Math 237 Notes

Syed Mustafa Raza Rizvi

August 10, 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 Professor 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 (https://github.com/meowstafa)

Week 1

Unit 1: Graphs of Scalar Functions

Scalar Functions

A review of basic vocabulary about functions in general

• A function $f : A \to B$ 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 $f(x_1, ..., x_n)$ which map subsets of $\mathbb{R}^n$ to $\mathbb{R}$

A **scalar function** $f(x_1, ..., x_n)$ 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 Function

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

Geometric Interpretation of $\mathbf{z = f(x, y)}$

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: Level Curves

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

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 cetain 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

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

Definition
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 parameterzing $f(x,y,z)$

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

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

Definition
Level Sets

So a level set is a generalisation of what we learned for level curves (case of $f:\mathbb{R}^2\to\mathbb{R}$) and level surfaces (case of $f:\mathbb{R}^3\to\mathbb{R}$) for $f:\mathbb{R}^n\to\mathbb{R}$
∘ For an f defined by: $f(x_1,...,x_n)=x_1^2+...+x_n^2$, 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 $\lim_{x\to a} f(x) = L$ to mean that the values of $f(x)$ can be made arbitrarily close to L by taking x sufficiently close to a.
More precisely,

Definition
Single Variable definition of a Limit

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

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

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

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

Definition
Open Interval

An **open interval** is defined as

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

where $r \in \mathbb{R}$

Definition
Euclidian Distance

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

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

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

Definition
Neighbourhood

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

`(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

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

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

Limit
Theorem 1

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

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

3. $$\lim_{(x,y)\to(a,b)} \frac{f(x,y)}{g(x,y)} = \frac{\lim\limits_{(x,y)\to(a,b)} f(x,y)}{\lim\limits_{(x,y)\to(a,b)} g(x,y)},\ \textit{provided}\ \lim_{(x,y)\to(a,b)} g(x,y) \neq 0$$

If $\lim_{(x,y)\to(a,b)} f(x,y)$ 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 occuring value of L is the actual limit or not.

Proving a Limit Exists

Squeeze Theorem

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 $\lim_{(x,y)\to(a,b)} B(x, y) = 0$ then

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

Proof:
(Our hypothesis says that $B(x, y) \geq 0$ for all $(x, y) \neq (a, b)$)
Let $\epsilon > 0$
Since $\lim_{(x,y)\to(a,b)} B(x, y) = 0$, by definition of limit, there exists $\delta > 0$ such that

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

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

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

Generalizations

The concept of neighbourhood, the definition of a limit, the Squeeze Theorm 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 $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{a} = (a_1, ..., a_n)$ are in $\mathbb{R}^n$, then the Euclidean distance from $\mathbf{x}$ and $\mathbf{a}$ is

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

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

Definition
Continuity
Single
Variable
Function

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

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

Definition
Continuous

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

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

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

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 trignometric functions, $\sin(\cdot), \cos(\cdot)$, etc.
6. the inverse trigonometric functions, $\arcsin(\cdot)$, etc.
7. the absolute value function $|\cdot|$

Definition Operations on Functions

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

1. the **sum** $f+g$ is defined by
$$(f+g)(x,y) = f(x,y) + g(x,y)$$
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
$$\left(\frac{f}{g}\right)(x,y) = \frac{f(x,y)}{g(x,y)}, \text{ if } g(x,y) \neq 0$$

Definition Composite Functions

For scalar functions $g(t)$ and $f(x,y)$ the **composite function** $g \circ f$ is defined by
$$(g \circ f)(x,y) = g(f(x,y))$$
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 thorems that will be of use. (Most proofs in course notes)

Continuity Theorem 1

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

Continuity Theorem 2

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 3

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)

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

Definition
Partial Derivatives

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

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

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

provided these limits exist

Sometimes it's convenient to use the **operator notation** $D_1 f$ and $D_2 f$ for the partial derivatives of $f(x, y)$, where $D_1 f$ 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

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

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

$$\frac{\partial^2 f}{\partial x^2} = f_{xx} = D_1^2 f \ , \ \frac{\partial^2 f}{\partial y \partial x} = f_{xy} = D_2 D_1 f$$

$$\frac{\partial^2 f}{\partial x \partial y} = f_{yx} = D_1 D_2 f \ , \ \frac{\partial^2 f}{\partial y^2} = f_{yy} = D_2^2 f$$

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

Definition
Hessian matrix

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

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 parital 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 $f_{i_1,\ldots,i_k} = f_{j_1,\ldots,j_k}$, whenever $(i_1,\ldots,i_k)$ and $(j_1,\ldots,j_k)$ are tuples (ordered sets/sequences) of indices (variable symbols) which are arrangements of each other.

E.g., If the partial derivatives of f satisfy Caliraut'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 $f(x_1,\ldots,x_n) = f(\mathbf{x})$ 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
Tangent Plane

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

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 $P(a, b, f(a, b))$.

Definition Linearization and Linear Approximation

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

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

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 $\mathbf{a}$

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 $\mathbf{a} \in \mathbb{R}^3$.
The **linearization** of f at $\mathbf{a}$ is defined by

Definition
Linearizaton, Linear Approximation

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

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

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

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

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

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 $\lim_{x\to a} \frac{|R_{1,a}(x)|}{|x-a|} = 0$ where

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

Theorem 5.1.1

(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 $x \to a$)

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 wether a function is differentiable.[1] 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

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

where

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

Definition Differentiable

[1] 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

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

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

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

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,

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

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 pont 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 Δt, 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 Δx and Δy is approximated by

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

for Δx and Δy sufficiently small. Dividing by Δt, and letting $\Delta t \to 0$, 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 Δx and $\Delta y \to 0$, 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 fuction 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

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

provided the derivative exists

Definition Directional Derivative

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

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

Theorem 7.1.1

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 approxmation 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

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

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

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

Definition 2nd Degree Taylor Polynomial

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

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

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

Theorem 8.2.1

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

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

where

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

The 2-D case

Theorem 8.2.2

(Taylor's Theorem)

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 existance theorem. That is, it only tells us that the point (c,d) exists, but not how to find it.

Corollary 8.2.3

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

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

$$[(x-a)D_1 + (y-b)D_2]^2 = (x-a)^2D_1^2 + 2(x-a)(y-b)D_1D_2 + (y-b)^2D_2^2$$

Note $D_1^2 = D_1D_2$, 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 $k = 2, 3, \ldots$ we recursively define the kth 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 8.3.1

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

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

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 8.3.2

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

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

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

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, Local Minimum

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)

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

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

for (x, y) sufficiently close to (a, b). The sigh 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

Definition
Quadratic Form

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 2×2 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**

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

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

so that

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

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

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

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 9.2.1

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

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

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 $f(x,y) - f(a,b)$ is 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 $i, j = 1, 2, \ldots, n$. 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 it;s Taylor polynomial $P_{2,(a,b)}(x,y)$, calculated at the critical point

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

The constant terms $f(a,b)$ and factor $1/2$ do not signifcantly 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 9.2.3

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 $x \in (a,b)$

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 $\begin{bmatrix} \tilde{a} & \tilde{b} \\ \tilde{b} & \tilde{c} \end{bmatrix}$ The lemma is also true if positive definite is replaced with negative definite or indefinite. Lemma 9.3.1

Unit 10: Optimization Problems

The Extreme Value Theorem

Given a function $f(x,y)$ and a set $S \subseteq \mathbb{R}^2$

Definition
Absolute Minimum and Maxmimum

1. a point $(a,b) \in S$ is an **absolute maximum point** of f on S if

$$f(x,y) \leq f(a,b) \quad \text{for all } (x,y) \in S$$

The value $f(a,b)$ is called the **absolute maximum value** of f on S

2. a point $(a,b) \in S$ is an **absolute minimum point** of f on S if

$$f(x,y) \geq f(a,b) \quad \text{for all } (x,y) \in S$$

The value $f(a,b)$ is called the **absolute minimum value** of f on S

The Extreme Value Theorem

The existence of a maximum/minimum value of f on S depends on f and on the set S. In single variable calculus, the Extreme Value Theorem gave conditions that implied the existence of a maximum and minimum value of f on an interval I

(The Extreme Value Theorem)
If $f(x)$ is continuous on a finite closed interval I, then there exists $c_1, c_2 \in I$ such that

Theorem 10.1.1

$$f(c_1) \leq f(x) \leq f(c_2) \quad \text{for all } x \in I$$

In order to generalize this theorem to two variable functions, we need to generalize the concept of a finite closed interval to sets in $\mathbb{R}^2$

A set $S \subset \mathbb{R}^2$ is siad to be bounded if and only if it is contained in some neighbourhood of the origin

Definition
Bounded Set

The definition implies that every point in S must have a finite distance from the origin.
The origin is just a point that we know to be $\mathbf{0}$, there isn't anything special about it in this case, it has just been selected because it's a point with a name. Any other point can be used to replace the origin in the definition.

Given a set $S \subset \mathbb{R}^2$, a point $(a,b) \in \mathbb{R}^2$ is said to be a **boundary point** of S if and only if every neighbourhood of (a,b) contains at least one point in S and one point not in S

Definition
Boundary Point

The set $B(S)$ of all boundary points of S is called the **boundary** of S

Definition
Boundary of S

A set $S \subseteq \mathbb{R}^2$ is said to be **closed** if S contains all of its boundary points

Definition
Closed Set

Now, for the generalization of the Extreme value theorem in $\mathbb{R}^2$

If $f(x,y)$ is continuous on a closed and bounded set $S \subset \mathbb{R}^2$, then there exist points $(a,b),(c,d) \in S$ such that

Theorem 10.1.2

$$f(a,b) \leq f(x,y) \leq f(c,d) \quad \text{for all } (x,y) \in S$$

Remark: A function may have an absolute maximum and/or an absolute minimum on a set $S \subseteq \mathbb{R}^2$ even if the conditions are not satisfied. We just cannot guarantee the existence with the theorem.

Algorithm for Extreme Values

Let $S \subset \mathbb{R}^2$ be closed and bounded. To find the maximum and/or minimum value of a function $f(x,y)$ that is continuous on S,

Algorithm

1. Find all critical points of f that are contained in S. Evaluate f at each such point.
2. Find the maximum and minimum points of f on the boundary $B(S)$.
3. The maximum value of f on S is the largest value of the function found in steps 1 and 2. The minimum value of f on S is the smallest value of the function found in steps 1 and 2.

Remark: The absolute maximum and/or minimum value may occur at more than one point in S. Also, it is not necessary to determine whether the critical points are local maximum or minimum points.

Optimization with Constraints

As we would do in step 2 of the algorithm for finding extreme values of a function f on the boundary of the set S, we would have certain constraints so that we evaluate f at $B(S)$, usually in the form $g(x,y) = k$. Here we may use parameters to find extema.
We now define an algorithm which will allow us to find the extrema of a differentiable function f on a smooth curve $g(x,y) = k$ without having to parameterize the curve.

Method of Lagrange Multipliers

We want to find the maximum (minimum) value of a differentiable function $f(x, y)$ subject to the constraint $g(x, y) = k$ where $g \in C^1$, or, more geometrically, find the maximum (minimum) value of $f(x, y)$ on the level set $g(x, y) = k$ (on all $(x, y) \mid g(x, y) = k$)

Algorithm Lagrange Multiplier Algorithm

Assume that $f(x, y)$ is a differentiable function and $g \in C^1$. To find the maximum value and minimum value of f subject to the constraint $g(x, y) = k$, evaluate $f(x, y)$ at all point (a, b) which satisfy one of the following conditions.

1. $\nabla f(a, b) = \lambda \nabla g(a, b)$ and $g(a, b) = k$
2. $\nabla g(a, b) = (0, 0)$ and $g(a, b) = k$
3. (a, b) is an end point of the curve $g(x, y) = k$

The maximum/minimum value of $f(x, y)$ is the largest/smallest value of f obtained at the points found in 1-3

Note, to find the points (a, b) in 1 we have to solve three equations in three unknowns.

$$f_x(x, y) = \lambda g_x(x, y)$$
$$f_y(x, y) = \lambda g_y(x, y)$$
$$g(x, y) = k$$

for x and y. Be sure to solve systematically to find all possible points.

Remarks: The variable λ, called the **Lagrange multiplier**, is not required for our purposes and so should be eliminated. However, it can be extremely useful in some real world applications.

Case 2 must be included since we assumed $\nabla g(a, b) \neq (0, 0)$ in the derivation.

If the curve $g(x, y) = k$ is unbounded, then one must consider $\lim_{||(x,y)|| \to \infty} f(x, y)$ for (x, y) satisfying $g(x, y) = k$

Functions of Three Variables

Algorithm

To find the maximum/minimum value of a differentiable function $f(x, y, z)$ subject to $g(x, y, z) = k$ such that $g \in C^1$, we evaluate $f(x, y, z)$ at all points (a, b, c) which satisfy one of the following:

1. $\nabla f(a, b, c) = \lambda \nabla g(a, b, c)$ and $g(a, b, c) = k$
2. $\nabla g(a, b, c) = (0, 0, 0)$ and $g(a, b, c) = k$
3. (a, b, c) is a boundary point of the surface $g(x, y, z) = k$

The maximum/minimum value of $f(x, y, z)$ is the largest/smallest value of f obtained at the points found in 1-3

Remark: If condition 1 holds, it follows that the level surface $f(x, y, z) = f(a, b, c)$ and the constraint surface $g(x, y, z) = k$ are tangent at the point (a, b, c), since their normals coincide (see Theorem 7.3.1)

Generalization

The method of the Lagrange multipliers can be generalized to functions of n variables $f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n$ and with r constraints of the form

$$g_1(\mathbf{x}) = 0, \quad g_2(\mathbf{x}) = 0, \quad \ldots, \quad g_r(\mathbf{x}) = 0$$

In order to find the possible maximum and minimum points of f subject to the constraints above, one has to find all points $\mathbf{a}$ such that

$$\nabla f(\mathbf{a}) = \lambda_1 \nabla g_1(\mathbf{a}) + \cdots + \lambda_r \nabla g_r(\mathbf{a}), \quad \text{and} \quad g_i(\mathbf{a}) = 0, \quad 1 \leq i \leq r$$

The scalars $\lambda_1, \ldots, \lambda_r$ are the Lagrange multiplies. When $r = 1$, and $n = 2$ or 3, this reduces to the previous algorithms

Remark: Sometimes you may need to get creative to solve using the lagrange multipliers method. Since you have a system of equations, you may need to inspect them as that, by checking consistency etc.

Unit 11: Coordinate Systems

A **coordinate system** is a system for representing the location of a point in a space by an ordered n-tuple. We call the elements of the n-tuple the **coordinates** of the point.

Polar Coordinates

The frame of reference for this coordinate system consists of a point O called the **pole** (or origin). From O we draw a ray called the **polar axis**, (usually horizontal to the right, to match the x-axis in the Cartesian coordinate system).

Letting P be any point in the plane, we represent the position of a point by the ordered pair (r, θ) where $r \geq 0$ is the length of the line OP and θ is the angle between the polar axis and OP. We call r and θ the polar coordinates of P

Remarks:

1. θ is considered positive if it is measured in the counterclockwise direction from the polar axis, and negative if clockwise.
2. We represent the point O by $(0, \theta)$ for any θ
3. The restriction of $r \geq 0$ is not common, we just use it so that it coincides with the interpretation of r as distance.
4. Since we use distance r from the pole in our representations, polar coordinates are useful in solving problems with symmetry about the pole.
5. Polar coordinate representations are not unique, since $(r, \theta) = (r, \theta + 2\pi k), k \in \mathbb{Z}$

Relationship to Cartesian Coordinates

For any (x, y) represented by the Cartesian coordinate system, we can find that

$$x = r\cos\theta$$

$$y = r\sin\theta$$

$$r = \sqrt{x^2 + y^2}$$

$$\tan\theta = \frac{y}{x}$$

Remark: The equation $\tan\theta = \frac{y}{x}$ does not uniquely determine θ, since over $0 \leq \theta \leq 2\pi$ the value of $\tan\theta$ occurs twice, so we have to use the one that lies in the correct quadrant.

Area in Polar Coordinates

If θ_1 and θ_2 with $\theta_2 > \theta_1$ are two angles in circle of radius r, then the area between them is

$$\frac{\theta_2 - \theta_1}{2\pi} \cdot \pi r^2 = \frac{1}{2} r^2 (\theta_2 - \theta_1)$$

We now derive the area. Dividing the region bounded into $\theta = a, \theta = b$ and $r = f(\theta)$ into subregions $\theta_0, \ldots, \theta_n$ of equal difference $\Delta\theta$, then for each subregion $\theta_i, 0 \leq i \leq n$ we pick some point θ_i^* with $\theta_i \leq \theta_i^* \leq \theta_{i+1}$. We then from the sector between θ_i and θ_{i+1} with radius $f(\theta_i^*)$. The area of this sector is

$$\frac{1}{2}[f(\theta_i^*)]^2 \Delta\theta$$

Letting the number of subdivisions go to infinity, hence letting each of the $\Delta\theta_i$ going to zero, and summing up the areas of all the sectors. We get

$$A = \lim_{||\Delta\theta_i|| \to 0} \sum_{i=0}^{n-1} \frac{1}{2}[f)\theta_i^*]^2 \Delta\theta = \int_a^b \frac{1}{2}[f(\theta)]^2 d\theta$$

Remark: Always have a rough sketch of the polar curves you may be solving for as there may be intersections or points you can miss.

Cylindrical Coordinates

An extension of polar coordinates to 3-dimensional space is by adding another axis, called the **axis of symmetry**, through the pole perpendicular to the polar plane. Any point P in the space is represented as (r, θ, z) where r and θ are as in the polar coordinates and z is the height above or below the polar plane. The restrictions are the same as Polar coordinates; $r \geq 0, 0 \leq \theta < 2\pi$ (or $-\pi < \theta \leq \pi$)

The relationship of Cylinderical Coordinates to Cartesian coordinates can be found to be

$$x = r\cos\theta$$

$$y = r\sin\theta$$

$$z = z$$

$$r = \sqrt{x^2 + y^2}$$

$$\tan\theta = \frac{y}{x}$$

Remark: As with functions $z = f(x, y)$, the graphs of functions $z = f(r, \theta)$, or more generally, $f(r, \theta, z) = 0$ are surfaces in $\mathbb{R}^3$

Spherical Coordinates

We learned that polar coordinates were useful for problems in 2-dimensional space which were symmetric about the origin. We now extend this to 3-dimensional space with **spherical coordinates**

Let P be any point in 3-dimensional space. We represent P by the coordinates (ρ, ϕ, θ), where $\rho \geq 0$ is the length of the line OP, θ is the same angle as in cylindrical coordinates, and ϕ is the angle between the positive z-axis and the line OP.

Since ρ and θ can span the 2-dimensional plane like in the cylindrical coordinates, ϕ only dictates the tilt of the graph with respect to the z- axis, so we restrict $0 \leq \phi \leq \pi$
We observe with examples that θ sort of controls which quadrant the point is in (its rotation around the z-axis) and ϕ only controls whether the point will be above or below the xy-plane.
So the restrictions for Spherical Coordinates are $\rho \geq 0, 0 \leq \theta < 2\pi$ (or $-\pi < \theta \leq \pi$), and $0 \leq \phi \leq \pi$

The relationship of Spherical Coordinates with Cartesian coordinates can be found to be

$$x = \rho \sin\phi \cos\theta$$

$$y = \rho \sin\phi \sin\theta$$

$$z = \rho \cos\phi$$

$$\rho = \sqrt{x^2 + y^2 + z^2}$$

$$\tan\theta = \frac{y}{x}$$

$$\cos\phi = \frac{z}{\sqrt{x^2 + y^2 + z^2}}$$

Unit 12: Mappings of $\mathbb{R}^2$ into $\mathbb{R}^2$

A function whose domain is a subset of $\mathbb{R}^n$ and whose codomain is $\mathbb{R}^m$ is called a **vector-valued function**

Definition
Vector-Valued Function

We have already seen the simplest type of vector-valued functions. Consider the parametric equations $x = f(t), y = g(t)$ for a curve in $\mathbb{R}^2$: These two scalar equations can be written as a vector equation:

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

The function F maps t to $F(t)$, so the domain of F is a subset of $\mathbb{R}$ and its codomain is $\mathbb{R}^2$. $F : \mathbb{R} \to \mathbb{R}^2$. Consequently, F is a vector-valued function.

A vector-valued function whose domain is a subset of $\mathbb{R}^n$ and whose codomain is a subset of $\mathbb{R}^n$ is called a **mapping** (or transformation)

Definition
Mapping

The Geometry of Mappings

A pair of equations

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

associates with each point $(x, y) \in \mathbb{R}^2$ a single point $(u, v) \in \mathbb{R}^2$, and this defines a vector-valued function

$$(u, v) = F(x, y) = (f(x, y), g(x, y))$$

The scalar functions f and g are called the **component functions** of the mapping.

Mappings of $\mathbb{R}^n$ into $\mathbb{R}^n$ have many applications, such as defining curvilinear coordinate systems (e.g. polar coordinates), and performing a change of variables in multiple integrals (chapters 14 and 15). They are used in applied mathematics, in statistics, and in computer graphics for simplifying problems in two or more variables.

In general, if a mapping $F : \mathbb{R}^2 \to \mathbb{R}^2$ acts on a curve C in its domain, it will determine its range, denoted by $F(C)$ and called the **image of** C **under** F

More generally, if a mapping $F : \mathbb{R}^2 \to \mathbb{R}^2$ acts on any subset S in its domain it will determine a set $F(S)$ in its range, called the **image of** S **under** F

Remark: The mapping of polar coordinates to Cartesian coordinates is non-linear. The image of a straight line is not necessarily a straight line. In general, a mapping will deform a given curve or set.

The Linear Approximation of a Mapping

Consider a mapping F defined by $u = f(x, y), v = g(x, y)$. We assume that F has continuous partial derivatives. By this we mean that the component functions f and g have continuous partial dertivatives.
The image of a point (a, b) in the xy-plane is the point (c, d) in the uv-plane, where $c = f(a, b), d = g(a, b)$
As usual, we want to approximate the image $(c+\Delta u, d+\Delta v)$ of a nearby point $(a+\Delta x, b+\Delta y)$.

We do this by using the linear approximation for $f(x, y)$ and $g(x, y)$ separately. We get

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

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

for Δx and Δy sufficiently small. This can be written as

$$\begin{bmatrix}\Delta u \\ \Delta v\end{bmatrix} \approx \begin{bmatrix}\frac{\partial f}{\partial x}(a, b) & \frac{\partial f}{\partial y}(a, b) \\ \frac{\partial g}{\partial x}(a, b) & \frac{\partial g}{\partial y}(a, b)\end{bmatrix} \begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix}$$

Definition
Derivative Matrix

The **derivative matrix** of a mapping defined by

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

is denoted DF and defined by

$$DF = \begin{bmatrix}\frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y}\end{bmatrix}$$

If we introduce the column vectors

$$\Delta\mathbf{u} = \begin{bmatrix}\Delta u \\ \Delta v\end{bmatrix}, \Delta\mathbf{x} = \begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix}$$

then the **increment form of the linear approximation of mappings** becomes

$$\Delta\mathbf{u} \approx DF(a, b)\Delta\mathbf{x}$$

for $\Delta\mathbf{x}$ sufficiently small. Thus, the **linear approximation of mappings** is

$$F(x, y) \approx F(a, b) + DF(a, b)\Delta\mathbf{x}$$

The geometrical interpretation of the linear approximation for mappings is that the derivative matrix $DF(a, b)$ acts as a linear mapping on the displacement vector $\Delta\mathbf{x}$ to give an approximation of the image $\Delta\mathbf{u}$ of the displacement under F

Generalization

The mapping $F : \mathbb{R}^n \to \mathbb{R}^m$ is defined by a set of m component functions:

$$\begin{aligned} u_1 &= f_1(x_1, \ldots, x_n) \\ &\vdots \\ u_m &= f_m(x_1, \ldots, x_n) \end{aligned}$$

Or, in vector notation

$$\mathbf{u} = F(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})), \quad \mathbf{x} \in \mathbb{R}^n$$

If we assume that F has continuous partial derivatives, then the derivative matrix of F is the $m \times n$ matrix defined by

$$DF(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial f_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

As expected, the linear approximation of F at $\mathbf{a}$ is

$$F(\mathbf{x}) \approx F(\mathbf{a}) + DF(\mathbf{a})\Delta\mathbf{x}$$

where

$$\Delta\mathbf{u} = \begin{bmatrix} \Delta u_1 \\ \vdots \\ \Delta u_m \end{bmatrix} \in \mathbb{R}^m, \qquad \Delta\mathbf{x} = \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix} \in \mathbb{R}^n$$

Composite Mappings and the Chain Rule

We now study the composition of two mappings.
Consider the successive mappings $F : \mathbb{R}^2 \to \mathbb{R}^2$ and $G : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$F : \begin{cases} p = p(u, v) \\ q = q(u, v) \end{cases} \qquad G : \begin{cases} u = u(x, y) \\ v = v(x, y) \end{cases}$$

We define the composite mapping of the two mappings

$$F \circ G : \begin{cases} p = p(u(x, y), v(x, y)) \\ q = q(u(x, y), v(x, y)) \end{cases}$$

maps the xy-plane directly into the pq-plane

We ask the question: how is the derivative matrix $D(F \circ G)$ of the composite mapping related to the derivative matrices DF and DG of the individual mappings?

$D(F \circ G)(x, y)$ is the matrix product of $DF(u, v)$ and $DG(x, y)$, where $(u, v) = G(x, y)$.

(Chain Rule in Matrix Form) Theorem 12.3.1

Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ and $G : \mathbb{R}^2 \to \mathbb{R}^2$. If G has continuous partial derivatives at (x, y) and F has continuous partial derivatives at $(u, v) = G(x, y)$, then the composite mappings $F \circ G$ has continuous derivatives at (x, y) and

$$D(F \circ G)(x, y) = DF(u, v)DG(x, y)$$

Unit 13: Jacobians and Inverse Mappings

The Inverse Mapping Theorem

We will now work to find a condition which will guarantee that a mapping $(u, v) = F(x, y)$ has an inverse. We define inverse mappings in the expected way

Let F be a mapping from a set D_{xy} onto a set D_{uv}. If there exists a mapping F^{-1}, called the **inverse of** F which maps D_{uv} onto D_{xy} such that

$$(x, y) = F^{-1}(u, v) \quad \text{if and only if} \quad (u, v) = F(x, y)$$

then F is said to be **invertible** on D_{xy}

Definition Invertible Mapping, Inverse Mapping

As usual, we have

$$(F^{-1} \circ F)(x, y) = (x, y) \qquad \text{for all } (x, y) \in D_{xy}$$
$$(F \circ F^{-1})(u, v) = (u, v) \qquad \text{for all } (u, v) \in D_{uv}$$

We know that the invertibility of a function is related to it being one-to-one

A mapping $F : \mathbb{R}^2 \to \mathbb{R}^2$ is said to be **one-to-one** on a set D_{xy} if and only if $F(a, b) = F(c, d)$ implies $(a, b) = (c, d)$, for all $(a, b), (c, d) \in D_{xy}$

Definition One-to-One

Let F be a mapping from a set D_{xy} onto a set D_{uv}. If F is one-to-one on D_{xy}, then F is invertible on D_{xy}

Theorem 13.1.1

Recall from Single Variable Calculus that if $f'(x) > 0$ or $f'(x) < 0$ for all $x \in [a, b]$, then f is one-to-one and has an inverse on $[a, b]$. So let's investigate the relationship of the derivative matrix DF for a mapping F being invertible. We start with the following theorem.

Consider a mapping F which maps D_{xy} onto D_{uv}. If F has continuous partial derivatives at $\mathbf{x} \in D_{xy}$ and there exists an inverse mapping F^{-1} of F which has continuous partial derivatives at $\mathbf{u} = F(\mathbf{x}) \in D_{uv}$, then

Theorem 13.1.2

$$DF^{-1}(\mathbf{u})DF(\mathbf{x}) = I$$

Seeing examples, we find that we can solve and obtain unique solutions for inverse mappings for simple problems. Hence, we develop a test to determine if a mapping F has an inverse mapping. The determinant of the derivative matrix plays an important role in the study of mappings and in their application to multiple integrals.

The **Jacobian** of a mapping

Definition
Jacobian

$$(u, v) = F(x, y) = (u(x, y), v(x, y))$$

is denoted $\dfrac{\partial(u, v)}{\partial(x, y)}$, and is defined by

$$\frac{\partial(u, v)}{\partial(x, y)} = \det[DF(x, y)] = \det \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}$$

We can interpret Theorem 13.1.2 as asserting that if a mapping F is invertible, then its derivative matrix $DF(x, y)$ is invertible, and its inverse matrix is the derivative matrix $DF^{-1}(u, v)$ of the inverse map. From Linear Algebra we know that a square matrix has an inverse if and only if its determinant is non-zero. So we have that if the mapping F has an inverse mapping F^{-1} (and both mappings have continuous partial derivatives), then the Jacobian of F is non-zero. This is stated as a Corollary

Consider a mapping defined by

Corollary
13.1.3

$$(u, v) = F(x, y) = (f(x, y), g(x, y))$$

which maps a subset D_{xy} onto a subset D_{uv}. Suppose that f and g have continuous partial derivatives on D_{xy}. If F has an inverse mapping F^{-1}, with continuous partial derivatives on D_{uv}, then the Jacobian of F is non-zero:

$$\frac{\partial(u, v)}{\partial(x, y)} \neq 0, \quad \text{on } D_{xy}$$

Note: Converse is NOT always true, but soon we will see how to restrict the domain so that it becomes true

Remark: Recall from Linear Algebra that $\det(AB) = \det A \det B$ for all $n \times n$ matrices A, B. So, we can deduce from Theorem 2 a simple relationship between the Jacobian of a mapping and the Jacobian of the inverse mapping.

(Inverse Property of the Jacobian)

Corollary
13.1.4

If the hypotheses of Theorem 2 hold, then

$$\frac{\partial(x, y)}{\partial(u, v)} = \frac{1}{\frac{\partial(u,v)}{\partial(x,y)}}$$

(Inverse Mapping Theorem)

Theorem
13.1.5

If a mapping $(u, v) = F(x, y)$ has continuous partial derivatives in some neighborhood of (a, b) and $\frac{\partial(u,v)}{\partial(x,y)} \neq 0$ at (a, b), then there is a neighborhood of (a, b) in which F has an inverse mapping $(x, y) = F^{-1}(u, v)$ which has continuous partial derivatives

Geometrical Interpretation of the Jacobian

This interpretation is based on the following result from Linear Algebra. The area of a parallelogram defined by the vectors $\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ and $\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ is given by

$$\text{Area } = \left| \det \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \right|$$

We calculate the area of the image in the uv-plane, of a small rectangle in the xy-plane under a mapping F

We approximate the image of the rectangle refined by the vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$, where $PR = \Delta y$ and $PQ = \Delta x$, as a parallelogram defined by the vectors $\overrightarrow{P'Q'}$ and $\overrightarrow{P'R'}$, and use the linear approximation to approximate $\overrightarrow{P'Q'}$ and $\overrightarrow{P'R'}$

Since $\overrightarrow{PQ} = \begin{bmatrix} \Delta x \\ 0 \end{bmatrix}$ and $\overrightarrow{PQ} = \begin{bmatrix} 0 \\ \Delta y \end{bmatrix}$ we get

$$\overrightarrow{P'Q'} \approx \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \begin{bmatrix} \Delta x \\ 0 \end{bmatrix} = \begin{bmatrix} u_x \Delta x \\ v_x \Delta x \end{bmatrix}$$

$$\overrightarrow{P'R'} \approx \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \begin{bmatrix} 0 \\ \Delta y \end{bmatrix} = \begin{bmatrix} u_y \Delta y \\ v_y \Delta y \end{bmatrix}$$

for Δx and Δy sufficiently small. Note that the partial derivatives are evaluated at P. We have

$$\Delta A_{xy} = \Delta x \Delta y$$

And so

$$\Delta A_{uv} \approx \left| \det \begin{bmatrix} u_x \Delta x & u_y \Delta y \\ v_x \Delta x & v_y \Delta y \end{bmatrix} \right| = \left| \det \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \right| \Delta x \Delta y$$

since Δx and Δy are positive. Thus, by definition of the Jacobian,

$$\Delta A_{uv} \approx \left| \frac{\partial(u, v)}{\partial(x, y)} \right| \Delta A_{xy}$$

where the Jacobian is evaluated at P.

In words, the Jacobian of a mapping F describes the extent to which F increases or decreases areas. We can think of it as a magnification factor for very small areas that are mapped by F. Keep in mind this relation is an approximation and works increasingly accurately for smaller Δx and Δy

Remark: For a linear mapping $(u, v) = F(x, y) = (ax+by, cx+dy)$ where a, b, c, d are constants, the derivative matrix is

$$DF(x, y) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and thus the linear approximation is exact by Taylor's Theorem since all second partial derivaitves are zero. Therefore, for a linear mapping the approximation of small change in area becomes an exact relation. (As the Jacobian, the 'magnification factor', would be the same for all points since it doesn't depend on x or y, so it's not an extrapolation).

Generalization

At the end of Section 12.2, we generalized the concept of a mapping $F : \mathbb{R}^2 \to \mathbb{R}^2$ to a mapping $F : \mathbb{R}^n \to \mathbb{R}^m$, and defined the $m \times n$ derivative matrix $DF(\mathbf{x})$. If $m = n$, then we can define the Jacobian of the mapping, as follows.

Definition Jacobian

For a mapping defined by

$$\mathbf{u} = F(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_n(\mathbf{x}))$$

where $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{x} = (x_1, \ldots, x_n)$, the **Jacobian** of F is

$$\frac{\partial(u_1, \ldots, u_n)}{\partial(x_1, \ldots, x_n)} = \det[DF(\mathbf{x})] = \det \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

We also note that the inverse property of the Jacobian also generalizes:

$$\frac{\partial(u_1, \ldots, u_n)}{\partial(x_1, \ldots, x_n)} = \frac{1}{\frac{\partial(x_1, \ldots, x_n)}{\partial(u_1, \ldots, u_n)}}$$

where $\frac{\partial(x_1, \ldots, x_n)}{\partial(u_1, \ldots, u_n)}$ is the Jacobian of the inverse mapping of F.

Geometrical Interpretation of the Jacobian in 3-D

The interpretation is based on the following result from Linear Algebra. The volume of a parallelepiped which is defined by three vectors $\mathbf{a} = (a_1, a_2, a_3), \mathbf{b} = (b_1, b_2, b_3), \mathbf{c} = (c_1, c_2, c_3)$ is given by

$$\text{Volume} = \left| \det \begin{bmatrix} \mathbf{a} & \mathbf{b} & \mathbf{c} \end{bmatrix} \right|$$

Consider a mapping defined by

$$\mathbf{u} = F(\mathbf{x}) = (f(\mathbf{x}), g(\mathbf{x}), h(\mathbf{x}))$$

The image of a small rectangular block of volume $\Delta V_{xyz} = \Delta x \Delta y \Delta z$ in xyz-space under this mapping can be approximated by a small parallelepiped in uvw-space. As in the 2-D case we can use the linear approximation and the formula above to approximate the volume ΔV_{uvw} of the image. The result is

$$\Delta V_{uvw} = \left| \frac{\partial(\mathbf{u})}{\partial(\mathbf{x})} \right| \Delta V_{xyz}$$

where $\frac{\partial(\mathbf{u})}{\partial(\mathbf{x})}$ is the Jacobian of the mapping F evaluated at P.

Constructing Mappings: Read section from course notes

Unit 14: Double Integrals

Defintion of Double Integrals

In single variable calculus, to find the area under a curve $y = f(x)$ over a closed interval $[a, b]$ we used a single integral defined as a limit of Riemann sums:

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\Delta x_i$$

where Δx_i was the length of the i-th subinterval.
To find the volumes of complicated regions, or masses of 2D objects etc, we use double integrals.

Let D be a closed and bounded set in $\mathbb{R}^2$ whose boundary is a piecewise smooth closed curve. Let $f(x, y)$ be a function which is bounded on D, that is, there exists a number $M : |f(x, y)| \leq M$ for all $(x, y) \in D$

Subdivide D by means of straight lines parallel to the axes, forming a partition P of D. Label the n rectangles that lie completely in D, in some specific order, and denote their areas by $\Delta A_i, i = 1, \ldots, n$. Choose a point (x_i, y_i) in the i-th rectangle and form the Riemann sum

$$\sum_{i=1}^{n} f(x_i, y_i)\Delta A_i$$

Let $D \subset \mathbb{R}^2$ be closed and bounded. Let P be a partition of D as described above, and let $|\Delta P|$ denote the length of the longest side of all rectangles in the partition P. A function $f(x, y)$ which is bounded on D is **integrable** on D if all Riemann sums approach the same value as $|\Delta P| \to 0$

Definition
Integrable

If $f(x, y)$ is integrable on a closed bounded set D, then we define the **double integral** of f on D as

$$\iint_D f(x, y)dA = \lim_{\Delta P\to 0} \sum_{i=1}^{n} f(x_i, y_i)\Delta A_i$$

Definition
Double
Integral

If f is continuous on D, it can be proved that f is integrable on D, i.e. the double integral of f does exist. Functions which are discontinuous on D may be integrable on D. For example, if f is continuous on D except at points which lie on a curve C (f is piecewise continuous), then f is integrable. The proofs are beyond the scope of this course.

Interpretation of the Double Integral

You can think of the symbol $\iint_D f(x,y)dA$ as the limit of a sum. There can be many interpretations depending on the meaning you assign to the integrand $f(x,y)$. The symbol dA should remind you of the area of a rectangle in a partition of D

Double Integral as Area:
The simplest interpretation is when you define f to be the constant function with the value unity:

$$f(x,y) = 1, \text{ for all } (x,y) \in D$$

Then the Riemann sums just sum the areas of all the rectangles in D, and the double integral serves to define the area $A(D)$ of the set D:

$$A(D) = \iint_D 1\, dA$$

Double Integral as Volume:
If $f(x,y) \geq 0$ for all $(x,y) \in D$, then the double integral

$$\iint_D f(x,y)\, dA$$

can be interpreted as the volume $V(S)$ of the region defined by

$$S = \{(x,y,z) \,|\, 0 \leq z \leq f(x,y),\ (x,y) \in D\}$$

which represents the solid below the surface $z = f(x,y)$ and above the set D in the xy-plane.

The justification: the partition P of D decomposes the solid S into vertical "columns". The height of the columns above the i-th rectangle is approximately $f(x_i, y_i)$, so its volume is approximately

$$f(x_i, y_i)\Delta A_i$$

Thus, the Riemann sum approximates the volume $V(S)$:

$$V(S) \approx \sum_{i=1}^{n} f(x_i, y_i)\Delta A_i$$

As $|\Delta P| \to 0$ the partition becomes increasingly fine, so the error in the approximation tends to zero. Thus, the volume $V(S)$ is

$$V(S) = \iint_D f(x,y)\, dA$$

Double Integral as Mass:

If we have a thin metal plate whose density varies with position, we can think of it as varying "area density," i.e. function $f(x, y)$ gives the mass per unit area at position (x, y). So the approximate mass of a small rectangle of area ΔA_i at (x_i, y_i) is given by

$$\Delta M_i \approx f(x_i, y_i)\Delta A_i$$

Hence, the Riemann sum approximates the total mass of the plate D

$$M = \iint_D f(x, y)\, dA$$

Double Integral as Probability

Let $f(x, y)$ be the probability density of a continuous 2-D random variable (X, Y). The probability that $(X, Y) \in D$, a given subset of $\mathbb{R}^2$, is

$$P((X, Y) \in D) = \iint_D f(x, y)\, dA$$

Average Value of a Function

The double integral is also used to define the value of a function $f(x, y)$ over a set $D \subset \mathbb{R}^2$

As we recall that the average value of a single variable function, $f(x)$, over an interval $[a, b]$, f_{av}, is defined by

$$f_{av} = \frac{1}{b-a}\int_a^b f(x)\, dx$$

We have for a function of two variables, $f(x, y)$, we can define the average value of a f over a closed and bounded subset D of $\mathbb{R}^2$ by

$$f_{av} = \frac{1}{A(D)}\iint_D f(x, y)\, dA$$

Properties of the Double Integral

(Linearity) Theorem 14.1.1

If $D \subset \mathbb{R}^2$ is a closed and bounded set and f and g are two integrable functions on D, then for any constant c:

$$\iint_D (f + g)\, dA = \iint_D f\, dA + \iint_D g\, dA$$

$$\iint_D cf\, dA = c\iint_D f\, dA$$

(Basic Inequality) Theorem 14.1.2

If $D \subset \mathbb{R}^2$ is a closed and bounded set and f and g are two integrable functions on D such that $f(x,y) \leq g(x,y)$ for all $(x,y) \in D$, then

$$\iint_D f\,dA \leq \iint_D g\,dA$$

(Absolute Value Inequality) Theorem 14.1.3

If $D \subset \mathbb{R}^2$ is a closed and bounded set and f is an integrable function on D, then

$$\left| \iint_D f\,dA \right| \leq \iint_D |f|\,dA$$

(Decomposition) Theorem 14.1.4

Assume $D \subset \mathbb{R}^2$ is a closed and bounded set and f is an integrable function on D. If D is decomposed into two closed and bounded subsets D_1 and D_2 by a piecewise smooth curve C, then

$$\iint_D f\,dA = \iint_{D_1} f\,dA + \iint_{D_2} f\,dA$$

Remarks:

1. The Basic Inequality can be used to obtain an extimate for a double integral that cannot be evaluated exactly,
2. The decomposition property is essential for dealing with complicated regions of integration and with discontinuous integrands.

Iterated Integrals

We can approximate double integrals using programs written to evaluate suitable Riemann sums, where the accuracy depends on the size of the partition. However, we can also exactly calculate double integrals of sufficiently simple functions, by writing the double integrals as a succession of two single integrals, called an **iterated integral**. We will derive a method for doing this by using the interpretation of the double integral as volume.

(Best to look at the diagrams in the course notes)
Let D be a region in the xy-plane and let f be a function such that $f(x,y) \geq 0$ for all $(x,y) \in D$. If V denotes the volume of the solid above D and below the surface $z = f(x,y)$, then we have

$$V = \iint_D f(x,y)\,dA$$

Assume that the region D lies between vertical lines $x = x_l$ and $x = x_u$ with $x_l \leq x_u$ and has top curve $y = y_u(x)$ and bottom curve $y = y_l(x)$. That is, the bounded and closed region D is described by the inequalities

$$y_l(x) \leq y \leq y_u(x), \text{ and } x_l \leq x \leq x_u$$

From Integral Calculus, we know that we can find the volume of a region by integrating over all possible cross-sectional areas. That is,

$$V = \int_{x_l}^{x_u} A(x)\, dx$$

where $A(x)$ is the cross-sectional area of the solid for any *fixed* value of x. But, we know that the cross-sectional area $A(x)$ for any fixed value of x is the area under the cross-section $z = f(x, y)$ in the zy-plane, and thus is given by a single integral

$$A(x) = \int_{y_l(x)}^{y_u(x)} f(x, y)\, dy$$

Hence, the volume of the region is

$$V = \int_{x_l}^{x_u} \left(\int_{y_l(x)}^{y_u(x)} f(x, y)\, dy \right) dx$$

This, we have

$$\iint_D f(x, y)\, dA = \int_{x_l}^{x_u} \int_{y_l(x)}^{y_u(x)} f(x, y)\, dy\, dx$$

as desired.

Let $D \subset \mathbb{R}^2$ be defined by

$$y_l(x) \leq y \leq y_u(x), \text{ and } x_l \leq x \leq x_u$$

where $y_l(x)$ and $y_u(x)$ are continuous for $x_l \leq x \leq x_u$. If $f(x, y)$ is continuous on D then

$$\iint_D f(x, y)\, dA = \int_{x_l}^{x_u} \int_{y_l(x)}^{y_u(x)} f(x, y)\, dy\, dx$$

Theorem 14.2.1

The proof is beyond the scope of the course

Remark:
We evaluate the inner integral first, even if we omit the parentheses. Moreover, as in our interpretation of volume, when we integrate with respect to y while holding x constant, we are using **partial integration**

We can have the set D described by inequalities of the form

$$x_l(y) \leq x \leq x_u(y), \text{ and } y_l \leq y \leq y_u$$

and so we get the reverse of Theorem 1

$$\iint_D f(x,y)\,dA = \int_{y_l}^{y_u}\int_{x_l(y)}^{x_u(y)} f(x,y)\,dx\,dy$$

Remark: one must be mindful of how to describe the iterated integral. Two factors that must be taken into account the shape of the region D, and the form of the integrand $f(x,y)$

Change of Variable Theorem

A mapping $F:\mathbb{R}^2\to\mathbb{R}^2$ can be used to simplify a double intergral

$$\iint_{D_{xy}} G(x,y)\,dA$$

either by changing the integrand $G(x,y)$, or by deforming the set D_{xy} in the xy-plane into a simpler shape D_{uv} in the uv-plane. The process is called a **change of variables** in the double integral. In this type of calculation it is convenient to replace the symbol "dA" in the double integral by "$dx\,dy$" if one is working in the xy-plane, and by "$du\,dv$" if one is working in the uv-plane

To derive the change of variable formula for double integrals, we need the formula which describes how areas are related under a mapping F given by

$$(x,y) = F(u,v) = (f(u,v), g(u,v))$$

The geometric interpretation of the Jacobian gives us

$$\Delta A_{xy} \approx \left|\frac{\partial(x,y)}{\partial(u,v)}\right| \Delta A_{uv}$$

for $\Delta u, \Delta v$ sufficiently small where the Jacobian $\frac{\partial(x,y)}{\partial(u,v)}$ is evaluated at a point in the region. Notice the interchanged roles of (x,y) and (u,v) in the equations as compared to unit 13.

(Change of Variable Theorem) Theorem 14.3.1

Let each of D_{uv} and D_{xy} be a closed bounded set whose boundary is a piecewise smooth closed curve. Let

$$(x,y) = F(u,v) = (f(u,v), g(u,v))$$

be a one-to-one mapping of D_{uv} onto D_{xy}, with $f, g \in C^1$, and $\frac{\partial(x,y)}{\partial(u,v)} \neq 0$ except possibly on a finite collection of piecewise-smooth curves in D_{uv}. If $G(x,y)$ is continuous on D_{xy}, then

$$\iint_{D_{xy}} G(x,y)\,dx\,dy = \iint_{D_{uv}} G(f(u,v), g(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv$$

Double Integrals in Polar Coordinates

If the boundary of a region is a circle centered on the origin or a circle that passes through

the origin, it will often help to transform from polar to Cartesian coordinates. Recall that the mapping from polar to Cartesian coordinates is

$$(x, y) = F(r, \theta) = (r\cos\theta, r\sin\theta)$$

which has Jacobian

$$\frac{\partial(x, y)}{\partial(r, \theta)} = r$$

Hence, we must restrict $r > 0$ so that the mapping is one-to-one and the Jacobian is non-zero so that the we can apply the Change of Variable Theorem. Note that we can make this restriction even if the origin is in the region as the integral over a single point is 0.

Remark: Since polar coordinates have a simple geometric interpretation, we can directly obtain the r and θ limits of integration from the diagram in the xy-plane without having to draw the region $D_{r\theta}$.

Unit 15: Triple Integrals

Definition of Triple Integrals

Let D be a closed bounded set in $\mathbb{R}^3$ whose boundary consists of a finite number of surface elements which are smooth except possibly at isolated points. Let $f(x, y, z)$ be a function which is bounded on D. Subdivide D by means of three families of planes which are parallel to the $xy-$, $yz-$, and $xz-$ planes respectively, forming a partition P of D

Label the N rectangular blocks that lie completely in D in some specific order, and denote their volumes by $\Delta V_i, i = 1, \ldots, n$. Choose an arbitrary point (x_i, y_i, z_i) in the i-th block, $i = 1, \ldots, n$, and form the Riemann sum

$$\sum_{i=1}^{n} f(x_i, y_i, z_i)\Delta V_i$$

Let ΔP denote the maximum of the dimensions of all rectangular blocks in the partition P.

A function $f(x, y, z)$ which is bounded on a closed bounded set $D \subset \mathbb{R}^3$ is said to be **integrable** on D if and only if all Riemann sums approach the same value as $\Delta P \to 0$

Definition Integrable

If $f(x, y, z)$ is integrable on a closed bounded set D, then we define the **triple integral** of f over D, as

$$\iiint_D f(x, y, z)\, dV = \lim_{\Delta P \to 0} \sum_{i=1}^{n} f(x_i, y_i, z_i)\Delta V_i$$

Definition Triple Integral

Again, if f is continuous on D, it can be proved that f is integrable on D. Functions with discontinuities on D may be integrable on D. E.g. if f is continuous on D expect at points which lie on a surface or curve in D, then f is integrable on D. Proofs are beyond the scope of this course.

Interpretation of the Triple Integral

Again, the symbol $\iiint_D f(x, y, z)\, dV$ should be thought of as the "limit of a sum." The interpretations can depend on the meaning you assign the integrand $f(x, y, z)$. The "dV" should remind you of the volume of a rectangular block in the partition of D.

Triple Integral as Volume

Letting $f(x, y, z) = 1$, for all $(x, y, z) \in D$, the RIemann sum sums the volumes of all the rectangular blocks in D, so the triple integral defines the volume $V(D)$ of D

$$V(D) = \iiint_D 1\, dV$$

Triple Integral as Mass

For a 3D object with varying density, we can let $f(x, y, z)$ represent the density (mass per unit

volume) at position (x, y, z). The mass of a small cuboid at (x_i, y_i, z_i) will be approximately

$$\Delta M_I \approx f(x_i, y_i, z_i)\Delta V_i$$

So the Riemann sum corresponding to a partition P of D

$$\sum_1^n f(x_i, y_i, z_i)\Delta V_i$$

will approximate the total mass M of the object. So the triple integral gives the total mass of the object

$$M = \iiint_D f(x, y, z)\, dV$$

Average Value of a Function

Let $D \subset \mathbb{R}^3$ be closed and bounded with volume $V(D) \neq 0$, and let $f(x, y, z)$ be a bounded and integrable function on D. The average value of f over D is defined by

Definition
Average Value

$$f_{\text{avg}} = \frac{1}{V(D)} \iiint_D f(x, y, z)\, dV$$

(Linearity)

Theorem 15.1.1

If $D \subset \mathbb{R}^3$ is a closed and bounded set and f and g are two integrable functions on D, then for any constant c:

$$\iiint_D (f + g)\, dV = \iiint_D f\, dV + \iiint_D g\, dV$$

$$\iiint_D cf\, dV = c \iiint_D f\, dV$$

(Basic Inequality)

Theorem 15.1.2

If $D \subset \mathbb{R}^3$ is a closed and bounded set and f and g are two integrable functions on D such that $f(x, y, z) \leq g(x, y, z)$ for all $(x, y, z) \in D$, then

$$\iiint_D f\, dV \leq \iiint_D g\, dV$$

(Absolute Value Inequality)

Theorem 15.1.3

If $D \subset \mathbb{R}^3$ is a closed and bounded set and f is an integrable function on D, then

$$\left| \iiint_D f\, dV \right| \leq \iiint_D |f|\, dV$$

(Decomposition) Theorem 15.1.4

Assume $D \subset \mathbb{R}^3$ is a closed and bounded set and f is an integrable function on D. If D is decomposed into two closed and bounded subsets D_1 and D_2 by a piecewise smooth curve C, then

$$\iiint_D f\,dV = \iiint_{D_1} f\,dV + \iiint_{D_2} f\,dV$$

Iterated Integrals

Consider a set $D \subset \mathbb{R}^3$ which is described by inequalities of the form

$$z_l(x,y) \le z \le z_u(x,y)$$

and

$$(x,y) \in D_{xy}$$

Here D_{xy} is a closed bounded subset of $\mathbb{R}^2$ whose boundary is a piecewise smooth closed curve, and z_l, z_u are continuous functions on D_{xy}. Think of the set D as being the 3-D region with botton surface $z = z_l(x,y)$ and top surface $z = z_u(x,y)$, where the extent is defined by the 2-D set D_{xy}

In order to write the triple integral as an iterated integral, take an arbitrary point $(x,y) \in D_{xy}$. Then integrate $f(x,y,z)$ with respect to z from $z_l(x,y)$ to $z_u(x,y)$, and integrate the result over D_{xy}, as a double integral.

This procedure sums over all rectangular blocks in a partition of D, and hence gives the triple integral of $f(x,y,z)$ over D

Let D be the subset of $\mathbb{R}^3$ defined by Theorem 15.2.1

$$z_l(x,y) \le z \le z_u(x,y) \quad \text{and} \quad (x,y) \in D_{xy}$$

where z_l and z_u are continous functions on D_{xy}, and D_{xy} is a closed bounded subset in $\mathbb{R}^2$, whose boundary is a piecewise smooth closed curve. If $f(x,y,z)$ is continous on D, then

$$\iiint_D f(x,y,z)\,dV = \iint_{D_{xy}} \int_{z_l(x,y)}^{z_u(x,y)} f(x,y,z)\,dz\,dA$$

Remark: for the inner integral, we carry out partial integration with respect to z.

Similar to the double integral case, if we have defined D in terms of inequalities of the form $x(y,z)$ then the inner integral is with respect to x and simlar for $y(x,z)$

The Change of Variable Theorem

A mapping $F : \mathbb{R}^3 \to \mathbb{R}^3$ can be used to simplify a triple integral

$$\iiint_{D_{xyz}} G(x, y, z)\, dV$$

Either by changing the integrand $G(x, y, z)$ or by deforming the set D_{xyz} in xyz-space into a simplar shape D_{uvw} in uvw-space, thereby simplifying the limits of integration. It is convenient to replace the "dV" by "$dx\, dy\, dz$" for the xyz-space and "$du\, dv\, dw$" for the uvw-space

(Change of Variable Theorem) Theorem 15.3.1

Let

$$(x, y, z) = \big(f(u, v, w), g(u, v, w), h(u, v, w)\big)$$

be a one-to-one mapping of D_{uvw} onto D_{xyz}, which f, g, h having continous partial derivatives, and

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} \neq 0 \text{ on } D_{uvw}$$

If $G(x, y, z)$ is continuous on D_{xyz}, then

$$\iiint_{D_{xyz}} G(x, y, z)\, dV = \iiint_{D_{uvw}} G\big(f(u, v, w), g(u, v, w), h(u, v, w)\big) \left|\frac{\partial(x, y, z)}{\partial(u, v, w)}\right| dV$$

Just like in the case of Polar Coordinates, mappings to Cylindrical and Spherical coordinates will be useful as we can exploit symmetry about the z-axis or origin in $\mathbb{R}^3$

Triple Integrals in Cylindrical Coordinates

The mapping from Cartesian to Cylindrical coordinates is given by

$$x = r\cos\theta, \quad y = r\sin\theta, \quad z = z$$

with $r \geq 0, 0 \leq \theta \leq 2\pi$, and the Jacobian is $\frac{\partial(x,y,z)}{\partial(r,\theta,z)} = r$. Since we need the Jacobian to be non-zero, we restrict $r > 0$. Hence the formula in the Change of Variable Theorem reads

$$\iiint_{D_{xyz}} G(x, y, z)\, dV = \iiint_{D_{r\theta z}} G(r\cos\theta, r\sin\theta, z) r\, dr\, d\theta\, dz$$

Triple Integrals in Spherical Coordinates

The mapping from Spherical to Cartesian coordinates is given by

$$x = \rho\sin\phi\cos\theta, \quad y = \rho\sin\phi\sin\theta, \quad z = \rho\cos\phi$$

with $\rho \geq 0, 0 \leq \phi \leq \pi, 0 \leq \theta \leq 2\pi$. The Jacobian is

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} = \rho^2 \sin \phi$$

We must restrict $\rho > 0, 0 < \phi < \pi$ so that the Jacobian is non-zero and the mapping is one-to-one. Notice that we are not just removing one point, but the entire z-axis. Though, this will not affect the result as the triple integral over the z-axis is 0. So the Change of Variable theorem reads

$$\iiint_{D_{xyz}} G(x, y, z)\, dV = \iiint_{D_{\rho\theta\phi}} G(\rho \sin \phi \cos \theta, \rho \sin \phi \sin \theta, \rho \cos \phi) \rho^2 \sin \phi \, d\rho \, d\theta \, d\phi$$