we proceed is to start with simple smooth curves (for example straight lines, parabolas or cubics) approaching the point from different directions and if we can find two curves which give a different limit, then the limit (of the function of several variables) doesn't exist.

On the other hand, if we want to show that a limit does exist, then we will usually have to find some sort of cancellation or simplification of the rule of the function that makes it clear what the limit is.

Now that we know what a limit of a function of several variables is, the definition of continuity follows as you would expect.

Definition 2.2.5. A function of several variables f is said to be *continuous* at $(a_1, a_2, \ldots, a_n)$ if

$$\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f(x_1, x_2, \ldots, x_n) = f(a_1, a_2, \ldots, a_n).$$

Remark 2.2.6. We have not defined limits at boundary points of domains, so this definition does not apply at such point either. It is possible to define continuity at boundary points, but we won't do this in MA1132.

With vector-valued functions, it follows from Theorem 1.2.3 that a multiple of a continuous function is continuous and that the sum of two continuous functions is continuous. However we didn't mention anything about the products or quotients of vector-valued functions; this is simply because multiplication and division of vectors is not defined. With functions of several variables we are in a much better situation and many nice theorems carry over from the ones we are familiar with from studying functions of one variable.

Theorem 2.2.7.

(a) If f is continuous at $(a_1, a_2, \ldots, a_n)$ and $a \in \mathbb{R}$, then af is also continuous at $(a_1, a_2, \ldots, a_n)$.
(b) If f and g are continuous at $(a_1, a_2, \ldots, a_n)$, then $f + g$ is also continuous at $(a_1, a_2, \ldots, a_n)$.
(c) If f and g are continuous at $(a_1, a_2, \ldots, a_n)$, then fg is also continuous at $(a_1, a_2, \ldots, a_n)$.
(d) If f and g are continuous at $(a_1, a_2, \ldots, a_n)$ with $g(a_1, a_2, \ldots, a_n) \neq 0$, then $\dfrac{f}{g}$ is also continuous at $(a_1, a_2, \ldots, a_n)$.
(e) If f is continuous at $(a_1, a_2, \ldots, a_n)$ and g is continuous at $f(a_1, a_2, \ldots, a_n)$, then $g \circ f$ is also continuous at $(a_1, a_2, \ldots, a_n)$.
(f) If f_i is continuous at a_i for $i = 1, 2, \ldots, n$, and g is continuous at $(f_1(a_1), f_2(a_2), \ldots f_n(a_n))$, then $g(f_1(\cdot), f_2(\cdot), \ldots, f_n(\cdot))$ is also continuous at $(a_1, a_2, \ldots, a_n)$.

Remark 2.2.8.

(a) In Part (e), g is a function of one variable.
(b) In Part (f), the f_i are functions of one variable.

2.3. **Partial Derivatives.**

Having had a look at limits and continuity in the last section, we will now turn our attention to the differentiability of functions of several variables.

For functions of one variable, the derivative has a geometrical description as the gradient of the tangent to the graph of the function at the point in question. However when we are dealing with functions of several variables there are infinitely many directions to consider (even for functions of two variables). In Section 2.6 we will look at the general case, but in this section, we will start by looking at the gradients of tangents in the directions of the coordinate axes.

Let us start with the definition.

Definition 2.3.1. A function of n variables f defined on an open ball $\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \colon \|(x_1, x_2, \ldots, x_n) - (a_1, a_2, \ldots, a_n)\| < r\}$ centred at $(a_1, a_2, \ldots, a_n)$ is said to be *partially differentiable* at $(a_1, a_2, \ldots, a_n)$, *with respect to the variable* x_i if the limit

$$\lim_{h \to 0} \frac{f(a_1, \ldots, a_{i-1}, a_i + h, a_{i+1}, \ldots, a_n) - f(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n)}{h}$$

exists. If this is the case, then we define the *partial derivative* of f at $(a_1, a_2, \dots, a_n)$, *with respect to the variable* x_i to be

$$\frac{\partial f}{\partial x_i}(a_1, a_2, \dots, a_n)$$
$$= \lim_{h \to 0} \frac{f(a_1, \dots, a_{i-1}, a_i + h, a_{i+1}, \dots, a_n) - f(a_1, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_n)}{h}.$$

Remark 2.3.2.

(a) As we would expect, the function $\dfrac{\partial f}{\partial x_i}(x_1, x_2, \dots, x_n)$ is a function of n variables.
(b) As is the case with real-valued functions, there are different notations used instead of $\dfrac{\partial f}{\partial x_i}$. For example $\dfrac{\partial}{\partial x_i}(f)$ or f_{x_i}. $\dfrac{\partial f}{\partial x_i}(a_1, a_2, \dots, a_n)$ is also sometimes written as $\left.\dfrac{\partial f}{\partial x_i}\right|_{(a_1, a_2, \dots, a_n)}$

Similarly to the cases of functions of one variable and vector-valued functions, we don't usually use Definition 2.3.1 to calculate derivatives. For this we use the following proposition.

Proposition 2.3.3. *If it exists, the partial derivative* $\dfrac{\partial f}{\partial x_i}(a_1, a_2, \dots, a_n)$ *is equal to the derivative of the function of one variable* $g(x_i) := f(a_1, \dots, a_{i-1}, x_i, a_{i+1}, \dots, a_n)$ *at the point* $x_i = a_i$.

Remark 2.3.4.

(a) All this means is that to find $\dfrac{\partial f}{\partial x_i}(a_1, a_2, \dots, a_n)$, we differentiate f treating all the variables except x_i as constants.
(b) This implies that the Product and Quotient Rules carry over to functions of many variables:
If f and g are differentiable functions of n variables then
$$\frac{\partial (fg)}{\partial x_i} = \frac{\partial f}{\partial x_i} g + f \frac{\partial g}{\partial x_i}$$
and, whenever g is non-zero,
$$\frac{\partial (f/g)}{\partial x_i} = \frac{\dfrac{\partial f}{\partial x_i} g - f \dfrac{\partial g}{\partial x_i}}{g^2}.$$
(c) It also means that we can differentiate functions of several variables implicitly with respect to any of the variables. As you might expect, the method is simply to treat all the other variables as constants.
(d) In the special case where g is a function of the n variables $x_1, x_2, \dots, x_n$ and f is a function of g (so f is a function of one variable, but also depends on

$x_1, x_2, \ldots, x_n$ through g), then the Chain Rule carries over in the form

$$\frac{\partial f}{\partial x_i} = \frac{df}{dg} \cdot \frac{\partial g}{\partial x_i}.$$

There are more complicated versions of the Chain Rule, where we might have a function of n variables which depends on n other functions of m variables. We will look at this situation in Section 2.5.

Warning 2.3.5. Remark 2.3.4(a) does **NOT** mean that all the variables except x_i disappear. They do if they are in a term that does not contain x_i, but if they appear in a term that also contains x_i then they must be treated as constants.

As we mentioned earlier, partial derivatives have a geometric interpretation. The partial derivative $\frac{\partial f}{\partial x_i}(a_1, a_2, \ldots, a_n)$ is the slope of the tangent line to the graph (which is a surface) of the function f at the point $(a_1, a_2, \ldots, a_n, f(a_1, a_2, \ldots, a_n))$ in the direction of the x_i axis.

Warning 2.3.6. It might be expected that if the partial derivative with respect to all the variables exist at a point, then the function is continuous at that point. This is **NOT** true as the example

$$f(x,y) = \begin{cases} -\dfrac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

Using the limit definition, it can be shown that f is differentiable with respect to x and y at $(0,0)$, but f is not continuous there.

There is a condition involving partial derivatives which ensures the continuity of the function and we shall return to this in Section 2.6.

As with functions of one variable and vector-valued functions, we can define higher order partial derivatives of functions of several variables. Since we can differentiate the function with respect to any of the n variables the first time and also the second time, there are n^2 second order partial derivatives. These are defined as follows.

Definition 2.3.7.

$$\frac{\partial^2 f}{\partial x_j \partial x_i} := \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right).$$

Remark 2.3.8. If $i \neq j$ then $\frac{\partial^2 f}{\partial x_j \partial x_i}$ is called a mixed second order partial derivative.

Warning 2.3.9. If we are using the f_{x_i} notation then we write $\frac{\partial^2 f}{\partial x_j \partial x_i}$ as $f_{x_i x_j}$. In particular, note the different order of the variables. This is since we are differentiating f_{x_i} with respect to x_j, we write $(f_{x_i})_{x_j} = f_{x_i x_j}$.

While there are n^2 second order partial derivatives of a function of n variables, for most of the functions we are dealing with, a lot of these will be equal.

Theorem 2.3.10. *If* $\dfrac{\partial^2 f}{\partial x_j \partial x_i}$ *and* $\dfrac{\partial^2 f}{\partial x_i \partial x_j}$ *are continuous on some open ball*

$$B = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \colon \|(x_1, x_2, \ldots, x_n) - (a_1, a_2, \ldots, a_n)\| < r\},$$

then $\dfrac{\partial^2 f}{\partial x_j \partial x_i} = \dfrac{\partial^2 f}{\partial x_i \partial x_j}$ *on* B.

2.4. **Differentiability, Differentials, and Local Linearity.**

When studying calculus of real-valued functions of one variable, you probably learnt about the following important facts about a function f that is differentiable at a point $x = a$.

- f is continuous at a.
- The change in $f(x)$ can be approximated by a linear function near $x = a$.
- The graph of f has a non-vertical tangent line at the point $(a, f(a))$.

Remark 2.4.1. The linear function referred to in the second bullet point is the function $y = f'(a)x$. So, in other words, the function f may be approximated by

$$f(x) \approx f(a) + f'(a)(x - a)$$

near the point $x = a$. Note that although this is the equation of a line, it is not a linear function unless its graph goes through the origin.

In this section we want to give conditions that will ensure that similar properties hold for functions of several variables. We would like to say that if a function of n variables is differentiable at a point $(x_1, x_2, \ldots, x_n) = (a_1, a_2, \ldots, a_n)$, then

- f is continuous at $(a_1, a_2, \ldots, a_n)$.
- The change in $f(x_1, x_2, \ldots, x_n)$ can be approximated by a linear function near $(x_1, x_2, \ldots, x_n) = (a_1, a_2, \ldots, a_n)$.
- The graph of f has a non-vertical tangent 'plane' at the point $(a_1, a_2, \ldots, a_n, f(a_1, a_2, \ldots, a_n))$.

Remark 2.4.2.

(a) In the second bullet point, the linear function will be a function of n variables.
(b) In the last bullet point, the word plane is in inverted commas since it will be a plane for functions of two variables, but in general, for a function of n variables, it will be a 'hyperplane' of dimension $n - 1$.

It follows from Warning 2.3.6 that partial differentiability with respect to all the variables is not a sufficiently strong condition to ensure these points hold, so we will have to find a suitable stronger condition.

Let us first return to functions of one variable and rewrite the definition of differentiability in a different way. If we are considering differentiability at a point $x = a$

and write $\Delta f = f(a+h) - f(a)$ and $\Delta x = h$, then we can rewrite the definition of the derivative as

$$f'(a) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}. \tag{1}$$

Now Equation (1) holds if and only if

$$\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} - f'(a) = 0 \quad \Longleftrightarrow \quad \lim_{\Delta x \to 0} \frac{\Delta f - f'(a)\Delta x}{\Delta x} = 0,$$

and it is this last condition that we will generalize to get a suitable differentiation of differentiability for functions of several variables.

If we write $\Delta x_i = x_i - a_i$ and use the f_{x_i} notation for partial derivatives, we get the following definition.

Definition 2.4.3. A function f of n variables is said to be *differentiable* at a point $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ if all the partial derivatives f_{x_i} exist at $\mathbf{a}$ and

$$\lim_{(\Delta x_1, \Delta x_2, \ldots, \Delta x_n) \to (0,0,\ldots,0)} \frac{\Delta f - f_{x_1}(\mathbf{a})\Delta x_1 - f_{x_2}(\mathbf{a})\Delta x_2 - \cdots - f_{x_n}(\mathbf{a})\Delta x_n}{\sqrt{\Delta x_1^2 + \Delta x_2^2 + \cdots \Delta x_n^2}} = 0. \tag{2}$$

Remark 2.4.4. If we write the change in $\mathbf{x}$ as a vector $\Delta\mathbf{x} = (\Delta x_1, \Delta x_2, \ldots, \Delta x_n)$, then we can write Equation (2) more concisely as

$$\lim_{\Delta\mathbf{x} \to \mathbf{0}} \frac{\Delta f - f_{x_1}(\mathbf{a})\Delta x_1 - f_{x_2}(\mathbf{a})\Delta x_2 - \cdots - f_{x_n}(\mathbf{a})\Delta x_n}{\|\Delta\mathbf{x}\|} = 0.$$

We could also write the partial derivatives as a column vector

$$\mathbf{J_f}(\mathbf{a}) = \begin{bmatrix} f_{x_1}(\mathbf{a}) \\ f_{x_2}(\mathbf{a}) \\ \vdots \\ f_{x_n}(\mathbf{a}) \end{bmatrix}$$

and then write

$$\lim_{\Delta\mathbf{x} \to \mathbf{0}} \frac{\Delta f - \Delta\mathbf{x} \cdot \mathbf{J_f}(\mathbf{a})}{\|\Delta\mathbf{x}\|} = 0,$$

if we wanted to be even more concise, where the multiplication between $\Delta\mathbf{x}$ and $\mathbf{J_f}(\mathbf{a})$ is matrix multiplication.

The vector $\mathbf{J_f}(\mathbf{a})$ is called the *Jacobian*. Note that if we were studying a function that maps $\mathbb{R}^n$ to $\mathbb{R}^m$, then the Jacobian would be an $n \times m$ matrix, and this also allows us to approximate such a (differentiable) function by a linear function near a point (recall that a matrix represents a linear transformation). We will not do this in MA1132, but we will meet Jacobians again when we integrate with respect to different coordinate systems in Chapter 3.

As was the case with functions of one variable, we don't usually use the definition to determine if a function of n variables is differentiable or not. A very easy to work with sufficient condition for a function to be differentiable is given in the following theorem.

Theorem 2.4.5. *If all the first order partial derivatives of a function of n variables exist and are continuous at a point* $\mathbf{a}$*, then f is differentiable at* $\mathbf{a}$*.*

Of course, we have not really done much yet, we have introduced a definition of a property we call 'differentiability', and we have a nice way to prove a function is differentiable, but we don't know what this will imply for the function. A first important step in this direction is the next theorem.

Theorem 2.4.6. *If a function of n variables is differentiable at a point* $\mathbf{a}$*, then f is continuous at* $\mathbf{a}$*.*

Remark 2.4.7.

(a) The converse of Theorem 2.4.6 is false (as it must be since a function of one variable is a function of n variables with $n = 1$). For an example of where the converse is false (that is an example of a function which is continuous but not differentiable at a point) for a function that isn't a function of one variable, we have the function $f(x, y) = \sqrt{x^2 + y^2}$ at the point $(0, 0)$, or more generally $f(\mathbf{x}) = \|\mathbf{x}\|$ at the point $\mathbf{0}$. Another way of putting this is that the Euclidean norm on $\mathbb{R}^n$ is continuous but not differentiable at $\mathbf{0}$. In general there are many different norms that we can put on a vector space, and an important area of study is determining where the norm is differentiable, since that tells us something about the 'shape' of the space under that particular norm.
(b) If the partial derivatives of the function in Warning 2.3.6 were continuous then Theorem 2.4.5 and Theorem 2.4.6 would imply that it would be continuous at $(0, 0)$. Since it is not continuous, we know that at least one of its partial derivatives must be discontinuous at $(0, 0)$ (and in view of the symmetric definition of the function, both the partial derivatives must be discontinuous at $(0, 0)$).

Now that we have dealt with continuity, let us turn our attention to approximations. We will start with a definition.

Definition 2.4.8. Given a function of n variables f, then its *total differential* at a point $\mathbf{a}$ is defined by

$$df(dx_1, dx_2, \ldots, dx_n) = f_{x_1}(\mathbf{a})dx_1 + f_{x_2}(\mathbf{a})dx_2 + \cdots + f_{x_n}(\mathbf{a})dx_n.$$

Remark 2.4.9.

(a) There is a different function for each point $\mathbf{a}$.
(b) In this definition, the n variables $dx_1, dx_2, \ldots, dx_n$ may be regarded as (signed) differences in each coordinate from the point $\mathbf{a} = (a_1, a_2, \ldots, a_n)$. In other words, at a point $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $dx_i = x_i - a_i$ for $i = 1, 2, \ldots, n$.

If f were a linear function then we would have

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= f(a_1, a_2, \ldots, a_n) + df(dx_1, dx_2, \ldots, dx_n) \\ &= f(a_1, a_2, \ldots, a_n) + df(x_1 - a_1, x_2 - a_2, \ldots, x_n - a_n). \end{aligned}$$

In other words $df(dx_1, dx_2, \ldots, dx_n) = df(x_1 - a_1, x_2 - a_2, \ldots, x_n - a_n)$ is the difference between $f(x_1, x_2, \ldots, x_n)$ and $f(a_1, a_2, \ldots, a_n)$ (for a linear function). For a differentiable function f, we still have that $df(dx_1, dx_2, \ldots, dx_n)$ is a good approximation to the difference between $f(x_1, x_2, \ldots, x_n)$ and $f(a_1, a_2, \ldots, a_n)$, provided $(x_1, x_2, \ldots, x_n)$ is 'sufficiently' close to $(a_1, a_2, \ldots, a_n)$. Of course what 'sufficiently' means will depend on the particular function and the particular point. We can write this in terms of the Δ notation as

$$\Delta f \approx df,$$

or

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &\approx f(a_1, a_2, \ldots, a_n) + df(dx_1, dx_2, \ldots, dx_n) \\ &= f(a_1, a_2, \ldots, a_n) + f_{x_1}(\mathbf{a})\Delta x_1 + f_{x_2}(\mathbf{a})\Delta x_2 + \cdots + f_{x_n}(\mathbf{a})\Delta x_n \qquad (3) \end{aligned}$$

Remark 2.4.10.

(a) Equation (3) is sometimes called the *local linear approximation to f at $\mathbf{a}$*.
(b) If $n = 1$ then Equation (3) reduces to the familiar

$$f(x) \approx f(a) + f'(a)(x - a).$$

We have now generalised the continuity and approximable properties of a differentiable function of one variable to a differentiable function of n variables. We will return to the tangent line/plane property in Section 2.7.

2.5. **The Chain Rule.**

In Section 2.2, we mentioned a special version of the chain rule for differentiating functions of the from $g \circ f$, where f is a function of n variables and g is a function of one variable. In this section, we will look at more general versions of the chain rule.

For the first generalization, we will look at the case where f is a function of two variables x and y, and each of these variables depends on another variable t.

Theorem 2.5.1. *If $x = x(t)$ and $y = y(t)$ are differentiable at a point t_0, and if $f(x, y)$ is differentiable at the point $(x(t_0), y(t_0))$, then $f(t)$ is differentiable at t_0 and*

$$\frac{df}{dt}(t_0) = \frac{\partial f}{\partial x}(x(t_0)) \cdot \frac{dx}{dt}(t_0) + \frac{\partial f}{\partial y}(y(t_0)) \cdot \frac{dy}{dt}(t_0).$$

Remark 2.5.2. Since f is a function of the one variable t, we do not need a partial derivative symbol for $\frac{df}{dt}$, and similarly for $\frac{dx}{dt}$ and $\frac{dy}{dt}$. On the other hand, f is a function of the two variables x and y, so we do need the partial derivatives for $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$.

Theorem 2.5.1 can easily be further generalized to the situation where f is a function of the n variables $x_1, x_2, \ldots, x_n$ which each depend on the same variable t.

Theorem 2.5.3. *If* $x_1 = x_1(t), x_2 = x_2(t), \dots, x_n = x_n(t)$ *are differentiable at a point* t_0*, and if* $f(x_1, x_2, \dots, x_n)$ *is differentiable at the point* $(x_1(t_0)), x_2(t_0), \dots (x_n(t_0))$*, then* $f(t)$ *is differentiable at* t_0 *and*

$$\frac{df}{dt}(t_0) = \frac{\partial f}{\partial x_1}(x_1(t_0)) \cdot \frac{dx_1}{dt}(t_0) + \frac{\partial f}{\partial x_2}(x_2(t_0)) \cdot \frac{dx_2}{dt}(t_0) + \cdots + \frac{\partial f}{\partial x_n}(x_n(t_0)) \cdot \frac{dx_n}{dt}(t_0).$$

We can further generalize Theorem 2.5.1 to the case were x and y are also functions of two variables. This is the following theorem.

Theorem 2.5.4. *If* $x = x(u,v)$ *and* $y = y(u,v)$ *are differentiable at a point* (u_0, v_0)*, and if* $f(x,y)$ *is differentiable at the point* $(x(u_0,v_0), y(u_0,v_0))$*, then* f *has first order partial derivatives at* (u_0, v_0) *given by*

$$\frac{\partial f}{\partial u}(u_0,v_0) = \frac{\partial f}{\partial x}(x(u_0,v_0), y(u_0,v_0)) \cdot \frac{\partial x}{\partial u}(u_0,v_0) + \frac{\partial f}{\partial y}(x(u_0,v_0), y(u_0,v_0)) \cdot \frac{\partial y}{\partial u}(u_0,v_0)$$

and

$$\frac{\partial f}{\partial v}(u_0,v_0) = \frac{\partial f}{\partial x}(x(u_0,v_0), y(u_0,v_0)) \cdot \frac{\partial x}{\partial v}(u_0,v_0) + \frac{\partial f}{\partial y}(x(u_0,v_0), y(u_0,v_0)) \cdot \frac{\partial y}{\partial v}(u_0,v_0).$$

Remark 2.5.5. Here all the functions involved are functions of two variables, so we do need the partial derivatives everywhere.

Of course there is nothing special about functions of two variables, and Theorem 2.5.1 can be further generalised to the situation where f is a function of n functions, each of which is a function of m variables. This is the next theorem.

Theorem 2.5.6. *If*

$$\begin{aligned} x_1 &= x_1(u_1, u_2, \dots, u_m) = x_1(\mathbf{u}), \\ x_2 &= x_2(u_1, u_2, \dots, u_m) = x_2(\mathbf{u}), \\ &\vdots \\ x_n &= x_n(u_1, u_2, \dots, u_m) = x_n(\mathbf{u}) \end{aligned}$$

are differentiable at a point $(a_1, a_2, \dots, a_m) = \mathbf{a}$*, and if* $f(x_1, x_2, \dots, x_n) = f(\mathbf{x})$ *is differentiable at the point* $(x_1(\mathbf{a}), x_2(\mathbf{a}), \dots, x_n(\mathbf{a})) = \mathbf{x}(\mathbf{a})$*, then* f *has first order partial derivatives at* $\mathbf{a}$ *given by*

$$\begin{aligned} \frac{\partial f}{\partial u_1}(\mathbf{a}) &= \frac{\partial f}{\partial x_1}(\mathbf{x}(\mathbf{a})) \cdot \frac{\partial x_1}{\partial u_1}(\mathbf{a}) + \frac{\partial f}{\partial x_2}(\mathbf{x}(\mathbf{a})) \cdot \frac{\partial x_2}{\partial u_1}(\mathbf{a}) + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}(\mathbf{a})) \cdot \frac{\partial x_n}{\partial u_1}(\mathbf{a}) \\ \frac{\partial f}{\partial u_2}(\mathbf{a}) &= \frac{\partial f}{\partial x_1}(\mathbf{x}(\mathbf{a})) \cdot \frac{\partial x_1}{\partial u_2}(\mathbf{a}) + \frac{\partial f}{\partial x_2}(\mathbf{x}(\mathbf{a})) \cdot \frac{\partial x_2}{\partial u_2}(\mathbf{a}) + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}(\mathbf{a})) \cdot \frac{\partial x_n}{\partial u_2}(\mathbf{a}) \\ &\vdots \\ \frac{\partial f}{\partial u_m}(\mathbf{a}) &= \frac{\partial f}{\partial x_1}(\mathbf{x}(\mathbf{a})) \cdot \frac{\partial x_1}{\partial u_m}(\mathbf{a}) + \frac{\partial f}{\partial x_2}(\mathbf{x}(\mathbf{a})) \cdot \frac{\partial x_2}{\partial u_m}(\mathbf{a}) + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}(\mathbf{a})) \cdot \frac{\partial x_n}{\partial u_m}(\mathbf{a}). \end{aligned}$$

Remark 2.5.7.

(a) Theorem 2.5.6 contains all the other versions of the chain rule we have mentioned in this Chapter. For example, the version of the chain rule we used in Section 2.2 can be obtained by leaving m and letting $n = 1$, Theorem 2.5.1 can be obtained by letting $n = 2$ and $m = 1$, Theorem 2.5.3 can be obtained by leaving n and letting $m = 1$, and Theorem 2.5.4 can be obtained by letting $m = n = 2$. This is something that often happens in mathematics: we find a general theorem that contains several other theorems as special cases.

(b) There is nothing to prevent $x_i(u_1, u_2, \ldots, u_m) = u_i$ for one or more i.

(c) Sometimes the object of differentiating might not be to find the derivative on the LHS of the equation, it might be to obtain an expression for one of the derivatives on the right.

(d) The chain rule can also be applied to partial derivatives in order to find higher order partial derivatives.

2.6. **Directional Derivatives and Gradients.**

In Section 2.3, we saw how to partially differentiate a function of n variables with respect to any of the variables, and interpreted this derivative as the instantaneous rate at which the function is increasing at a point in the direction of one of the coordinate axes.

In this chapter, we will do something similar, but in this case, we will look at the instantaneous rate at which the function is increasing at a point in any direction. For a partial derivative we looked at the limit of the quotient

$$\frac{f(a_1, \ldots, a_{i-1}, a_i + h, a_{i+1}, \ldots, a_n) - f(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n)}{h}, \tag{4}$$

as h tends to zero.

Our approach here will be similar. Let us first examine exactly what we have done in Equation 4. We looked at the change in f as we moved h in the direction of the x_i axis and divided it by h. In particular, note that the (signed) distance we moved in the direction of the x_i axis and the number we divided the difference in f by were the same.

Now, suppose that we want to do the same thing in the direction $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ instead of the direction $(0, \ldots, 0, 1, 0, \ldots, 0)$ (where the 1 is in the i'th coordinate). If we just looked at the quotient

$$\frac{f(a_1 + u_1 h, a_2 + u_2 h, \ldots, , a_n + u_n h) - f(a_1, a_2, \ldots, a_n)}{h},$$

then the problem is that this quotient would depend on the magnitude of $\mathbf{u}$ as well as its direction, which is clearly not what we want. Accordingly, in the definition, we have to specify that $\mathbf{u}$ is a unit vector. Then the (signed) distance from $(a_1, a_2, \ldots, a_n)$ to

$$(a_1 + u_1 h, a_2 + u_2 h, \ldots, , a_n + u_n h) = (a_1, a_2, \ldots, a_n) + h(u_1, u_2, \ldots, u_n)$$

is h, as we want.

Definition 2.6.1. If f is a function of n variables defined on an open ball $\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : \|(x_1, x_2, \ldots, x_n) - (a_1, a_2, \ldots, a_n)\| < r\}$ centred at $(a_1, a_2, \ldots, a_n)$ and if $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ is a unit vector, then the *directional derivative of f in the direction* $\mathbf{u}$ is defined to be

$$D_{\mathbf{u}} f(a_1, a_2, \ldots, a_n) = \lim_{h \to 0} \frac{f(a_1 + u_1 h, a_2 + u_2 h, \ldots, , a_n + u_n h) - f(a_1, a_2, \ldots, a_n)}{h},$$

provided this limit exists.

Remark 2.6.2.

(a) The directional derivative at a point, $D_{\mathbf{u}} f(a_1, a_2, \ldots, a_n)$, is a number (rather than a vector).
(b) The directional derivative $D_{\mathbf{u}} f$, is a real valued function of n variables.
(c) Writing $\mathbf{a} = (a_1, a_2, \ldots, a_n)$, we can write the definition in vector form as

$$D_{\mathbf{u}} f(\mathbf{a}) = \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{u}) - f(\mathbf{a})}{h}.$$

(d) We can also adapt the definition for a non-unit vector $\mathbf{u}$ by having $h\|\mathbf{u}\|$ in the denominator instead of just h.
(e) An equivalent definition is

$$D_{\mathbf{u}} f(\mathbf{a}) = \frac{d}{dh}(f(\mathbf{a} + h\mathbf{u}))(0).$$

That is, we differentiate $f(\mathbf{a} + h\mathbf{u})$ (which is a function of the one real variable h) and evaluate the derivative at $h = 0$.
(f) As we want, $D_{\mathbf{u}} f(\mathbf{a})$ does have a geometric interpretation as the instantaneous rate the function is increasing in the direction of $\mathbf{u}$ at the point $\mathbf{a}$. For a function of two variables we can say equivalently that it is the slope of the tangent to the graph of f in the direction of $\mathbf{u}$.
(g) All the partial derivatives of f are special cases of the directional derivative; they are the directional derivatives in the directions of the coordinate axes.
(h) As was the case with partial derivatives, there are many different notations. For example, instead of $D_{\mathbf{u}} f$, we might write $\nabla_{\mathbf{u}} f$, or $\dfrac{df}{d\mathbf{u}}$ or any of these with any alternative vector notation replacing $\mathbf{u}$.

As is often the case in mathematics, the definition is not the thing that we use for calculation, and that is the case here. For actually finding the directional derivative, we use the following theorem.

Theorem 2.6.3. *If f is a a function of n variables that is differentiable at a point $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and if $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ is a unit vector, then the directional derivative of f in the direction $\mathbf{u}$ at $\mathbf{a}$ exists and is given by*

$$D_{\mathbf{u}} f(\mathbf{a}) = f_{x_1}(\mathbf{a}) u_1 + f_{x_2}(\mathbf{a}) u_2 + \cdots + f_{x_n}(\mathbf{a}) u_n. \tag{5}$$

Remark 2.6.4. Recall that if all the partial derivatives of a function exist and are continuous, then the function is differentiable. In this situation, we can use evaluate the directional derivative using Equation (5).

We can also write Equation (5) in terms of a dot product as follows:

$$D_{\mathbf{u}}f(\mathbf{a}) = (f_{x_1}(\mathbf{a}), f_{x_2}(\mathbf{a}), \ldots, f_{x_n}(\mathbf{a})) \cdot (u_1, u_2, \ldots, u_n).$$

So it seems like it would be a good idea to be able to easily refer to the vector $(f_{x_1}(\mathbf{a}), f_{x_2}(\mathbf{a}), \ldots, f_{x_n}(\mathbf{a}))$ and this leads to the following definition.

Definition 2.6.5. If f is a function of n variables is partially differentiable with respect to all the variables at a point $\mathbf{a} = (a_1, a_2, \ldots, a_n)$, then the *gradient* of f at $\mathbf{a}$ is defined to be

$$\nabla f(\mathbf{a}) = (f_{x_1}(\mathbf{a}), f_{x_2}(\mathbf{a}), \ldots, f_{x_n}(\mathbf{a})).$$

Remark 2.6.6.

(a) The symbol ∇ is read as 'del' or 'nabla'.
(b) Sometimes $\nabla f(\mathbf{a})$ is written as a column vector rather than a row vector.
(c) Clearly we can now write the directional derivative as

$$D_{\mathbf{u}}f(\mathbf{a}) = \nabla f(\mathbf{a}) \cdot (u_1, u_2, \ldots, u_n). \tag{6}$$

If we examine Equation (6) and use the properties of the dot product, we have

$$D_{\mathbf{u}}f(\mathbf{a}) = \|\nabla f(\mathbf{a})\| \cdot \|(u_1, u_2, \ldots, u_n)\| \cos(\theta) = \|\nabla f(\mathbf{a})\| \cos(\theta),$$

where θ is the angle between $\nabla f(\mathbf{a})$ and $(u_1, u_2, \ldots, u_n)$.

Thus, at a given point $\mathbf{a}$, $D_{\mathbf{u}}f(\mathbf{a})$ is a maximum when $\theta = 0$ and $D_{\mathbf{u}}f(\mathbf{a})$ is a minimum when $\theta = \pi$. Put another way, $\nabla f(\mathbf{a})$ points in the direction where f increases most rapidly, and $-\nabla f(\mathbf{a})$ points in the direction where f decreases most rapidly (provided $\nabla f(\mathbf{a}) \neq \mathbf{0}$). Moreover the rate of increase and decrease of f in these directions is $\|\nabla f(\mathbf{a})\|$ and $-\|\nabla f(\mathbf{a})\|$, respectively.

On the other hand, if $\|\nabla f(\mathbf{a})\| = 0$, then the directional derivative of f at $\mathbf{a}$ is zero in any direction. Or put another way, if all the partial derivatives at a point are zero, then all directional derivatives at this point are also zero.

2.7. **Tangent Planes and Normal Vectors.**

Let us now return to Equation (6) in Section 2.6 and see if we can extract any more information from it. Let us suppose that the gradient is not the zero vector. Then $D_{\mathbf{u}}f(\mathbf{a}) = 0$ if and only if the angle between $\nabla f(\mathbf{a})$ and $(u_1, u_2, \ldots, u_n)$ is $\frac{\pi}{2}$. However $D_{\mathbf{u}}f(\mathbf{a}) = 0$ if and only if $\mathbf{u}$ lies in a 'direction' of the level set $\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \colon f(x_1, x_2, \ldots, x_n) = f(a_1, a_2, \ldots, a_n)\}$. In other words, the gradient is 'perpendicular' to the level set
$\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \colon f(x_1, x_2, \ldots, x_n) = f(a_1, a_2, \ldots, a_n)\}$.

Remark 2.7.1. You may wonder why I have put the words direction and perpendicular in inverted commas above. This is since for general functions, we have no idea what the level set looks like, it could be some sort of fractal for example, in which case direction would not really have a meaning. It can be shown that if all the partial derivatives exist and are continuous at a point, then locally the level set is a 'surface'. We won't go into this proof, or exactly what a surface (manifold) is in MA1132, but we can use our intuitive idea of what a surface is. You will probably treat this more rigorously in a course about differential geometry in the future.

In MA1132, we will be dealing with 'nice' functions, and we will use the above discussion to define what we mean by a normal vector and a tangent plane.

Definition 2.7.2. Assume that f is a function of n variables and that it has continuous first order partial derivatives. Suppose that a point $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ lies on the level surface $S = \{(x_1, x_2, \ldots, x_n)\colon f(x_1, x_2, \ldots, x_n) = c\}$.
If $\nabla f(a_1, a_2, \ldots, a_n) \neq \mathbf{0}$, then $\mathbf{n} = \nabla f(a_1, a_2, \ldots, a_n)$ is a *normal vector* to the surface S at $(a_1, a_2, \ldots, a_n)$. We then define the *tangent plane* to the surface S at $(a_1, a_2, \ldots, a_n)$ to be the plane passing through $(a_1, a_2, \ldots, a_n)$ with $\mathbf{n}$ as its normal vector.

Remark 2.7.3.

(a) The equation of the tangent plane in Definition 2.7.2 is

$$f_{x_1}(\mathbf{a})(x_1 - a_1) + f_{x_2}(\mathbf{a})(x_2 - a_2) + \cdots + f_{x_n}(\mathbf{a})(x_n - a_n) = 0. \tag{7}$$

Alternatively, if we write $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, then we can write the equation as

$$\nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) = 0.$$

(b) While we have called it a tangent plane, if f is a function of two variables it will really be a tangent line (a translate of a one dimensional subspace of $\mathbb{R}^2$), if f is a function of three variables it will be a tangent plane (a translate of a two dimensional subspace of $\mathbb{R}^3$), and if f is a function of n variables it will be really be a tangent space (a translate of a $n - 1$ dimensional subspace of $\mathbb{R}^n$, or put another way a translate of a $n - 1$ dimensional hyperplane, or put yet another way, an affine hyperplane).

(c) In MA1132, we will deal with the cases where f is a function of two or three variables, but I still thought it would do no harm to write down the general case.

(d) Note the general equation of a tangent plane can also be written

$$b_1 x_1 + b_2 x_2 + \cdots + b_n x_n = c,$$

where the b_i and c are constants.

While we have defined tangent planes in terms of level surfaces, we can also find tangent planes to the graphs of functions of $n - 1$ variables of the form

$$x_n = g(x_1, x_2, \ldots, x_{n-1}),$$

by defining the function of n variables

$$f(x_1, x_2, \ldots, x_n) = x_n - g(x_1, x_2, \ldots, x_{n-1})$$

and applying the Definition 2.7.2 to the level curve

$$f(x_1, x_2, \dots, x_n) = x_n - g(x_1, x_2, \dots, x_{n-1}) = 0.$$

If $\mathbf{a} = (a_1, a_2, \dots, a_n)$ is a point on this surface, then we have

$$a_n = g(a_1, a_2, \dots, a_{n-1}) \tag{8}$$

and writing $\mathbf{a}' = (a_1, a_2, \dots, a_{n-1})$,

$$f_{x_1}(\mathbf{a}) = -g_{x_1}(\mathbf{a}'), f_{x_2}(\mathbf{a}) = -g_{x_2}(\mathbf{a}'), \dots, f_{x_{n-1}}(\mathbf{a}) = -g_{x_{n-1}}(\mathbf{a}'), f_{x_n}(\mathbf{a}) = 1.$$

Substituting these these into Equation (7), we obtain the equation

$$-g_{x_1}(\mathbf{a}')(x_1 - a_1) - g_{x_2}(\mathbf{a}')(x_2 - a_2) - \cdots - g_{x_{n-1}}(\mathbf{a}')(x_{n-1} - a_{n-1}) + (x_n - a_n) = 0.$$

That is

$$x_n = a_n + g_{x_1}(\mathbf{a}')(x_1 - a_1) + g_{x_2}(\mathbf{a}')(x_2 - a_2) + \cdots + g_{x_{n-1}}(\mathbf{a}')(x_{n-1} - a_{n-1}).$$

Using Equation (8), this becomes

$$x_n = g(\mathbf{a}') + g_{x_1}(\mathbf{a}')(x_1 - a_1) + g_{x_2}(\mathbf{a}')(x_2 - a_2) + \cdots + g_{x_{n-1}}(\mathbf{a}')(x_{n-1} - a_{n-1}).$$

Comparing this with (3), we see that the tangent plane to the graph of g at the point $\mathbf{a}'$ is in fact the graph of the local linear approximation to g at that point.

2.8. **Maxima and Minima of Functions of Two Variables.**

In this section we will extend some of the techniques for finding maxima and minima of functions of one variable to functions of two variables. Let us start with some definitions, which are almost identical to those you will have met when studying functions of one variable, just with intervals replaced by discs.

Definition 2.8.1.

- Given a set $S \subseteq \mathbb{R}^2$ and a function $f\colon S \to \mathbb{R}$, then we say f attains a *global maximum* at $(a, b) \in S$ if $f(x, y) \leqslant f(a, b)$ for all $(x, y) \in S$.
- Given a set $S \subseteq \mathbb{R}^2$ and a function $f\colon S \to \mathbb{R}$, then we say f attains a *global minimum* at $(a, b) \in S$ if $f(x, y) \geqslant f(a, b)$ for all $(x, y) \in S$.
- Given a set $S \subseteq \mathbb{R}^2$ and a function $f\colon S \to \mathbb{R}$, then we say f attains a *local maximum* at $(a, b) \in S$ if there exists some number $r > 0$ so that $f(x, y) \leqslant f(a, b)$ for all for all $(x, y) \in \{(x, y) \in \mathbb{R}^2 \colon \|(x-a, y-b)\| < r\} \cap S$.
- Given a set $S \subseteq \mathbb{R}^2$ and a function $f\colon S \to \mathbb{R}$, then we say f attains a *local minimum* at $(a, b) \in S$ if there exists some number $r > 0$ so that $f(x, y) \geqslant f(a, b)$ for all for all $(x, y) \in \{(x, y) \in \mathbb{R}^2 \colon \|(x-a, y-b)\| < r\} \cap S$.

Remark 2.8.2.

(a) If we replace $\leqslant$ and $\geqslant$ in the above definitions with $<$ and $>$ then we say the maxima and minima are *strict*.
(b) Sometimes we may use the word 'extremum' instead of either maximum or minimum (or both) or 'extrema' instead of either maxima or minima (or both).

One question we can ask is whether a maximum or minimum exists, without necessarily being able to find it algebraically. This is important since once we know that a maximum or minimum exists then we can attempt to use numerical methods to find it. There is a very important theorem that we have at our disposal, but before we can state it, we need some definitions.

Definition 2.8.3.

- We say a set in $\mathbb{R}^2$ is *bounded* if all points of it lie within a finite distance of the origin. Put an other way, a set S is bounded if we can find a positive number r such that
$$S \subseteq \{(x, y) \in \mathbb{R}^2 \colon \|(x, y)\| \leqslant r\}.$$
- A set in $\mathbb{R}^2$ is said to be *open* if around every point in the set there is at least one open ball contained within the set. In other words, a set S in $\mathbb{R}^2$ is open if for each $(a, b) \in S$, there exists a positive number r such that
$$\{(x, y) \in \mathbb{R}^2 \colon \|(x - a, y - b)\| < r\} \subseteq S.$$
- A set S in $\mathbb{R}^2$ is said to be *closed* if its complement (that is $\mathbb{R}^2 \backslash S$) is open.

We are now in a position to state the extreme value theorem.

Theorem 2.8.4. *If a function of two variables f is continuous on a closed and bounded subset of $\mathbb{R}^2$, then f attains a global maximum and a global minimum on the set.*

Remark 2.8.5. While Definition 2.8.3 and Theorem 2.8.4 have been stated for subsets and functions on $\mathbb{R}^2$, they can also be stated for much more general topological spaces. In this case the appropriate generalization of ‘closed and bounded’ is ‘compact’.

Now that we have looked at global maxima and minima, let us look at local maxima and minima. As was the case with global maxima and minima, the situation is very similar to the situation for functions of one variable.

Theorem 2.8.6. *If a function of two variables f has a local extremum at a point (a, b) and if the partial derivatives of f exist at (a, b) then $f_x(a, b) = f_y(a, b) = 0$.*

We also have a name for places where we have both partial derivatives zero.

Definition 2.8.7. If the partial derivatives of a function of two variables f exist and are zero at a point (a, b) then we say that (a, b) is a *critical point*.

It follows from Theorem 2.8.6 and Definition 2.8.7 that one place where we might look for local extrema is at critical points. However, as is the case for functions of one variable, not all critical points yield local extrema. For functions of one variable we had so called points of inflection. For functions of two variables, we have saddle points.

Definition 2.8.8. A function of two variables f has a saddle point at a point (a, b) if in one direction through (a, b) in the xy-plane f restricted to this line has a local maximum and in another direction through (a, b) in the xy-plane f restricted to this line has a local minimum.

The following theorem helps us find those critical points where we do have local extrema or saddle points and also classify whether they are maxima, minima or saddle points.

Theorem 2.8.9 (Second Derivative Test)**.** *Let f be a function of two variables which has continuous second order partial derivatives on some open disc centred at a critical point (a, b). If we let $D = f_{xx}(a, b)f_{yy}(a, b) - f_{xy}^2(a, b)$, then*

- *If $D > 0$ and $f_{xx}(a, b) > 0$ then f has a local minimum at (a, b).*
- *If $D > 0$ and $f_{xx}(a, b) < 0$ then f has a local maximum at (a, b).*
- *If $D < 0$ then f has a saddle point at (a, b).*

Definition 2.8.10.

- If $D = 0$ then the theorem does not tell us anything and we have to use other techniques to classify the critical point.
- The number $D = f_{xx}(a, b)f_{yy}(a, b) - f_{xy}^2(a, b)$ is in fact the determinant of the two by two matrix $\begin{bmatrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{yx}(a, b) & f_{yy}(a, b) \end{bmatrix}$ called the *Hessian*. This matrix can be generalized in the obvious way for functions of n variables and the second derivative test can be generalized to use this Hessian. Unfortunately, the generalization is a little bit complicated, involving the principal minors, so we won't cover it here.

Apart from critical points, local extrema can also occur where one or both of the partial derivatives don't exist. Sometimes these are called *singular* points, but the actual name can vary from text to text. In fact for a certain part of the domain, critical points and singular points are the only places where local extrema can occur, but before we state the relevant theorem we need the following definition.

Definition 2.8.11. A point (a, b) is said to be an interior point of a set $S \subseteq \mathbb{R}^2$ if there exists a positive number r such that

$$\{(x, y) \in \mathbb{R}^2 \colon \|(x - a, y - b)\| < r\} \subseteq S.$$

Remark 2.8.12. Definition 2.8.11 can be generalized to all topological spaces.

Now for the theorem.

Theorem 2.8.13. *If a function of two variables f has an global extremum at an interior point of its domain, then this extremum occurs at a critical point or a singular point.*

The importance of this theorem is that if we want to find the global maximum or minimum of a function on a closed bounded set, then since a global extremum must also be a local extremum, we only need look at the critical points and singular points in the interior of the domain and then points on the boundary (that is the domain minus the interior). Usually for the boundary points, the problem will reduce to a problem about a function of one variable.

Remark 2.8.14. If the set we are not interested in is not closed and bounded, then the function may not have a global maximum or minimum, but if it does then it still must occur at a boundary point, a critical point or a singular point.

2.9. **Lagrange Multipliers.**

In this final section of this chapter, we will again look at maximizing and minimizing functions of several variables, but this time we will look for maxima and minima subject to an additional constraint. For example, if we are driving through a mountain range, we may not be interested in the height of the mountains but we may be interested in the maximum height that the road reaches, and it is this sort of problem that the technique of Lagrange multipliers will enable us to attack.

So, suppose that we have a function $f(x, y)$ (in the example above, this is the height of the ground) and a constraint $g(x, y) = 0$ defined in terms of a level set of a function g (this level set traces out the road in the above example). The problem is to maximize or minimize f subject to the constraint $g(x, y) = 0$ (of course in the above example, we are interested in the maximum).

Let us also assume that we can smoothly parametrize the level set $g(x, y) = 0$ with a unit speed parametrization $x = x(s)$, $y = y(s)$. If f has a local extremum subject to the constraint at a point $(x(s_0), y(s_0))$, then it follows that we must have $\dfrac{df}{ds}(s_0) = 0$. However, using the chain rule,

$$\frac{df}{ds} = \frac{\partial f}{\partial x} \cdot \frac{dx}{ds} + \frac{\partial f}{\partial y} \cdot \frac{dy}{ds} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \cdot \left(\frac{dx}{ds}, \frac{dy}{ds}\right) = \nabla f \cdot \mathbf{T},$$

where $\mathbf{T}$ is the tangent vector to the path defined by $x = x(s)$, $y = y(s)$.

In other words, at a local extremum, subject to the constraint, either the gradient of f is $\mathbf{0}$ or it is at right angles to the tangent vector to the constraint path. This means that at such an extremum, either the gradient of f is $\mathbf{0}$ or it is parallel to the normal vector to the constraint path. In fact, both these situations are covered by saying that the gradient of f is a multiple of the normal vector to the constraint path, that is, the gradient of f is a multiple of the gradient of g. Let us summarize this discussion in a theorem.

Theorem 2.9.1. *Let f and g be functions of n variables with continuous first order partial derivatives on some open set containing the constraint curve $g(x, y) = 0$, and assume that $\nabla g \neq \mathbf{0}$ at any point on this curve. If f has a local extremum at a point $\mathbf{a} = (a_1, a_2, \ldots, a_n)$, subject to the constraint, then*

$$\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a}) \quad \textit{for some} \quad \lambda \in \mathbb{R}. \tag{9}$$

Remark 2.9.2.

(a) The number λ is called a *Lagrange multiplier*.

(b) Although our discussion only dealt with the case $n = 2$, the theorem is true for any n, and, in contrast to the situation for local extrema with no constraints, this does not increase the complexity of the theorem. It does increase the complexity when we use the theorem though. Note that the constraint $g(\mathbf{a}) = 0$ defines a $n - 1$ dimensional surface in the general case.

(c) Equation (9) can also be written $\nabla\mathcal{L} = 0$, where we define

$$\mathcal{L}(x_1, x_2, \ldots, x_n, \lambda) = f(x_1, x_2, \ldots, x_n) - \lambda g(x_1, x_2, \ldots, x_n).$$

This function $\mathcal{L}$ is called the *Lagrangian*.

(d) This method can easily be generalized to the case where we have more than one constraint, but we won't cover this generalization in MA1132.

(e) It can also be generalized to the case where the constraints are of the form $g(\mathbf{a}) \leqslant 0$ by using the *Karush-Kuhn-Tucker conditions* (often abbreviated KKT conditions), but this is not so straightforward.

(f) Theorem 2.9.1 gives necessary but not sufficient conditions for an extremum to occur. In practice, our procedure will be to solve the simultaneous equations $\nabla f(\mathbf{a}) = \lambda\nabla g(\mathbf{a})$ and $g(\mathbf{a}) = 0$ to find possible places where extrema can occur, and then evaluate f at these points. If the set of all points where $g(\mathbf{a}) = 0$ is closed and bounded, then we know by the Extreme Value Theorem that f must attain a maximum and a minimum on this set, so we can determine where these occur by using the values of f calculated above. If the set of all points where $g(\mathbf{a}) = 0$ is not closed and bounded, then things become more complicated.

(g) Solving the simultaneous equations $\nabla f(\mathbf{a}) = \lambda\nabla g(\mathbf{a})$ and $g(\mathbf{a}) = 0$, where f and g are functions of n variables amount to solving $n + 1$ equations in $n + 1$ unknowns.