

# Math 237 Notes

Syed Mustafa Raza Rizvi

July 4, 2020

These are my 2nd year Calculus 3 notes at the University of Waterloo (MATH 237). They are pretty similar to the content you may see in the course notes by J. Wainwright, J. West, D. Wolczuk.

You will find that these aren't very useful as notes, in the sense that they are not significantly shorter than the content in the course notes, they're really just a way for me to type down the content I am learning and absorb it. Also, I won't be including the proofs, it's best to read the course notes for that.

Thanks to Prof. Dan Wolczuk for providing me with the macros to typeset this LaTeX document. If the university or staff feel that I should take down this document, please feel free to contact me on github

# Week 1

## Unit 1: Graphs of Scalar Functions

### Scalar Functions

A review of basic vocabulary about functions in general

- A function  $f : A \rightarrow B$  associates with each element  $a \in A$  a unique element in  $f(a) \in B$  called the **image** of  $a$  under  $f$
- The set  $A$  is called the **domain** of  $f$  and is denoted by  $D(f)$
- The set  $B$  is called the **codomain** of  $f$
- The subset of  $B$  consisting of all  $f(a)$  is called the **range** of  $f$  and is denoted by  $R(f)$

We will usually look at functions whose domain is a subset of  $\mathbb{R}^2$  and whose codomain is  $\mathbb{R}$ . I.e. we consider functions  $f$  which map points  $(x, y) \in \mathbb{R}^2$  to a real scalar  $f(x, y) \in \mathbb{R}$ . We write  $z = f(x, y)$ . We will also consider more general functions  $f(x_1, \dots, x_n)$  which map subsets of  $\mathbb{R}^n$  to  $\mathbb{R}$

A **scalar function**  $f(x_1, \dots, x_n)$  of  $n$  variables is a functions whose domain is a subset of  $\mathbb{R}^n$  and whose range is a subset of  $\mathbb{R}$

Definition

Scalar

May also be denoted by  $f(\mathbf{x})$  or  $f(\vec{x})$

Function

### Geometric Interpretation of $z = f(x, y)$

When we graph a function  $y = f(x)$ , we plot points  $(a, f(a))$  in the  $xy$ -plane. Observe that we can think of  $f(a)$  as representing the height of the graph  $y = f(x)$  above (or below if negative) the  $x$ -axis at  $x = a$

We define the **graph** of a function  $f(x, y)$  as the set of all points

$$\{(a, b, f(a, b)) \in \mathbb{R}^3 : (a, b) \in D(f)\}$$

We think of  $f(a, b)$  as representing the height of the graph  $z = f(x, y)$  above (or below if negative) the  $xy$ -plane at the point  $(x, y) = (a, b)$

◦ When  $f$  is defined as  $f(x, y) = c_1x + c_2y + c_3$ , where  $c_1, c_2, c_3 \in \mathbb{R}$ , the graph of  $z = f(x, y)$  is a **plane**. (Note: it is of the form  $z - c_1x + c_2y = c_3$ )

**Level curves** are 2-dimensional slices of a surface, sort of like a top-down view of what the curve looks like for a fixed  $z$  value.

The level curves of a function  $f(x, y)$  are the curves

Definition

$$f(x, y) = k, k \in \mathbb{R}$$

Level Curves

The level curve of  $f(x, y) = k$  is the intersection of  $z = f(x, y)$  and the horizontal plane  $z = k$ . In our family of curves, each value of  $k$  represents a height above the  $xy$ -plane. Thus, the family of level curves is often called a **contour map** or a **topographic map**

(A little general info: Weather maps which show regions of constant temperatures are called **isotherms**, in barometric pressure charts curves of constant pressure are called **isobars**. Another example would be an MRI scan)

- A level curve that behaves unusually compared to other members of the family is called an **exceptional level curve**

A **cross section** of a surface  $z = f(x, y)$  is the intersection of  $z = f(x, y)$  with a plane

Definition

Cross

Sections

For sketching purposes, it is useful to consider cross sections formed by intersection  $z = f(x, y)$  with the *vertical* planes  $x = c$  and  $y = d$ , where  $c, d$  can take on multiple values as  $k$  did in level curves

- $f(x, y) = x^2 + y^2$  gives a **paraboloid** surface (level curves are circles; cross sections are parabolas)

- $f(x, y) = x^2 - y^2$  gives a **saddle surface** (level curves are hyperbolae, about  $x$ -axis above/below a certain  $k$  value, about  $y$ -axis below/above a certain  $k$  value; cross sections are parabolas (if **im not wrong**))

- $f(x, y) = x^2$  gives a **parabolic cylinder** (level curves are straight lines,  $x = \pm\sqrt{k}$ ; cross section is a parabola. Since it has the same cross section for all planes  $y = d$ , it is called a cylinder by definition)

A **level surface** of a scalar function  $f(x, y, z)$  is defined by

Definition

$$f(x, y, z) = k, \quad k \in R(f)$$

Level

Surfaces

A level surface, is the analogy case of a surface having a level curve, but one-dimension up.

So, we have a 4-dimensional shape (instead of a surface), which is made up of (layers of) surfaces, level surfaces. Each level surface can be found by parameterizing  $f(x, y, z)$

A **level set** of a scalar function  $f(\vec{x}), \vec{x} \in \mathbb{R}^n$  is defined by

Definition

$$\{\vec{x} \in \mathbb{R}^n | f(\vec{x}) = k, \text{ for } k \in R(f)\}$$

Level Sets

So a level set is a generalisation of what we learned for level curves (case of  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ) and level surfaces (case of  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ ) for  $f : \mathbb{R}^n \rightarrow \mathbb{R}$

- For an  $f$  defined by:  $f(x_1, \dots, x_n) = x_1^2 + \dots + x_n^2$ , the level sets for  $f(\vec{x}) = k, k > 0 \in \mathbb{R}^n$  are called **(n - 1)-spheres**, denoted by  $S^{n-1}$  (e.g. for  $n = 3$  we get a 2-sphere denoted by  $S^2$ )

# Week 2

## Unit 2: Limits

### Definition of a Limit for One Variable

For a real-valued function  $f(x)$  we defined  $\lim_{x \rightarrow a} f(x) = L$  to mean that the values of  $f(x)$  can be made arbitrarily close to  $L$  by taking  $x$  sufficiently close to  $a$ .

More precisely,

For every  $\epsilon > 0$  there exists a  $\delta > 0$  :

$$|f(x) - L| < \epsilon \text{ whenever } 0 < |x - a| < \delta (*)$$

$$\text{and } \lim_{x \rightarrow a} f(x) = L \iff \lim_{x \rightarrow a^-} f(x) = L = \lim_{x \rightarrow a^+} f(x)$$

This means no matter what  $\epsilon > 0$  value we choose, we can always find a corresponding  $\delta > 0$  value that would satisfy the condition (\*)

Definition

Single

Variable

definition of

a Limit

### Definition of a Limit for Functions of Two Variables

We define the limit for functions of two variables in a very similar way to the limit of functions of a single variable. For a scalar function  $f(x, y)$ , we want  $\lim_{(x, y) \rightarrow (a, b)} f(x, y) = L$ , to mean that the values of  $f(x, y)$  can be made arbitrarily close to  $L$  by taking  $(x, y)$  sufficiently close to  $(a, b)$

For a single variable we could approach the limit from either the left or the right.

For multivariable scalar functions our domain is multidimensional and so we can approach it from infinitely many directions, moreover, we aren't even restricted to straight lines either; we can approach  $(a, b)$  along any smooth curve.

An **open interval** is defined as

$$(-r, r) = \{x : |x| < r\}$$

where  $r \in \mathbb{R}$

Definition

Open

Interval

**Euclidian distance** in  $\mathbb{R}^2$  is defined as

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

Definition

Euclidian

Distance

An **r-neighbourhood** of a point  $(a, b) \in \mathbb{R}^2$  is a set

Definition

Neighbourhood

$$N_r(a, b) = \{(x, y) \in \mathbb{R}^2 \mid \|(x, y) - (a, b)\| < r, r \in \mathbb{R}\}$$

(Where  $r > 0$  (=??))

You may notice that the r-neighbourhood of  $(a, b)$  is simply a locus of distance  $r$  or less from the point  $(a, b)$

Assume  $f(x, y)$  is defined in a neighbourhood of  $(a, b)$ , except possibly at  $(a, b)$ . If, for every  $\epsilon > 0$  there exists a  $\delta > 0$  such that

Definition

Limit

$$0 < \|(x, y) - (a, b)\| < \delta \implies |f(x, y) - L| < \epsilon$$

Then

$$\lim_{(x, y) \rightarrow (a, b)} f(x, y) = L$$

Although we said that the we can approach the limits from infinitely many directions, note that the limit definition does not refer to any direction at all, and refers only to the distance between  $(x, y)$  and  $(a, b)$

## Limit Theorems

If  $\lim_{(x, y) \rightarrow (a, b)} f(x, y)$  and  $\lim_{(x, y) \rightarrow (a, b)} g(x, y)$  both exist, then

Limit

Theorem 1

1.

$$\lim_{(x, y) \rightarrow (a, b)} [f(x, y) + g(x, y)] = \lim_{(x, y) \rightarrow (a, b)} f(x, y) + \lim_{(x, y) \rightarrow (a, b)} g(x, y)$$

2.

$$\lim_{(x, y) \rightarrow (a, b)} [f(x, y)g(x, y)] = \left[ \lim_{(x, y) \rightarrow (a, b)} f(x, y) \right] \left[ \lim_{(x, y) \rightarrow (a, b)} g(x, y) \right]$$

3.

$$\lim_{(x, y) \rightarrow (a, b)} \frac{f(x, y)}{g(x, y)} = \frac{\lim_{(x, y) \rightarrow (a, b)} f(x, y)}{\lim_{(x, y) \rightarrow (a, b)} g(x, y)}, \text{ provided } \lim_{(x, y) \rightarrow (a, b)} g(x, y) \neq 0$$

If  $\lim_{(x, y) \rightarrow (a, b)} f(x, y)$  exists, then the limit is unique

Limit

Theorem 2

## Proving a Limit Does Not Exist

For a single variable function, we often showed a limit did not exist by showing the left-hand and right-hand limit did not equal each other, and used the fact that the limit is supposed to be unique. For multivariable functions, we will essentially do the same thing, only now we have to remember that we can approach  $(a, b)$  along any smooth curve.

One can approach a question like this by taking the equation  $y = mx$  or  $x = my$  (for any real coefficient  $m$ ) and if the limit turns out to be dependent on  $m$ , then we know that the limit is not unique.

Though, this approach does not always work, as  $y = mx$  does not describe all the lines, (it cannot represent vertical lines).

Sometimes trying out several straight lines will give the same limit, but using a continuous curve will show that the limit in-fact does not exist. The trick to use here would be to choose a curve in such a way that (if the function is a fraction) the numerator and denominator cancel out. (e.g of forms  $y = mx^k$  or  $y = mx^{p/q}$  etc)

**Caution:** Be sure to use lines or curves that actually approach the limit point in question.

**Note:** Finding two paths that show that a limit does not exist does indeed mean that it doesn't exist. But being unable to find a contradictory value for a limit does not necessarily mean that a limit exists. We then use other methods such as the Squeeze theorem to test if this consistently occurring value of  $L$  is the actual limit or not.

## Proving a Limit Exists

If there exists a function  $B(x, y)$  such that

$$|f(x, y) - L| \leq B(x, y), \text{ for all } (x, y) \neq (a, b)$$

in some neighbourhood of  $(a, b)$  and  $\lim_{(x,y) \rightarrow (a,b)} B(x, y) = 0$  then

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y) = L$$

**Proof:**

(Our hypothesis says that  $B(x, y) \geq 0$  for all  $(x, y) \neq (a, b)$ )

Let  $\epsilon > 0$

Since  $\lim_{(x,y) \rightarrow (a,b)} B(x, y) = 0$ , by definition of limit, there exists  $\delta > 0$  such that

$$0 < \|(x, y) - (a, b)\| < \delta \implies |B(x, y) - 0| < \epsilon$$

Squeeze  
Theorem

Hence, if  $0 < \|(x, y) - (a, b)\| < \delta$ , then since we have

$$|f(x, y) - L| \leq B(x, y) = |B(x, y)| < \epsilon$$

as our hypothesis requires that  $B(x, y) \geq 0$  for all  $(x, y) \neq (a, b)$  in the neighbourhood of  $(a, b)$ . Therefore, by definition of a limit, we have

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y) = L$$

## Generalizations

The concept of neighbourhood, the definition of a limit, the Squeeze Theorem and limit theorems are all valid for scalar functions  $f(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$ . In fact, to generalise these concepts, one only needs to know that  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbf{a} = (a_1, \dots, a_n)$  are in  $\mathbb{R}^n$ , then the Euclidean distance from  $\mathbf{x}$  and  $\mathbf{a}$  is

$$\|\mathbf{x} - \mathbf{a}\| = \sqrt{(x_1 - a_1)^2 + \dots + (x_n - a_n)^2}$$

With this adjustment and some more rephrasing, the previous section can be generalised for  $\mathbb{R}^n$

## Appendix: Inequalities

**Trichotomy Property:** For any real numbers  $a$  and  $b$ , one and only one of the follow holds

$$a = b, \quad a < b, \quad b < a$$

**Transitivity Property:** If  $a < b$  and  $b < c$ , then  $a < c$

**Addition Property:** If  $a < b$  then for all  $c$ ,  $a + c < b + c$

**Multiplication Property:** If  $a < b$  and  $c < 0$ , then  $bc < ac$

**Absolute value** of a real number  $a$  is defined by

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0 \end{cases}$$

A few useful results

1.  $|a| = \sqrt{a^2}$
2.  $|a| < b \iff -b < a < b$
3. The Triangle Inequality:  $|a + b| \leq |a| + |b|$ ,  $\forall a, b \in \mathbb{R}$

When using the Squeeze Theorem, the most commonly used inequalities are:

1. Triangle Inequality
2. If  $c > 0$ , then  $a < a + c$
3. The cosine inequality  $2|x||y| \leq x^2 + y^2$

# Unit 3: Continuous Functions

## Definition of a Continuous Function

A quick review of the definition of a continuous function in one variable

A function of a single variable  $f(x)$  is continuous at  $x = a$  if and only if

Definition

Continuity

Single

Variable

Function

1.  $f$  is defined at  $x = a$
2.  $\lim_{x \rightarrow a} f(x)$  exists, which means that
  - (a)  $\lim_{x \rightarrow a^-} f(x)$  exists; and
  - (b)  $\lim_{x \rightarrow a^+} f(x)$  exists; and
  - (c)  $\lim_{x \rightarrow a^-} f(x) = \lim_{x \rightarrow a^+} f(x)$
3.  $\lim_{x \rightarrow a} f(x) = f(a)$

A function  $f(x, y)$  is **continuous** at  $(a, b)$  if and only if

Definition

Continuous

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y) = f(a, b)$$

Additionally, if  $f$  is continuous at every point in a set  $D \subset \mathbb{R}^2$ , then we say that  $f$  is continuous on  $D$

**Remark:** Just like in single variable calculus, there are three requirements in this definition:

1.  $\lim_{(x,y) \rightarrow (a,b)} f(x, y)$  exists
2.  $f$  is defined at  $(a, b)$ , and
3.  $\lim_{(x,y) \rightarrow (a,b)} f(x, y) = f(a, b)$

## Basic functions

To make the process of verifying if a function is continuous, we will employ the use of simpler or "basic" functions, which we know are continuous and view functions we inspect as being made up of these basic functions

In this course, we can take the continuity of these functions on their domain as a given

1. the constant function  $f(x, y) = k$
2. the power functions  $f(x, y) = x^n$ ,  $f(x, y) = y^n$
3. the logarithm function  $\ln(\cdot)$



4. the exponential function  $e^{(\cdot)}$
5. the trigonometric functions,  $\sin(\cdot)$ ,  $\cos(\cdot)$ , etc.
6. the inverse trigonometric functions,  $\arcsin(\cdot)$ , etc.
7. the absolute value function  $|\cdot|$

If  $f(x, y)$  and  $g(x, y)$  are scalar functions and  $(x, y) \in D(f) \cap D(g)$ , then:

Definition

1. the **sum**  $f + g$  is defined by

Operations  
on Functions

$$(f + g)(x, y) = f(x, y) + g(x, y)$$

2. the **product**  $fg$  is defined by

$$(fg)(x, y) = f(x, y)g(x, y)$$

3. the **quotient**  $\frac{f}{g}$  is defined by

$$\left(\frac{f}{g}\right)(x, y) = \frac{f(x, y)}{g(x, y)}, \text{ if } g(x, y) \neq 0$$

For scalar functions  $g(t)$  and  $f(x, y)$  the **composite function**  $g \circ f$  is defined by

Definition

$$(g \circ f)(x, y) = g(f(x, y))$$

Composite  
Functions

for all  $(x, y) \in D(f)$  for which  $f(x, y) \in D(g)$

**Remark:** When composing multivariable functions, it is very important to make sure that the range of the inner function is a subset of the domain of the outer function.

## Continuity Theorems

With basic functions and operations on functions discussed, we now state some theorems that will be of use. (Most proofs in course notes)

If  $f$  and  $g$  are both continuous at  $(a, b)$ , then  $f + g$  and  $fg$  are continuous at  $(a, b)$

Continuity  
Theorem 1

If  $f$  and  $g$  are both continuous at  $(a, b)$  and  $g(a, b) \neq 0$ , then the quotient  $f/g$  is continuous at  $(a, b)$

Continuity  
Theorem 2

If  $f(x, y)$  is continuous at  $(a, b)$  and  $g(t)$  is continuous at  $f(a, b)$ , then the composition  $g \circ f$  is continuous at  $(a, b)$

Continuity  
Theorem 3

# Week 3

## Unit 4: The Linear Approximation

### Partial Derivatives

A scalar function  $f(x, y)$  can be differentiated in two natural ways, by treating  $y$  as a constant and differentiating with respect to  $x$  to get  $\frac{\partial f}{\partial x}$  or treating  $x$  as constant and differentiating with respect to  $y$  to get  $\frac{\partial f}{\partial y}$ . These are called the (first) **partial derivatives** of  $f$

The **partial derivatives** of  $f(x, y)$  are defined by

Definition

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

Partial  
Derivatives

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

provided these limits exist

Sometimes it's convenient to use the **operator notation**  $D_{1f}$  and  $D_{2f}$  for the partial derivatives of  $f(x, y)$ , where  $D_{1f}$  means to differentiate w.r.t. the variable in the first position, holding the others fixed. Sometimes  $\frac{\partial f}{\partial x}(x, y)$  is simply written as  $\frac{\partial f}{\partial x}$

### Higher Order Partial Derivatives

Partial derivatives of a scalar function of two variables are also a scalar function of two variables, so we can take partial derivatives of the partial derivatives of any scalar function. There are four possible second partial derivatives of  $f$

$$\begin{aligned} \bullet \frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial x} \right) & \bullet \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial y} \right) \\ \bullet \frac{\partial^2 f}{\partial y \partial x} &= \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial x} \right) & \bullet \frac{\partial^2 f}{\partial y^2} &= \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial y} \right) \end{aligned}$$

**Remark:** It is often convenient to use the subscript notation or the operator notation:

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= f_{xx} = D_1^2 f, \quad \frac{\partial^2 f}{\partial y \partial x} = f_{xy} = D_2 D_1 f \\ \frac{\partial^2 f}{\partial x \partial y} &= f_{yx} = D_1 D_2 f, \quad \frac{\partial^2 f}{\partial y^2} = f_{yy} = D_2^2 f \end{aligned}$$

The **Hessian Matrix** of  $f(x, y)$ , denoted by  $Hf(x, y)$ , is defined as

Definition

$$Hf(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix}$$

Hessian  
matrix

You will notice that sometimes  $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$ , this is in fact a general property of partial derivatives, subject to a continuity requirement, as follows.

If  $f_{xy}$  and  $f_{yx}$  are defined in some neighbourhood of  $(a, b)$  and are both continuous at  $(a, b)$ , then

Clairaut's  
Theorem

$$f_{xy}(a, b) = f_{yx}(a, b)$$

We can take higher-order partial derivatives in the expected way.  $f(x, y)$  has eight third partial derivatives.

Clairaut's theorem also extends to higher-order partial derivatives: if the higher-order partial derivatives are defined in a neighbourhood of a point  $(a, b)$  and are continuous at  $(a, b)$ , then  $f_{i_1, \dots, i_k} = f_{j_1, \dots, j_k}$ , whenever  $(i_1, \dots, i_k)$  and  $(j_1, \dots, j_k)$  are tuples (ordered sets/sequences) of indices (variable symbols) which are arrangements of each other.

E.g., If the partial derivatives of  $f$  satisfy Clairaut's theorem, then

$$f_{xxy}(a, b) = f_{xyx}(a, b) = f_{yxx}(a, b)$$

In many situations, we will want to require that a function have continuous partial derivatives of some order. Some terminology;

If the  $k$ -th partial derivatives of  $f(x_1, \dots, x_n) = f(\mathbf{x})$  are continuous, then we write

$$f \in C^k$$

and say  $f$  is in class  $C^k$

Having  $f(x, y) \in C^2$ , for example, means that  $f$  has continuous second partial derivatives, and therefore, by Clairaut's theorem, that  $f_{xy} = f_{yx}$ . More generally,  $f(x, y) \in C^k$  means that  $f$  has continuous  $k$ -th partial derivatives and that the mixed higher-order partial derivatives are equal regardless of the order in which they are taken

## The Tangent Plane

The **tangent plane** to  $z = f(x, y)$  at the point  $(a, b, f(a, b))$  is

Definition

$$z = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b)$$

Tangent  
Plane

## Linear Approximation

In the One-Dimensional case, for a function  $f(x)$  the tangent line to  $y = f(x)$  at the point  $(a, f(a))$  is  $y = f(a) + f'(a)(x - a)$ . The function  $L_a$  defined by  $L_a(x) = f(a) + f'(a)(x - a)$  is called the **linearization** of  $f$  at  $a$  since  $L_a(x)$  approximates  $f(x)$  for sufficiently close to  $a$ . For  $x$  sufficiently close to  $a$ , the approximation  $f(x) \approx L_a(x)$ , is called the **linear approximation** of  $f$  at  $a$

For a multivariable function  $f(x, y)$ , we can use the tangent plane to approximate the surface  $z = f(x, y)$  near a point of tangency  $P(a, b, f(a, b))$ .

For a function  $f(x, y)$  we define the **linearization**  $L_{(a,b)}(x, y)$  of  $f$  at  $(a, b)$  by

$$L_{(a,b)}(x, y) = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b)$$

We call the approximation

$$f(x, y) \approx L_{(a,b)}(x, y)$$

the **linear approximation** of  $f(x, y)$  at  $(a, b)$

In the case we want to know the change in the value of  $f(x, y)$  due to a change  $(\Delta x, \Delta y)$  away from the point  $(a, b)$ , where  $\Delta x = x - a$  and  $\Delta y = y - b$ , we can manipulate the linear approximation to get the **increment form** of the linear approximation

$$\Delta f \approx \frac{\partial f}{\partial x}(a, b)\Delta x + \frac{\partial f}{\partial y}(a, b)\Delta y$$

## Linear Approximation in $\mathbb{R}^3$

By analogy with the case of a function with two variables we can define the linearization of a function  $f(x, y, z)$  at  $\mathbf{a} = (a, b, c)$  by

$$L_{\mathbf{a}}(x, y, z) = f(\mathbf{a}) + f_x(\mathbf{a})(x - a) + f_y(\mathbf{a})(y - b) + f_z(\mathbf{a})(z - c)$$

To simplify the notation we can represent the final three terms as the dot product of the vectors

$$(x - a, y - b, z - c) = (x, y, z) - (a, b, c), \text{ and } \nabla f(\mathbf{a}) = (f_x(\mathbf{a}), f_y(\mathbf{a}), f_z(\mathbf{a}))$$

since

$$(x - a, y - b, z - c) \cdot (f_x(\mathbf{a}), f_y(\mathbf{a}), f_z(\mathbf{a})) = f_x(\mathbf{a})(x - a) + f_y(\mathbf{a})(y - b) + f_z(\mathbf{a})(z - c)$$

The vector  $\nabla f(\mathbf{a})$  is called the **gradient** of  $f$  at  $\mathbf{a}$

Definition

Linearization  
and Linear  
Approxima-  
tion

Suppose that  $f(x, y, z)$  has partial derivatives at  $\mathbf{a} \in \mathbb{R}^3$ . The **gradient** of  $f$  at  $\mathbf{a}$  is defined by

Definition  
Gradient

$$\nabla f(\mathbf{a}) = (f_x(\mathbf{a}), f_y(\mathbf{a}), f_z(\mathbf{a}))$$

Suppose that  $f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^3$ , has partial derivatives at  $\mathbf{a} \in \mathbb{R}^3$ .

Definition  
Linearization,  
Linear Ap-  
proximation

The **linearization** of  $f$  at  $\mathbf{a}$  is defined by

$$L_{\mathbf{a}}(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$$

The **linear approximation** of  $f$  at  $\mathbf{a}$  is

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$$

## Generalization for $\mathbb{R}^n$

The advantage of using vector notation is that the equations for **linearization** and **linear approximation** hold for a function of  $n$  variables  $f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n$ . For an arbitrary vector  $\mathbf{a} \in \mathbb{R}^n$ , we have

$$\Delta \mathbf{x} = \mathbf{x} - \mathbf{a} = (x_1 - a_1, \dots, x_n - a_n)$$

And we define the gradient of  $f$  at  $\mathbf{a}$  to be

$$\nabla f(\mathbf{a}) = (D_1 f(\mathbf{a}), D_2 f(\mathbf{a}), \dots, D_n f(\mathbf{a}))$$

Then, the increment form of the linear approximation for  $f(\mathbf{x})$  is

$$\Delta f \approx \nabla f(\mathbf{a}) \cdot \Delta \mathbf{x}$$

This is a true generalization as using the formula for  $n = 1$  we find out familiar equation (in increment form)  $\Delta g \approx \nabla g(a) \cdot \Delta \vec{x} = g'(a)(x - a)$

And for  $n = 2$ , we get  $\Delta f \approx \nabla f(a, b) \cdot \Delta(x, y) = f_x(a, b)(x - a) + f_y(a, b)(y - b)$

# Week 4

## Unit 5: Differentiable Functions

### Definition of Differentiability

In the case of single variable functions a function  $g(x)$  was differentiable at  $x = a$  simply if  $g'(a)$  existed. But in the case of partial derivatives a partial derivative can exist at a point without it being continuous at that point.

To define the concept of differentiability in  $f(x, y)$  we want to ensure that it has the same properties as the definition in single variable calculus. The function  $g(x)$  was differentiable at  $x = a$ , then the graph of  $g(x)$  was 'smooth' at  $x = a$  (no cusps or jumps) and that the linear approximation was a good approximation. So let's define the error in the linear approximation to be

$$R_{1,a}(x) = g(x) - L_a(x)$$

then we get the theorem

If  $g'(a)$  exists, then  $\lim_{x \rightarrow a} \frac{|R_{1,a}(x)|}{|x-a|} = 0$  where

Theorem  
5.1.1

$$R_{1,a}(x) = g(x) - L_a(x) = g(x) - g(a) - g'(a)(x - a)$$

(The proof is short, simply place the definition of  $R_{1,a}(x)$  into the fraction and see that it simplifies to  $|g'(x) - g'(a)|$  as  $x \rightarrow a$ )

You will notice that Theorem 1 says that the error  $R_{1,a}(x)$  tends to zero faster than the displacement  $|x - a|$  (and hence, the limit goes to 0). We can also verify that the tangent line (linearization) is the best through  $(a, g(a))$  for the approximation, as the property (limit) in Theorem 1 only goes to 0 for the tangent line.

With the simple existence of partial derivatives not being enough to consider a multivariable function differentiable, we consider this property of the error to make deduce whether a function is differentiable.<sup>1</sup> So to get a definition similar to the one for single variables, for two variables we make the definition

A function  $f(x, y)$  is **differentiable** at  $(a, b)$  if

Definition  
Differentiable

$$\lim_{(x,y) \rightarrow (a,b)} \frac{|R_{1,(a,b)}(x, y)|}{\|(x, y) - (a, b)\|} = 0$$

where

$$R_{1,(a,b)}(x, y) = f(x, y) - L_{(a,b)}(x, y)$$

---

<sup>1</sup>I do understand how this will draw a connection between differentiability in single and multivariable functions, but I don't yet see how this is well defined. Perhaps it has some analysis behind it

If a function  $f(x, y)$  satisfies

Theorem  
5.1.2

$$\lim_{(x,y) \rightarrow (a,b)} \frac{|f(x, y) - f(a, b) - c(x - a) - d(y - b)|}{\|(x, y) - (a, b)\|} = 0$$

then  $c = f_x(a, b)$  and  $d = f_y(a, b)$

Just like before, this implies that the tangent plane gives the best linear approximation. Moreover, it tells us that the linear approximation is a 'good approximation' if and only if  $f$  is differentiable at  $(a, b)$

**Remark:** Observe that for the linear approximation to exist at  $(a, b)$  both partial derivatives of  $f$  must exist at  $(a, b)$ . (In order to have the linear approximation and hence the error). However, both partial derivatives existing does not guarantee that  $f$  will be differentiable (one way implication). We say that the partial derivatives of  $f$  existing at  $(a, b)$  is necessary, but not sufficient

Consider a function  $f(x, y)$  which is *differentiable* at  $(a, b)$ . The **tangent plane** of the surface  $z = f(x, y)$  at  $(a, b, f(a, b))$  is the graph of the linearization. That is, the tangent plane is given by

Definition  
Tangent  
Plane

$$z = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b)$$

Since  $f$  is assumed to be differentiable, the tangent plane best approximates the surface at  $(a, b, f(a, b))$  by Theorem 5.1.2. In this case, we say that at the point  $(a, b, f(a, b))$  the surface of  $z = f(x, y)$  is **smooth**

## Differentiability and Continuity

From single variable calculus we know that if a function  $g(x)$  is differentiable at  $x = a$ , then it is continuous at  $x = a$ . This is true for scalar functions  $f(x, y)$

If  $f(x, y)$  is differentiable at  $(a, b)$ , then  $f$  is continuous at  $(a, b)$

Theorem  
5.2.1

## Continuous Partial Derivatives and Differentiability

We now present a theorem that states that if the partial derivatives of  $f(x, y)$  are continuous at  $(a, b)$ , then  $f$  is differentiable at  $(a, b)$ .

For the proof of this theorem, we will need the Mean value theorem from single variable calculus

If  $f(t)$  is continuous on the closed interval  $[t_1, t_2]$  and  $f$  is differentiable on the open interval

Mean  
Value  
Theorem  
(5.3.1)

$(t_1, t_2)$ , then there exists  $t_0 \in (t_1, t_2)$  such that

$$f(t_2) - f(t_1) = f'(t_0)(t_2 - t_1)$$

If  $\frac{\partial f}{\partial x}$  and  $\frac{\partial f}{\partial y}$  are continuous at  $(a, b)$ , then  $f(x, y)$  is differentiable at  $(a, b)$

Theorem  
5.3.2

(Proof in course notes)

**Remark:** The converse is not true.

## Linear Approximation Revisited

The error in the linear approximation of  $f(x, y)$  is defined by

$$R_{1,(a,b)}(x, y) = f(x, y) - L_{(a,b)}(x, y)$$

which we can re-write as

$$f(x, y) = f(a, b) + \nabla f(a, b) \cdot (x - a, y - b) + R_{1,(a,b)}(x, y)$$

The linear approximation

$$f(x, y) \approx f(a, b) + \nabla f(a, b) \cdot (x - a, y - b)$$

for  $(x, y)$  sufficiently close to  $(a, b)$  arises when we ignore the error term. In general, we have no information about  $R_{1,(a,b)}(x, y)$  so we're uncertain if the approximation is reasonable. However, we can use Theorem 5.3.2 such that;

If we know that the partial derivatives of  $f$  are continuous at  $(a, b)$ , then  $f$  is differentiable, and hence,

$$\lim_{(x,y) \rightarrow (a,b)} \frac{|R_{1,(a,b)}(x, y)|}{\|(x, y) - (a, b)\|} = 0$$

and so, in such a case the approximation is reasonable for  $(x, y)$  sufficiently close to  $(a, b)$  and so we say that  $L_{(a,b)}(x, y)$  is a good approximation for  $f(x, y)$  near  $(a, b)$ .

Approximations are very important in calculus and the equation

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + R_{1,(a,b)}(\mathbf{x})$$

is of fundamental importance. We will revisit the error term in chapter 8 while discussing second partial derivatives



# Week 5

## Unit 6: Chain Rule

### Basic Chain Rule in Two Dimensions

#### Review of the Chain Rule for $f(x(t))$

Let  $T = f(x)$  be the temperature of a rod as a function of position  $x$ . An ant's position on the rod is given by  $x = x(t)$  as a function of time  $t$ . To get an expression for the rate of change of temperature w.r.t. time experienced by the ant.

$$T(t) = f(x(t))$$

$$T'(t) = f'(x(t))x'(t)$$

Re-writing in Leibniz notation

$$\frac{dT}{dt} = \frac{dT}{dx} \frac{dx}{dt}$$

Notice that we have defined  $T$  as a composite function of  $t$ , but we've also showed its derivative with  $x$ , writing it as a function of  $x$ . This is abuse of notation

#### Chain Rule for $f(x(t), y(t))$

Suppose the surface temperature of a pond is  $T = f(x, y)$  as a function of position  $(x, y)$ . A duck's position is given by  $x = x(t)$ ,  $y = y(t)$  as a function of time  $t$ . The temperature experienced by the duck as a function of  $t$  is

$$T(t) = f(x(t), y(t))$$

For a given change  $\Delta t$ ,  $x$  and  $y$  change by,

$$\Delta x = x(t + \Delta t) - x(t) , \Delta y = y(t + \Delta t) - y(t)$$

By the increment form of the linear approximation, the change in  $T$  corresponding to changes  $\Delta x$  and  $\Delta y$  is approximated by

$$\Delta T \approx \frac{\partial T}{\partial x} \Delta x + \frac{\partial T}{\partial y} \Delta y$$

for  $\Delta x$  and  $\Delta y$  sufficiently small. Dividing by  $\Delta t$ , and letting  $\Delta t \rightarrow 0$ , using the definition of the derivative to get  $\frac{dT}{dt}$  on the left side of the equation. Assuming  $T(x, y)$  is differentiable at  $(x, y)$ , then as  $\Delta x$  and  $\Delta y \rightarrow 0$ , the error in the linear approximation tends to zero, increasing the accuracy of the approximation, giving

$$\frac{dT}{dt} = \frac{\partial T}{\partial x} \frac{dx}{dt} + \frac{\partial T}{\partial y} \frac{dy}{dt}$$

Again, notice that  $T$  is mentioned as a composite function of  $t$ , and also implied to be a function of  $x$  and  $y$ . This is also abuse of notation, a form without it would be

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

like we did in the single variable case. We would define the composite function  $T$  by

$$T(t) = f(x(t), y(t))$$

and write

$$T'(t) = f_x(x(t), y(t))x'(t) + f_y(x(t), y(t))y'(t)$$

**Note:**  $f_x(x(t), y(t))$  is the partial derivative of  $f(x, y)$  w.r.t.  $x$ , evaluated at  $(x(t), y(t))$

Let  $G(t) = f(x(t), y(t))$ , and let  $a = x(t_0)$  and  $b = y(t_0)$ . If  $f$  is differentiable at  $(a, b)$  and  $x'(t_0)$  and  $y'(t_0)$  exist, then  $G'(t_0)$  exists and is given by

Theorem  
6.1.1  
Chain Rule

$$G'(t_0) = f_x(a, b)x'(t_0) + f_y(a, b)y'(t_0)$$

## Vector Form of the Basic Chain Rule

If we have,

$$T(t) = f(x(t), y(t))$$

where  $f(x, y)$ ,  $x(t)$ , and  $y(t)$  are differentiable then

$$\begin{aligned} \frac{dT}{dt} &= \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} \\ &= \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \cdot \left( \frac{dx}{dt}, \frac{dy}{dt} \right) \\ &= \nabla f \cdot \frac{d\mathbf{x}}{dt} \end{aligned}$$

So, we have

$$\frac{d}{dt}f(\mathbf{x}(t)) = \nabla f(\mathbf{x}(t)) \cdot \frac{d\mathbf{x}}{dt}(t)$$

with  $\mathbf{x}(t) = (x(t), y(t))$

In this vector form, the Chain Rule holds for any differentiable function  $f(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$

## Extensions of the Basic Chain Rule

When we had that  $u = f(x, y)$  with  $x = x(t)$ ,  $y = y(t)$ . The rate of change of  $u$  should be the sum of the rate of change with respect to its  $x$ -component and with respect to its  $y$ -component. The term  $\frac{\partial u}{\partial x} \frac{dx}{dt}$  calculates the rate of change of  $u$  with respect to those  $t$ 's that affect  $u$  through  $x$ . And similar for  $y$ .

$$\frac{du}{dt} = \frac{\partial u}{\partial x} \frac{dx}{dt} + \frac{\partial u}{\partial y} \frac{dy}{dt}$$

Now, if we have that  $x = x(s, t)$ ,  $y = y(s, t)$ , which have first order partial derivative at  $(s, t)$  and that  $u = f(x, y)$ , where  $f$  is differentiable at  $(x, y) = (x(s, t), y(s, t))$ . Since  $u$  is a function of two variables, we want to write a chain rule for  $\frac{\partial u}{\partial s}$  and  $\frac{\partial u}{\partial t}$

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial s}$$

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial t}$$

### Remarks:

- $\frac{\partial u}{\partial x}$  means: regard  $u$  as a given function of  $x$  and  $y$ , differentiate wrt  $x$  and hold  $y$  constant
- $\frac{\partial u}{\partial s}$  means: regard  $u$  as a composite function of  $s$  and  $t$ , differentiate wrt  $s$  and hold  $t$  constant.
- Equations of the form  $x = x(s, t)$  and  $y = y(s, t)$  can be thought of as defining a change of coordinates in 2-space
- To show the functional dependence, we may need a more precise form of the chain rule. Let  $g$  denote the composite function of  $f(x, y)$  and  $x = x(s, t)$ ,  $y = y(s, t)$

$$g(s, t) = f(x(s, t), y(s, t))$$

Then the first partial derivative equation w.r.t.  $s$  would be

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

and a similar equation for  $\frac{\partial g}{\partial t}(s, t)$

For a function describing the temperature of a pond wrt position and time,  $T = T(x, y, t)$ , where  $x = x(t)$  and  $y = y(t)$ , all of which are differentiable. The rate of change of temperature experienced by a duck swimming around would be

$$\frac{dT}{dt} = \frac{\partial T}{\partial x} \frac{dx}{dt} + \frac{\partial T}{\partial y} \frac{dy}{dt} + \frac{\partial T}{\partial t}$$

**Note:** It is important to distinguish between:

$\frac{dT}{dt}$ : the ordinary derivative of  $T$  as a composite function of  $t$

$\frac{\partial T}{\partial t}$ : the partial derivative of  $T$  as the given function of  $x, y, t$  with  $x, y$  held constant

To emphasize what variables are fixed, one may write

$$\left(\frac{\partial T}{\partial t}\right)_{x,y}$$

To prevent abuse of notation, i.e. using  $T$  to denote two different functions, one can write

$$T(t) = f(x(t), y(t), t)$$

So that  $T(t)$  describes the temperature at the duck's position at time  $t$  and  $f(x, y, t)$  is the temperature of the water at position  $(x, y)$  at time  $t$ . Then the chain rule is

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

or more concisely,

$$T'(t) = f_x x' + f_y y' + f_t$$

# Unit 7: Directional Derivatives and Gradient Vector

## Directional Derivatives

The **directional derivative** of  $f(x, y)$  at a point  $(a, b)$  in the direction of a unit vector  $\hat{u} = (u_1, u_2)$  is defined by

Definition  
Directional  
Derivative

$$D_{\hat{u}}f(a, b) = \left. \frac{d}{ds} f(a + su_1, b + su_2) \right|_{s=0}$$

provided the derivative exists

If  $f(x, y)$  is differentiable at  $(a, b)$  and  $\hat{u} = (u_1, u_2)$  is a unit vector, then

Theorem  
7.1.1

$$D_{\hat{u}}f(a, b) = \nabla f(a, b) \cdot \hat{u}$$

### Remarks:

- Be sure to check the condition of Theorem 1 before applying it. If  $f$  is not differentiable at  $(a, b)$  then we must apply the definition of the directional derivative
- If we choose  $\hat{u} = \hat{i} = (1, 0)$  or  $\hat{u} = \hat{j} = (0, 1)$  then the directional derivative is equal to the partial derivatives  $f_x$  or  $f_y$  respectively

## The Gradient Vector in Two Dimensions

### The Greatest Rate of Change

If  $f(x, y)$  is differentiable at  $(a, b)$  and  $\nabla f(a, b) \neq (0, 0)$ , then the largest value of  $D_{\hat{u}}f(a, b)$  is  $\|\nabla f(a, b)\|$ , and occurs when  $\hat{u}$  is in the direction of  $\nabla f(a, b)$

Theorem  
7.2.1

**Note:** Very short proof in notes

This theorem also applies in the general case

### The Gradient and the Level Curves of $f$

If  $f(x, y) \in C^1$  in a neighbourhood of  $(a, b)$  and  $\nabla f(a, b) \neq (0, 0)$ , then  $\nabla f(a, b)$  is orthogonal to the level curve  $f(x, y) = k$  through  $(a, b)$

Theorem  
7.2.2

### The Gradient Vector Field

Given a function  $f(x, y)$  that is differentiable at  $(x, y)$ , the gradient of  $f$  at  $(x, y)$  is defined by

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

The gradient of  $f$  associates a vector with each point of the domain of  $f$  and is referred to as a **vector field**. It is represented graphically by drawing  $\nabla f(a, b)$  as a vector emanating from the corresponding point  $(a, b)$

By Theorem 1 and 2, the gradient vector field has important geometric properties:

1. It gives the direction in which the the function has its largest rate of change.
2. It gives the direction that is orthogonal to the level curves of the function

**Remark:** Vector fields and gradient vectors will be studied in detail in Calculus 4 (AMATH 231)

## The Gradient Vector in Three Dimensions

If  $f(x, y, z) \in C^1$  in a neighbourhood of  $(a, b, c)$  and  $\nabla f(a, b, c) \neq (0, 0, 0)$ , then  $\nabla f(a, b, c)$  is orthogonal to the level surface of  $f(x, y, z) = k$  through  $(a, b, c)$  Theorem 7.3.1

A quick way to find a tangent plane to the surface is if  $\mathbf{x} \in \mathbb{R}^3$  is an arbitrary point in the tangent plane to the surface at point  $\mathbf{a} \in \mathbb{R}^3$ , then the vector  $\mathbf{x} - \mathbf{a}$  lies in the tangent plane, and by Theorem 1, is orthogonal to  $\nabla f(\mathbf{a})$ , leading to

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

# Unit 8: Taylor Polynomials and Taylor's Theorem

For a single variable function  $f$ , the second derivative  $f''$  is important for approximating  $f$ . Since it geometrically tells if the graph is concave up or concave down, so if it is concave up, we know that the linear approximation is an underestimate, and vice versa. The second derivative can in fact be used to estimate the error through the Taylor's formula. If we define a quadratic approximation, the second degree Taylor polynomial, then the linear approximation becomes more accurate.

## The Taylor Polynomial of Degree 2

### The 1-D case

For a single variable function  $f(x)$ , the second degree Taylor polynomial at  $a$  is

$$P_{2,a}(x) = f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2$$

$P_{2,a}(x)$  is the sum of the linear approximation  $L_a(x)$  and a term which is of second degree in  $(x - a)$ . The coefficient of this term is determined by requiring that the second derivative of  $P_{2,a}(x)$  equals the second derivative of  $f$  at  $a$ :

$$P_{2,a}''(a) = f''(a)$$

### The 2-D case

Suppose that  $f(x, y)$  has continuous second partial derivatives  $(a, b)$ . Denoting the second degree Taylor polynomial of  $f$  at  $(a, b)$  by  $P_{2,(a,b)}(x, y)$ , we obtain the polynomial by adding appropriate 2nd degree terms in  $(x - a)$  and  $(y - b)$  to the linear approximation  $L_{(a,b)}(x, y)$ . Consider

$$P_{2,(a,b)}(x, y) = L_{(a,b)}(x, y) + A(x - a)^2 + B(x - a)(y - b) + C(y - b)^2$$

where  $A, B, C$  are constants. We find that

$$\frac{\partial^2 P_{2,(a,b)}}{\partial x^2} = 2A$$

since  $L_{(a,b)}(x, y)$  does not contribute to the second derivatives, it is of first degree. Similarly, finding the other second partial derivatives of  $P_{2,(a,b)}(x, y)$  gives

$$\frac{\partial^2 P_{2,(a,b)}}{\partial x \partial y} = B$$

$$\frac{\partial^2 P_{2,(a,b)}}{\partial y^2} = 2C$$

Requiring that the second partial derivatives of  $P_{2,(a,b)}$  equal the second partial derivatives of  $f(x, y)$  at  $(a, b)$  gives

$$2A = \frac{\partial^2 f}{\partial x^2}(a, b), \quad B = \frac{\partial^2 f}{\partial x \partial y}(a, b), \quad 2C = \frac{\partial^2 f}{\partial y^2}(a, b)$$

The **second degree Taylor polynomial**  $P_{2,(a,b)}$  of  $f(x, y)$  at  $(a, b)$  is given by

Definition  
2nd Degree  
Taylor  
Polynomial

$$P_{2,(a,b)}(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) + \frac{1}{2} \left[ f_{xx}(a, b)(x - a)^2 + 2f_{xy}(a, b)(x - a)(y - b) + f_{yy}(a, b)(y - b)^2 \right]$$

So in general, it approximates  $f(x, y)$  for  $(x, y)$  sufficiently close to  $(a, b)$

$$f(x, y) \approx P_{2,(a,b)}(x, y)$$

with a better accuracy than the linear approximation

## Taylor's Formula with Second Degree Remainder

### Review of the 1-D case

If  $f''(x)$  exists on  $[a, x]$ , then there exists a number  $c \in (a, x)$  such that

Theorem  
8.2.1

$$f(x) = f(a) + f'(a)(x - a) + R_{1,a}(x)$$

where

$$R_{1,a}(x) = \frac{1}{2} f''(c)(x - a)^2$$

### The 2-D case

#### (Taylor's Theorem)

Theorem  
8.2.2

If  $f(x, y) \in C^2$  in some neighbourhood  $N(a, b)$  of  $(a, b)$ , then for all  $(x, y) \in N(a, b)$  there exists a point  $(c, d)$  on the line segment joining  $(a, b)$  and  $(x, y)$  such that

$$f(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) + R_{1,(a,b)}(x, y)$$

where

$$R_{1,(a,b)}(x, y) = \frac{1}{2} \left[ f_{xx}(c, d)(x - a)^2 + 2f_{xy}(c, d)(x - a)(y - b) + f_{yy}(c, d)(y - b)^2 \right]$$

**Remark:** Like the one variable case, Taylor's Theorem for  $f(x, y)$  is an existence theorem. That is, it only tells us that the point  $(c, d)$  exists, but not how to find it.

If  $f(x, y) \in C^2$  in some closed neighbourhood  $N(a, b)$  of  $(a, b)$ , then there exists a positive

Corollary  
8.2.3



constant  $M$  such that

$$|R_{1,(a,b)}(x, y)| \leq M \|(x, y) - (a, b)\|^2, \text{ for all } (x, y) \in N(a, b)s$$

## Generalizations

To define the Taylor polynomial  $P_{k,(a,b)}(x, y)$  of degree  $k$ , in a concise manner, we introduce the differential operator

$$(x - a)D_1 + (y - b)D_2$$

where  $D_1 = \frac{\partial}{\partial x}$  and  $D_2 = \frac{\partial}{\partial y}$  are the partial derivative operators. Then, we formally write

$$[(x - a)D_1 + (y - b)D_2]^2 = (x - a)^2 D_1^2 + 2(x - a)(y - b)D_1 D_2 + (y - b)^2 D_2^2$$

Note  $D_1^2 = D_1 D_1$ , which means to apply  $D_1$  twice, taking the second partial derivative wrt to the first variable

Using this notation, the first degree Taylor polynomial  $P_{1,(a,b)}(x, y)$ , the linear approximation, is written as

$$P_{1,(a,b)}(x, y) = f(a, b) + [(x - a)D_1 + (y - b)D_2]f(a, b)$$

and the second degree polynomial is written as

$$P_{2,(a,b)} = P_{1,(a,b)}(x, y) + \frac{1}{2!}[(x - a)D_1 + (y - b)D_2]^2 f(a, b)$$

for  $k = 2, 3, \dots$  we recursively define the  $k$ th degree Taylor polynomial by

$$P_{k,(a,b)}(x, y) = P_{k-1,(a,b)}(x, y) + \frac{1}{k!}[(x - a)D_1 + (y - b)D_2]^k f(a, b)$$

### Taylor's Theorem of order $k$

Theorem

If  $f(x, y) \in C^{k+1}$  at each point on the line segment joining  $(a, b)$  and  $(x, y)$ , then there exists a point  $(c, d)$  on the line segment between  $(a, b)$  and  $(x, y)$  such that

8.3.1

$$f(x, y) = P_{k,(a,b)}(x, y) + R_{k,(a,b)}(x, y)$$

where

$$R_{k,(a,b)}(x, y) = \frac{1}{(k+1)!}[(x - a)D_1 + (y - b)D_2]^{k+1} f(c, d)$$

If  $f(x, y) \in C^k$  in some neighbourhood of  $(a, b)$ , then

Corollary

$$\lim_{(x,y) \rightarrow (a,b)} \frac{|f(x, y) - P_{k,(a,b)}(x, y)|}{\|(x, y) - (a, b)\|^k} = 0$$

8.3.2

If  $f(x, y) \in C^{k+1}$  in some closed neighbourhood  $N(a, b)$  of  $(a, b)$ , then there exists a constant  $M > 0$  such that Corollary 8.3.3

$$|f(x, y) - P_{k,(a,b)}(x, y)| \leq M \|(x, y) - (a, b)\|^{k+1}$$

To completely generalize, we consider functions of  $n$  variables,  $f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n$ . We just need to replace

$$[(x - a)D_1 + (y - b)D_2]$$

with

$$[(x_1 - a_1)D_1 + \cdots + (x_n - a_n)D_n]$$

Letting  $\nabla = (D_1, \dots, D_n)$ , we can write this concisely in vector notation as

$$\left[ (\mathbf{x} - \mathbf{a}) \cdot \nabla \right]$$

# Unit 9: Critical Points

In single variable calculus we had that  $x = a$  is a local extremum of  $f(x)$  if either  $f'(a) = 0$  or  $f'(a)$  does not exist. These points are called critical points of  $f$ . Critical points are not necessarily local extrema.

## Local Extrema and Critical Points

A point  $(a, b)$  is a **local maximum point** of  $f$  if  $f(x, y) \leq f(a, b)$  for all  $(x, y)$  in some neighbourhood of  $(a, b)$

Definition

Local  
Maximum,

A point  $(a, b)$  is a **local minimum point** of  $f$  if  $f(x, y) \geq f(a, b)$  for all  $(x, y)$  in some neighbourhood of  $(a, b)$

Local  
Minimum

Geometrically, if  $(a, b)$  is a local maximum/minimum of  $f$  and  $f$  has continuous partial derivatives, then  $(a, b)$  is a local extremum of the cross-sections  $f(x, b)$  and  $f(a, y)$ . And so both partial derivatives of  $f$  will be zero and the tangent plane will be horizontal.

If  $(a, b)$  is a local maximum or minimum point of  $f$ , then

Theorem

9.1.1

$$f_x(a, b) = 0 = f_y(a, b)$$

or at least one of  $f_x$  or  $f_y$  does not exist at  $(a, b)$

A point  $(a, b)$  in the domain of  $f(x, y)$  is called a **critical point** of  $f$  if  $\frac{\partial f}{\partial x}(a, b) = 0$  or  $\frac{\partial f}{\partial x}(a, b)$  does not exist, and  $\frac{\partial f}{\partial y}(a, b) = 0$  or  $\frac{\partial f}{\partial y}(a, b)$  does not exist

Definition

Critical

Point

A critical point  $(a, b)$  of  $f(x, y)$  is called a **saddle point** of  $f$  if in every neighbourhood of  $(a, b)$  there exist points  $(x_1, y_1)$  and  $(x_2, y_2)$  such that

Definition

Saddle Point

$$f(x_1, y_1) > f(a, b) \text{ and } f(x_2, y_2) < f(a, b)$$

To find critical points, one may set the two partial derivatives equal to zero and solve for all possible critical points. And categorize them into either local minima, local maxima, or saddle points.

## The Second Derivative Test

### Review of the 1-D case

For a single variable function  $f(x)$ , the second degree Taylor polynomial approximation is

$$f(x) \approx f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2$$

for  $x$  sufficiently close to  $a$ . If  $x = a$  is a critical point of  $f$ , then  $f'(a) = 0$ , and the approximation can be rewritten as

$$f(x) - f(a) \approx \frac{1}{2}f''(a)(x - a)^2$$

Thus, for  $x$  sufficiently close to  $a$ ,  $f(x) - f(a)$  has the same sign as  $f''(a)$ . If  $f''(a) > 0$  then  $f(x) - f(a) > 0$  for  $x$  sufficiently close to  $a$  and  $x = a$  is a local minimum point. If  $f''(a) < 0$ , then  $f(x) - f(a) < 0$  for  $x$  sufficiently close to  $a$  and  $x = a$  is a local maximum point. There is no conclusion if  $f''(a) = 0$

### The 2-D case

For  $f(x, y) \in C^2$ , the second degree Taylor polynomial approximation is

$$f(x, y) \approx P_{2,(a,b)}(x, y)$$

for  $(x, y)$  sufficiently close to  $(a, b)$ . If  $(a, b)$  is a critical point of  $f$  such that

$$f_x(a, b) = 0 = f_y(a, b)$$

then the approximation can be re-arranged to yield

$$f(x, y) - f(a, b) \approx \frac{1}{2}[f_{xx}(a, b)(x - a)^2 + 2f_{xy}(a, b)(x - a)(y - b) + f_{yy}(a, b)(y - b)^2]$$

for  $(x, y)$  sufficiently close to  $(a, b)$ . The sign of the right hand side will determine the sign of the left hand side, and hence whether  $(a, b)$  is a local maximum or local minimum or a saddle point.

The expression on the right is called a **quadratic form**, and at this stage it is necessary to discuss some properties of these objects. (We will return to the equation above)

### Quadratic Forms

A function  $Q$  of the form

$$Q(u, v) = a_{11}u^2 + 2a_{12}uv + a_{22}v^2$$

where  $a_{11}, a_{12}$ , and  $a_{22}$  are constants, is called a **quadratic form on  $\mathbb{R}^2$** . One can use matrix notation to write

$$Q(u, v) = \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

so that a quadratic form on  $\mathbb{R}^2$  is determined by a  $2 \times 2$  matrix

We classify quadratic forms on  $\mathbb{R}^2$  in the following way:

1. If  $Q(u, v) > 0$  for all  $(u, v) \neq (0, 0)$ , then  $Q(u, v)$  is said to be **positive definite**
2. If  $Q(u, v) < 0$  for all  $(u, v) \neq (0, 0)$ , then  $Q(u, v)$  is said to be **negative definite**

Definition

Quadratic

Form

3. If  $Q(u, v) < 0$  for some  $(u, v) \neq (0, 0)$  and  $Q(u, v) > 0$  for some  $(u, v) \neq (0, 0)$ , then  $Q(u, v)$  is said to be **indefinite**.
4. If  $Q(u, v)$  does not satisfy any of 1 – 3, then  $Q(u, v)$  is said to be **semidefinite**

These terms are also used to describe the corresponding symmetric matrices

**Remark:** Semidefinite quadratic forms may be split into two classes, positive semidefinite and negative semidefinite. For example,  $Q(u, v) = 2u^2 \geq 0$  for all  $(u, v)$ , and  $Q(0, v) = 0$  for all  $v$ .

If the matrix of the quadratic form is a diagonal matrix, the nature is easy to deduce, however, when it is not diagonal, one may need to do some algebra to classify it (usually, completing the square). (Even if all terms of the matrix are positive, it is still not necessarily a positive definite matrix)

Now that quadratic forms have been introduced, we return to the equation earlier. Let

$$u = x - a, \quad v = y - b$$

so that

$$f(x, y) - f(a, b) \approx \frac{1}{2} [f_{xx}(a, b)u^2 + 2f_{xy}(a, b)uv + f_{yy}(a, b)v^2]$$

The matrix of the quadratic form on the right is the Hessian matrix of  $f$  at  $(a, b)$  :

$$Hf(a, b) = \begin{bmatrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{xy}(a, b) & f_{yy}(a, b) \end{bmatrix}$$

It is thus plausible if  $Hf(a, b)$  is positive definite, then

$$f(x, y) - f(a, b) > 0$$

for all  $(u, v) \neq (0, 0)$  i.e. for all  $(x, y) \neq (a, b)$  (assuming that  $(x, y)$  is sufficiently close to  $(a, b)$  so that the approximation is sufficiently accurate).

## Second Partial Derivative Test

Theorem

Suppose that  $f(x, y) \in C^2$  in some neighbourhood of  $(a, b)$  and that

9.2.1

$$f_x(a, b) = 0 = f_y(a, b)$$

1. If  $Hf(a, b)$  is positive definite, then  $(a, b)$  is a local minimum point of  $f$
2. If  $Hf(a, b)$  is negative definite, then  $(a, b)$  is a local maximum point of  $f$
3. If  $Hf(a, b)$  is indefinite, then  $(a, b)$  is a saddle point of  $f$ .

**Remark:** Note the analogy with the second derivative test for single variable functions. The requirement that  $g''(a) > 0$ , which implies a local minimum, is replaced by the requirement that the matrix of the second partial derivatives  $Hf(a, b)$  be positive definite.

If  $Q(u, v) = a_{11}u^2 + 2a_{12}uv + a_{22}v^2$  and  $D = a_{11}a_{22} - a_{12}^2$ , then

Theorem  
9.2.2

1.  $Q$  is positive definite iff  $D > 0$  and  $a_{11} > 0$
2.  $Q$  is negative definite iff  $D > 0$  and  $a_{11} < 0$
3.  $Q$  is indefinite iff  $D < 0$
4.  $Q$  is semidefinite iff  $D = 0$

Notice that  $D$  is the determinant of the associated symmetric matrix

**Remark:** Another way of classifying the Hessian matrix is by finding its eigenvalues. If they are all positive then it is positive definite, if they are negative then it is negative definite, and indefinite if it has both positive and negative eigenvalues.

### Degenerate Critical Points

The second partial derivative test gives conclusions for positive definite, negative definite, and indefinite cases but not the semidefinite case. If  $Hf(a, b)$  is semidefinite, the critical point  $(a, b)$  may be a local maximum or minimum point, or a saddle point.

If  $Hf(a, b)$  is semidefinite, so that the second partial derivative test gives no conclusion, we say that the critical point  $(a, b)$  is **degenerate**. So in order to classify the critical point, one has to investigate the sign of  $f(x, y) - f(a, b)$  in a small neighbourhood of  $(a, b)$

### Generalizations

The definitions of local maximum point, local minimum point, and critical point can be generalized for functions  $f$  of  $n$  variables. The Hessian matrix of  $f$  at  $\mathbf{a}$  is the  $n \times n$  symmetric matrix given by

$$Hf(\mathbf{a}) = \left[ \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) \right]$$

where  $i, j = 1, 2, \dots, n$ . The Hessian matrix can be classified as positive definite, negative definite, indefinite or semidefinite by considering the associated quadratic form in  $\mathbb{R}^n$ :

$$Q(\mathbf{u}) = \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) u_i u_j$$

as in  $\mathbb{R}^2$ . The second partial derivative test as stated in  $\mathbb{R}^2$  now holds in  $\mathbb{R}^n$ . It can be justified heuristically by using the second degree Taylor polynomial approximation,

$$f(\mathbf{x}) \approx P_{2,\mathbf{a}}(\mathbf{x})$$

which leads to

$$f(\mathbf{x}) - f(\mathbf{a}) \approx \frac{1}{2!} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) (x_i - a_i)(x_j - a_j)$$

## Level Curves Near Critical Points

Consider a function  $f \in C^2$ . In Unit 7.2 we discussed the fact that if  $\nabla f(a, b) \neq (0, 0)$ , then the level curve through  $(a, b)$  is a smooth curve (at least sufficiently close to  $(a, b)$ ). Also, by continuity,  $\nabla f(x, y) \neq (0, 0)$  for all  $(x, y)$  in some neighbourhood of  $(a, b)$ . This, if  $\nabla f(a, b) \neq (0, 0)$ , there will be some neighbourhood of  $(a, b)$  in which the level curves of  $f$  are smooth non-intersecting curves.

A point at which  $\nabla f(a, b) \neq (0, 0)$  is called a **regular point** of  $f$

Assuming  $f$  has continuous second partial derivatives, we can approximate  $f$  by its Taylor polynomial  $P_{2,(a,b)}(x, y)$ , calculated at the critical point

$$f(x, y) \approx f(a, b) + \frac{1}{2} [f_{xx}(a, b)(x - a)^2 + 2f_{xy}(a, b)(x - a)(y - b) + f_{yy}(a, b)(y - b)^2]$$

The constant terms  $f(a, b)$  and factor  $1/2$  do not significantly effect the shape of the level curves, so we can approximate the level curves of  $f$  by the level curves of  $P_{2,(a,b)}(x, y)$  for  $(x, y)$  sufficiently close to  $(a, b)$ . By using the translation  $u = x - a$  and  $v = y - b$  we get the quadratic form

$$Q(u, v) = a_{11}u^2 + 2a_{12}uv + a_{22}v^2$$

Therefore, to approximate the level curves of  $f$  near a critical point, we can sketch the level curves of the associated quadratic form  $Q(u, v)$ . The matter of sketching quadratic forms is covered in MATH 235, since it requires more Linear Algebra.

## Convex Functions

### 1-D Case

We say a twice differentiable function  $f(x)$  is **strictly convex** if  $f''(x) > 0$  for all  $x$  and  $f$  is **convex** if  $f''(x) \geq 0$  for all  $x$ . Thus the term convex means *concave up*.

If  $f(x)$  is twice continuously differentiable and strictly convex, then

Theorem

1.  $f(x) > L_a(x) = f(a) + f'(a)(x - a)$  for all  $x \neq a$ , for any  $a \in \mathbb{R}$
2. For  $a < b$ ,  $f(x) < f(a) + \frac{f(b)-f(a)}{b-a}(x - a)$  for  $x \in (a, b)$

9.2.3

**Remark:** (1) says that the graph of  $f$  lies above any tangent line, and (2) says that any secant line lies above the graph of  $f$

### 2-D Case

Suppose  $f(x, y)$  has continuous second partial derivatives. We say that  $f$  is **strictly convex** if  $Hf(x, y)$  is positive definite for all  $(x, y)$ . By Theorem 2,  $f$  is strictly convex means  $f_{xx} > 0$  and  $f_{xx}f_{yy} - f_{xy}^2 > 0$  for all  $(x, y)$ . We get a result that is analogous to the previous theorem

If  $f(x, y)$  has continuous second partial derivatives and is strictly convex, then

Theorem  
9.2.4

1.  $f(x, y) > L_{(a,b)}(x, y)$  for all  $(x, y) \neq (a, b)$ , and
2.  $f(a_1 + t(b_1 - a_1), a_2 + t(b_2 - a_2)) < f(a_1, a_2) + t[f(b_1, b_2) - f(a_1, a_2)]$   
for  $0 < t < 1, (a_1, a_2) \neq (b_1, b_2)$

**Remark:** (1) says that the graph of  $f$  lies above the tangent plane and (2) says that the cross section of the graph of  $f$  above the line segment from  $(a_1, a_2)$  to  $(b_1, b_2)$  lies below the secant line.

If  $f(x, y) \in C^2$  is strictly convex and has a critical point  $(c, d)$ , then  $f(x, y) > f(c, d)$  for all  $(x, y) \neq (c, d)$  and  $f$  has no other critical point.

Theorem  
9.2.5

## Proof of the Second Partial Derivative Test

Refer to the course notes to see the proof, I have only mentioned the lemma used in it.

Let  $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$  be a positive definite matrix. If  $|\tilde{a} - a|, |\tilde{b} - b|$  and  $|\tilde{c} - c|$  are sufficiently small, then

Lemma  
9.3.1

$\begin{bmatrix} \tilde{a} & \tilde{b} \\ \tilde{b} & \tilde{c} \end{bmatrix}$  The lemma is also true if positive definite is replaced with negative definite or indefinite.