Lecture Notes 2021

Analysis 2

Lecturer: Upendra Kulkarni
Scribe: Soham Chatterjee

Introduction

This is the lecture notes scribed by me. If you find any mistakes in the notes please email me at `sohamc@cmi.ac.in`.

The whole course is taken by Prof. Upendra Kulkarni, online. If you want the lecture videos then you can find them in this link. Sir mainly followed Prof. Pramath Sastry's Notes (`https://www.cmi.ac.in/~pramath/teaching.html#ANA2`). You can find all the assignments problems in the following drive link. Through out the course the book we followed is Principles of Mathematical Analysis by Walter Rudin.

Contents

Chapter 1

Normed Linear Space

Definition 1.0.1: Limit of Sequence in $\mathbb{R}$

Let $\{s_n\}$ be a sequence in $\mathbb{R}$. We say

$$\lim_{n\to\infty} s_n = s$$

where $s \in \mathbb{R}$ if $\forall$ real numbers $\epsilon > 0$ $\exists$ natural number N such that for $n > N$

$$s - \epsilon < s_n < s + \epsilon \text{ i.e. } |s - s_n| < \epsilon$$

Want to generalize this to a sequence in $\mathbb{R}^n$ i.e. $s_n \in \mathbb{R}^n\ \forall\ n \in \mathbb{N}$. Now the $s - s_n$ makes no sense. So it is useful to have a notion of whether vectors are big or small. We have magnitude of a vector. So lets revisit this

Definition 1.0.2: Limit of Sequence in $\mathbb{R}^n$

Let $\{s_n\}$ be a sequence in $\mathbb{R}^n$. We say

$$\lim_{n\to\infty} s_n = s$$

where $s \in \mathbb{R}^n$ if $\forall$ real numbers $\epsilon > 0$ $\exists$ natural number N such that for $n > N$

$$\|s - s_n\| < \epsilon$$

The same definition works if we interpret $\|v\| =$ length of the vector v.

From school, for v=$v_1, v_2, \cdots, v_n$ we had

$$\|v\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

But it will be useful to have a more general notion of length (of which the above will be an example)

1.1 Defination

Definition 1.1.1: Normed Linear Space and Norm $\|\cdot\|$

Let V be a vector space over $\mathbb{R}$ (or $\mathbb{C}$). A norm on V is function $\|\cdot\|\ V \to \mathbb{R}_{\geq 0}$ satisfying

1. $\|x\| = 0 \iff x = 0\ \forall\ x \in V$
2. $\|\lambda x\| = |\lambda|\|x\|\ \forall\ \lambda \in \mathbb{R}(\text{or } \mathbb{C}),\ x \in V$
3. $\|x + y\| \leq \|x\| + \|y\|\ \forall\ x, y \in V$ (Triangle Inequality/Subadditivity)

And V is called a normed linear space.

• Same definition works with V a vector space over $\mathbb{C}$ (again $\|\cdot\| \to \mathbb{R}_{\geq 0}$) where 2 becomes $\|\lambda x\| = |\lambda|\|x\|$ $\forall\ \lambda \in \mathbb{C},\ x \in V$, where for $\lambda = a + ib$, $|\lambda| = \sqrt{a^2 + b^2}$

Example 1.1.1 ($p-$Norm)
$V=\mathbb{R}^m$, $p\in\mathbb{R}_{\geq 0}$. Define for $x=(x_1,x_2,\cdots,x_m)\in\mathbb{R}^m$

$$\|x\|_p=\Big(|x_1|^p+|x_2|^p+\cdots+|x_m|^p\Big)^{\frac{1}{p}}$$

(In school $p=2$)

Special Case $p=1$: $\|x\|_1=|x_1|+|x_2|+\cdots+|x_m|$ is clearly a norm by usual triangle inequality.
Special Case $p\to\infty$ ($\mathbb{R}^m$ with $\|\cdot\|_\infty$): $\|x\|_\infty=\max\{|x_1|,|x_2|,\cdots,|x_m|\}$
For $m=1$ these $p-$norms are nothing but $|x|$. Now exercise

Question 1

Prove that triangle inequality is true if $p\geq 1$ for $p-$norms. (What goes wrong for $p<1$?)

Solution: **For Property 3 for norm-2**

When field is $\mathbb{R}$:

We have to show

$$\sum_i(x_i+y_i)^2\leq\left(\sqrt{\sum_i x_i^2}+\sqrt{\sum_i y_i^2}\right)^2$$

$$\implies\sum_i(x_i^2+2x_iy_i+y_i^2)\leq\sum_i x_i^2+2\sqrt{\left[\sum_i x_i^2\right]\left[\sum_i y_i^2\right]}+\sum_i y_i^2$$

$$\implies\left[\sum_i x_iy_i\right]^2\leq\left[\sum_i x_i^2\right]\left[\sum_i y_i^2\right]$$

So in other words prove $\langle x,y\rangle^2\leq\langle x,x\rangle\langle y,y\rangle$ where

$$\langle x,y\rangle=\sum_i x_iy_i$$

Note:-

- $\|x\|^2=\langle x,x\rangle$
- $\langle x,y\rangle=\langle y,x\rangle$
- $\langle\cdot,\cdot\rangle$ is $\mathbb{R}-$linear in each slot i.e.

$$\langle rx+x',y\rangle=r\langle x,y\rangle+\langle x',y\rangle\text{ and similarly for second slot}$$

Here in $\langle x,y\rangle$ x is in first slot and y is in second slot.

Now the statement is just the Cauchy-Schwartz Inequality. For proof

$$\langle x,y\rangle^2\leq\langle x,x\rangle\langle y,y\rangle$$

expand everything of $\langle x-\lambda y,x-\lambda y\rangle$ which is going to give a quadratic equation in variable λ

$$\begin{aligned}\langle x-\lambda y,x-\lambda y\rangle&=\langle x,x-\lambda y\rangle-\lambda\langle y,x-\lambda y\rangle\\&=\langle x,x\rangle-\lambda\langle x,y\rangle-\lambda\langle y,x\rangle+\lambda^2\langle y,y\rangle\\&=\langle x,x\rangle-2\lambda\langle x,y\rangle+\lambda^2\langle y,y\rangle\end{aligned}$$

Now unless $x=\lambda y$ we have $\langle x-\lambda y,x-\lambda y\rangle>0$ Hence the quadratic equation has no root therefore the discriminant is greater than zero.

When field is $\mathbb{C}$:

Modify the definition by

$$\langle x, y\rangle = \sum_i \overline{x_i} y_i$$

Then we still have $\langle x, x\rangle \geq 0$

1.2 Open and Closed Ball

Definition 1.2.1: Open and Closed Ball in Normed Linear Space

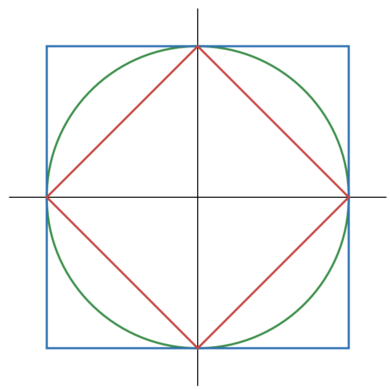
An open Ball of radius r with center x in Normed Linear Space V is the set

$$\{y \in V \mid \|x - y\| < r\} = B_r(x)$$

and Closed ball is the set

$$\{y \in V \mid \|x - y\| \leq r\} = \overline{B_r(x)}$$

Now take $B_r(0)$ w.r.t $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$. Now imagine a sequence converging to origin. So if I



draw an ordinary circle around the origin then no matter how small the circle the points of the sequence are eventually land inside the circle. If instead of that circle can same be said for diamond w.r.t norm 2. Then i can take circle that is inside that diamond. Same is true for $\infty-$norm. Hence convergence with respect to all norm 1 and norm 2 and even ∞ results for convergence.

Now there is no reason why we can not consider a norm on an infinite dimensional vector space. It will work. Perhaps i can define only for some sequences where the morm converges.

Example 1.2.1

Suppose for set of all bounded infinite sequences a vector space because every number in a vector is less than some number so if you add two vectors then add the bound and if you scale then scale the bound. Now the ∞ norm works on that.
Now suppose you take all continuous real valued functions on closed interval $[0, 1]$, such a function is bounded and this is a vector space and we can define $\infty-$norm even for that because for all f in this space attains its maximum value so just take that maximum value. Its an extremely infinite dimensional space.

Note:-

$\mathbb{R}^\infty$ is the space of all sequences.

Question 2

Modify the above proof for field $\mathbb{C}$

Question 3

Show that the following are normed linear spaces.

(a) l^∞ = Set of all bounded infinite sequences $(x_1, x_2, \cdots)$ $x_i \in \mathbb{R}$ with norm $\|x\| = \sup |x_i|$

(b) $C[0,1]$ = Set of all continuous functions $[0,1] \to \mathbb{R}$ with norm $\|f\| = \sup_{x\in[0,1]} |f(x)|$

1.3 Limit of a Sequence

Definition 1.3.1: Limit of Sequence in Normed Linear Space

A sequence $\{s_n\}$ in a normed linear space V converge to s means $\forall$ real number $\epsilon > 0$ $\exists$ natural number N such that for $\forall$ $n > N$ $\|s - s_n\| < \epsilon$

1.4 Continuity

Definition 1.4.1: Continuity in Normed Linear Space

Let S be a subset of V and $f: S \to W$ where V, W are normed linear space. f is continuous at $v \in V$ means $\forall$ $\epsilon > 0$, $\exists$ $\delta > 0$, st whenever $\|x - v\| < \delta$ for $x \in S$ one has $\|f(x) - f(v)\| < \epsilon$

Distance in a normed linear space for $x, y \in V$ is

$$d(x,y) = \|x, y\|$$

Hence properties of this d are

1. $d(x,y) = 0 \iff x = y$
2. $d(\lambda x, \lambda y) = |\lambda| d(x,y)$ for any scalar λ
3. $d(u,v) + d(u,v) \geq d(u,w)$

Chapter 2

Metric Space

2.1 Definition

Definition 2.1.1: Metric Space X

A set X with a function $d \ X \times X \to \mathbb{R}_{\geq 0}$ such that

1. $d(x,y) = 0 \iff x = y$
2. $d(x,y) = d(y,x)$
3. $d(x,z) \leq d(x,y) + d(y,z)$

Notice that there is no homogeneity condition, and it does ot make sense as we don't have a field. In fact there is no notion of addition. But the condition 1 of norm has to be satisfied by this distance. Also we don't have a translational condition i.e. distance between x, y and distance between $x+v, y+v$ has to be same. Hence

Note:-

A metric space need not be a vector space. So it doesn't need a zero, or a notion of addition or scalar multiplication.

If I take a metric space and take any subset of it. And those three conditions of distance functions are still satisfied.

Note:-

Any subset of metric space is a metric space under the same distance function.

2.2 Open and Closed Ball and Set

Definition 2.2.1: Open Ball and Closed Ball in a Metric Space

An open ball of radius r with center $c \in X$ in a metric space X is

$$B_r(c) = \{x \in X \mid d(c,x) < r\}$$

and a closed ball is

$$\overline{B_r(c)} = \{x \in X \mid d(c,x) \leq r\}$$

Definition 2.2.2: Open Set and Closed Ball in a Metric Space

An open set in a metric space X is one of the form of union of some open balls and a closed set in a metric space X is one of the form of $X \setminus$some open sets

Note:-

We will do topology in Normed Linear Space (Mainly $\mathbb{R}^n$ and occasionally $\mathbb{C}^n$)using the language of Metric Space

Example 2.2.1 (Open Set and Close Set)

Open Set:
- ϕ
- $\bigcup\limits_{x \in X} B_r(x)$ (Any $r > 0$ will do)
- $B_r(x)$ is open

Closed Set:
- $X, \ \phi$
- $\overline{B_r(x)}$

$x-$axis $\cup$ $y-$axis

Question 4

Is the set $x-$axis$\setminus\{$Origin$\}$ a closed set

Solution: We have to take its complement and check whether that set is a open set i.e. if it is a union of open balls

Now this works well for points which are above or below the $x-$axis. But for origin no matter how small the ball we take it willl have points from $x-$axis. Hence the set is not a closed set.

Question 5

Any continuous path in $\mathbb{R}^2$ is closed where path $= f : [0, 1] \to \mathbb{R}^2$

Solution: This is true. To be proved later.
Analogous to: For continuous function $f : [0, 1] \to \mathbb{R}$, the image is a closed interval

Question 6

If i take $X = x-$axis $\cup$ $y-$axis then is it open

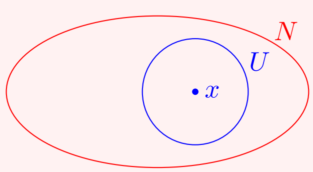
Solution: Yes because here the space is only the union of those two axis. So any ball would be like a cross or line but it just as the metric space given to us. [It is open for this metric space but not open in $\mathbb{R}^2$]

Note:-

If $S \subset X$, then S itself has a collection of open sets of S by containing S as a metric space.

Definition 2.2.3: Neighborhood

For a point x in metric space X, a neighborhood of x is a set N such that $x \in$an open set $U \subset N$

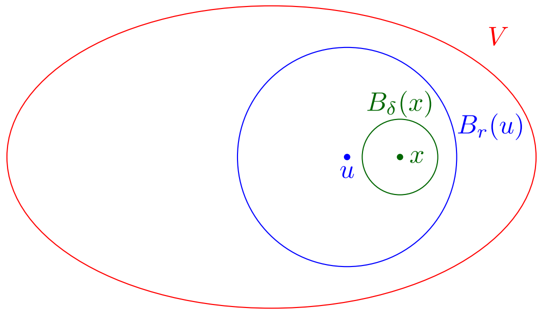


If N itself is open. then we say that N is an open neighborhood of x

Theorem 2.2.1

If $x \in$ open set V then $\exists\, \delta > 0$ such that $B_\delta(x) \subset V$

Proof: By openness of V, $x \in B_r(u) \subset V$



Given $x \in B_r(u) \subset V$, we want $\delta > 0$ such that $x \in B_\delta(x) \subset B_r(u) \subset V$. Let $d = d(u, x)$. Choose δ such that $d + \delta < r$ (e.g. $\delta < \frac{r-d}{2}$)

If $y \in B_\delta(x)$ we will be done by showing that $d(u, y) < r$ but

$$d(u, y) \leq d(u, x) + d(x, y) < d + \delta < r$$

$\square$

Note:-

V is open $\iff \bigcup\limits_{x \in V} B_r(x)$ (where r depends on x)

Theorem 2.2.2

Let X be a metric space.

1. Union of open sets is open
2. Intersection of two open sets is open

Analogues to these as we are just taking complement of the open sets

$1'$. Arbitrary intersection of closed sets is closed

$2'$. Finite union of closed sets is closed.

Proof:

1. Let $\{V_\alpha\}_{\alpha \in I}$ be a collection of open sets where I is an index set. We want ti show $\bigcup\limits_{\alpha \in I} V_\alpha$ is open in X. Since each V_α is open $V_\alpha = \bigcup\limits_{\beta \in J_\alpha} B_{r_\beta}(c_\beta)$ Then

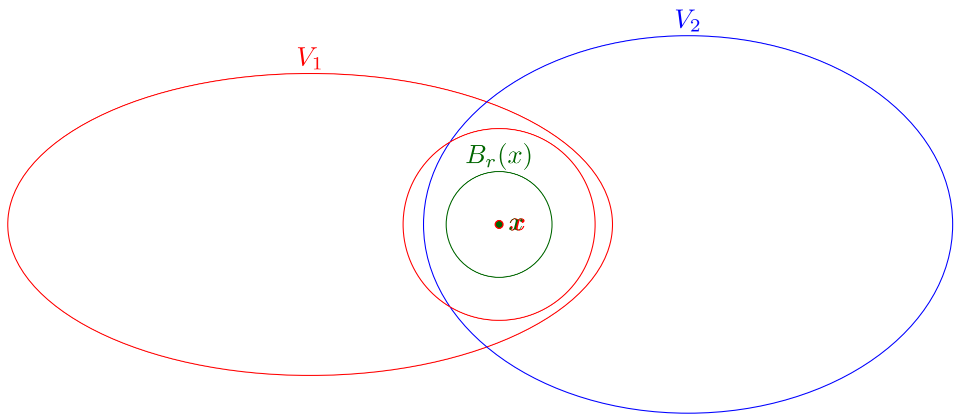
$$\begin{aligned} \bigcup_{\alpha \in I} V_\alpha &= \bigcup_{\alpha \in I} \bigcup_{\beta \in J_\alpha} B_{r_\beta}(c_\beta) \\ &= \bigcup_{\beta \in \sqcup J_\alpha} B_{r_\beta}(c_\beta) \end{aligned}$$

which is still a union of balls

$\square$

2. The statement implies intersection of finite number of open sets is open. We can prove this by induction.

We will do by showing that for each $x \in V_1 \cap V_2\ \exists\, r > 0$ s.t. $B_r(x) \subset V_1 \cap V_2$



As $x \in V_1$ $\exists$ r_1 such that $x \in B_{r_1}(x) \subset V_1$. Similarly $x \in V_2$ $\exists$ r_2 such that $x \in B_{r_2}(x) \subset V_2$. Take $r = \min\{r_1, r_2\}$. Thus we have $x \in B_r(x) \subset V_1 \cap V_2$

The second part for closed sets are left as exercise □

2.3 Topological Space

Definition 2.3.1: Topological Space

A topological space is a set X together with a collection of subsets of X (i.e. a subset of the power set of X) that is closed under taking arbitrary unions and finite intersections. This collection is called a topology on X

Note:-

Union means $\bigcup_{\alpha \in I} S_\alpha = \{x \in X \mid \exists\ \alpha \text{ s.t. } x \in S_\alpha\}$

Intersection means $\bigcap_{\alpha \in I} S_\alpha = \{x \in X \mid \forall \alpha,\ x \in S_\alpha\}$

Question 7

Suppose i have a topological space X under given some topology. Is the entire set open ? And that the empty set is open ?

Solution: If $I = \phi$, $\bigcup_{\alpha \in I} S_\alpha = \{x \in X \mid \exists\ \alpha \in I \text{ s.t. } x \in S_\alpha\}$ gives ϕ and $\bigcap_{\alpha \in I} S_\alpha = \{x \in X \mid \forall \alpha \in I,\ x \in S_\alpha\}$ gives X because $\forall\ \alpha \in I$ condition is vacuously true for each $x \in X$.

Note:-

Intersection of empty families are not defined in set theory. This brings a very important point. In a set theory you have to have a universe. (Set theory have to avoid paradoxes, Russel Paradox) At the beginning you construct a large enough universe and you taking subsets only from that universe. Notice all subsets we are considering here are subsets of X and here we defined how we union and intersection mean. Though it still this asks what our axioms of set theory. So you can change the part of the definition of topological space like this "...with a collection of subsets of X including the empty set and the whole space..."

(If you don't like this as it is)

Note:-

If S is a subset of metric space X, then S is itself a metric space and as such open/closed sets as subsets of metric space

Question 8

Is there any connection between being open in X and being open in S (Similar question for closed)

Solution: Let $x \in S$. Now, Ball of radius r in $S = S\cap$ Ball of radius r in X. Therefore

$$\begin{aligned}\text{Open Set in } S &= \bigcup \text{ Balls in } S \\ &= \bigcup (\text{Balls in } X \cap S) \\ &= \left(\bigcup \text{ Balls in } X\right) \cap S \\ &= \text{Open set } X \cap S\end{aligned}$$

Part 2 is left as exercise

Corollary 2.3.1

If $S \subset X$ is open in X then a subset T of S is open in $S \iff T$ is open in X

Corollary 2.3.2

If $S \subset X$ is closed in X then a subset T of S is closed in $S \iff T$ is closed in X

Definition 2.3.2: Subspace of a Topological Space X

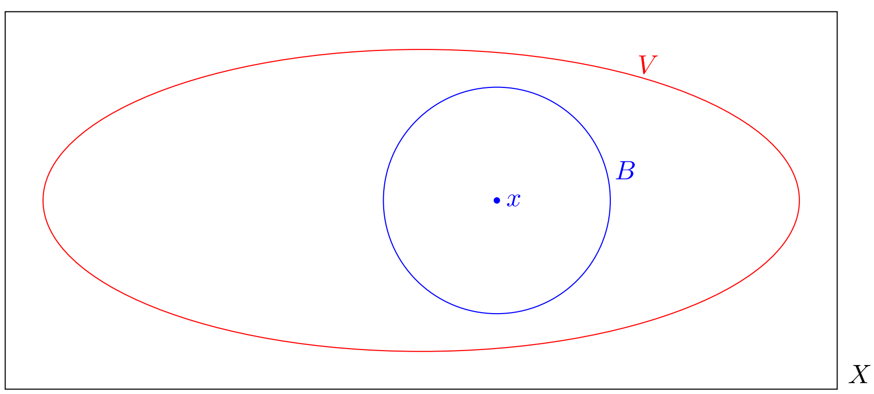
For any subset S of a topological space X, the collection $S \cap U$, U open in X is called a subspace.

Question 9

Prove that subspace of a metric space X defines a topology on X

Wrong Concept 2.1

If $x \in$ open V then there exists $r > 0$ such that $x \in B_r(x) \subset V$



Idea: Why not we take $r = \inf\{\text{distance from } x \text{ to boundary of ball } B\}$.

Now we first have to ensure $r > 0$. Suppose that's true.

Then we have to define boundary. What is boundary, We can give a reasonable definition (Boundary has already a definition but we don't know that for now). Let boundary of $B = \{x \in X \mid d(c, x) = \delta\}$ Now this definition is not proper for our purpose.Because if we take union of all balls in V then we will have lots of points as boundary but part of them should not be considered as boundary. Even if we take this definition.

Then the big question comes/ We are taking a infimum of a certain set of real numbers. The very first question arises is whether this set is nonempty. For example if we take B to be the metric space it self we have no boundary.

Questions which come thorough this.

- Is there a meaningful way to define boundary
- Can we modify the idea

Chapter 3

Continuity in Metric Space

3.1 Limit Point and Closure

Definition 3.1.1: Limit Point

$S \subset X$ is a metric space. We say that $x \in X$ is a limit point of S if $\exists$ a sequence $\{s_n\}$ with all $s_n \in S \setminus \{x\}$ such that $s_n \to x$ (each s_n is different from x)

Theorem 3.1.1

x is a limit point of $S \iff$ every neighborhood of x in X contains a point of S other than S.

Proof: **If Part:**

Let x be a limit point of S. Therefore take a sequence $\{s_n\}$ in $S \setminus \{x\}$ with $s_n \to x$.

To prove what we want it is enough to show that $B_r(x) \cap S$ contains a point other than x. As $s_n \to x$, $\exists$ N s.t. $\forall\ n > N$ $d(x, s_n) < r$ i.e. $s_n \in B_r(x)$. In particular $s_n]inB_r(x) \cap (S \setminus \{x\})$

Only If Part:

We need to produce a sequence $\{s_n\} \in S \setminus \{x\}$ with $\lim s_n = x$. Take $s_n \in B_{\frac{1}{n}}(x) \cap (S \setminus \{x\})$ See that $\lim\limits_{n\to\infty} s_n = x$. This is essentially because $\frac{1}{n} \to 0$.

Complete the rest of the proof.

□

Definition 3.1.2: Closure

Given a topological space X and $S \subset X$, the closure of the set S is $\overline{S}$ the smallest closed set containing S.

Theorem 3.1.2

$$\begin{aligned}\overline{S} &= \text{Smallest closed set of } X \text{ containing } S = \text{A}\\ &= S \cup (\text{limit points of } S) = \text{B}\\ &= \{x \in X \mid x = \lim_{n\to\infty} \text{ for some sequence } \{s_n\} \text{ in } S\} = \text{C}\\ &= \{x \in X \mid \text{Every neighborhood of } x \text{ intersects } S\} = \text{D}\end{aligned}$$

Proof: ***$A \subset D$***

$$A^c = \bigcup (\text{All open set } V \text{ s.t. } V \cap S = \phi)$$
$$D^c = \{x \in X \mid \exists \text{ open neighborhood of } x, B \text{ s.t. } B \cap S = \phi\}$$

Clearly for all $x \in D^c$, $x \in A^c$. Hence $D^c \subset A^c \implies A \subset D$

D $\subset$ ***B***

Take $x \in D$. Suppose $x \notin S$. Now any neighborhood of x intersects S in a point hence it has to be a different point from x since $x \notin S$. Therefore x is a limit point of S. $D \subset B$

B $\subset$ ***C***

If $x \in S$ then take a sequence

□

Question 10

What does it mean to be smallest closed set containing the set S here ?

Solution: $\bigcap$ All closed sets containing S is automatically closed and hence the smallest closed set containing S.

Proof: For proof of Theorem 3.1.2 notice A,B,C,D all contains S (obvious).

Note:-

We don't need to show B,C,D are closed. We can also take the sets element wise and show each set is a subset of the other. This may simplify our way of proof. (exercise)

Now see A and D completely deal with topology. A is about closed sets and D is about open sets. So A and D close to each other. Now by the 3.1.1 we have equivalence of C and D. So we can prove like this

$$A \iff D \iff B \,\&\, C$$

Left as exercise □

Note:-

For these kind of proofs instead of looking for the most efficient way try to find a path that allows you to go from anywhere to anywhere

3.2 Continuity

Definition 3.2.1: Continuity

$f : X \to Y$ function between metric spaces is continuous at $a \in X$ if $\forall\, \epsilon > 0\ \exists\, \delta > 0$ s.t.

$$\begin{array}{ccc} d(x,a) < \delta & \implies & d(f(a), f(x)) < \epsilon \\ \Updownarrow & & \Updownarrow \\ x \in B_\delta(x) & \implies & f(x) \in B_\epsilon(f(a)) \end{array}$$

That means f^{-1}(Any ball around $f(a)$) $\supset$ Ball around a.

So $f : X \to Y$ is continuous at all points $\iff$ f^{-1}(Any ball intersecting the range) $\supset$ A ball

Note:-

We can not say f^{-1}(Any ball) because because we need a ball that contains a point in the range

Theorem 3.2.1

f is continuous $\iff$ f^{-1}(Any open set in Y) is open in X

Proof: **If Part:-**

It is enough to show f^{-1}(Any ball) is open on X because f^{-1} preserves unions $f^{-1}\left(\bigcup_\alpha V_\alpha\right) = \bigcup_\alpha \left(f^{-1}(V_\alpha)\right)$

Let B is any open set (as its conceptually simpler to take open set here instead of a ball) in Y. Let $a \in f^{-1}(B)$. Hence we can say $f(a) \in B$. Since B is an open set we can say there is a ball $B_\epsilon(f(a)) \subset B$. Since f is continuous $\exists\ \delta$ such that $f(x) \in B_\epsilon(f(a))$ whenever $x \in B_\delta(a)$. Now $f^{-1}(B) \supset f^{-1}\Big(B_\epsilon(f(a))\Big) \supset B_\delta(a)$ Hence $f^{-1}(B)$ is open.

Only If Part:-

Lets prove continuity ar $a \in X$. We are given that $f^{-1}\Big(B_\epsilon\big(f(a)\big)\Big)$ is open and obviously contains a. Therefore $f^{-1}\Big(B_\epsilon\big(f(a)\big)\Big)$ contains a ball around a. Take $\delta =$ Radius of the ball. $\square$

Question 11

For a metric space X, show that $\overline{S} = \{x \in X \mid \lim\limits_{n\to\infty} s_n = x\}$ for some sequence $\{s_n\}$ in S.

Question 12

For a function $f : X \to Y$ between metric spaces, show that the followings are equivalent.

1. f is continuous
2. f^{-1}(Open Set) is open
3. f^{-1}(Closed Set) is closed
4. $f(\overline{S}) = \overline{f(S)}$
5. $x_n \to x \implies f(x_n) \to f(x)$

One or more of the above are wrong so check if they are true and if not then find the true statement.

Solution: 4 is wrong. How to correct and rest is left as exercise

Question 13

For $f : X \to Y$ any set map

(i) f^{-1} preserves unions, intersections, complements

(ii) Is there any condition on f under which f possesses the property above ?

Example 3.2.1 (Continuous Function)

1. Any constant function.
2. $X \xrightarrow{f} Y \xrightarrow{g} Z$ f, g continuous $\implies g \cdot f$ is continuous
3. Is $S \subset X$ then $S \xrightarrow{\text{Inclusion}} X$ is continuous
4. Projection $\underset{(x_1,x_2,\cdots,x_n)\mapsto x_i}{\mathbb{R}^n \to \mathbb{R}}$

   More generally for example $\underset{(x,y,z)\mapsto(x,x,y,y)}{\mathbb{R}^3 \to \mathbb{R}^4}$
5. Map from metric space to euclidean space.

$$\left.\begin{array}{l} X \to \mathbb{R}^n \\ x \mapsto (f_1(x), f_2(x), \cdots, f_n(x)) \end{array}\right\} \begin{array}{c} f \text{ is continuous} \\ \Longleftrightarrow \\ \text{each } f_i \text{is continuous} \end{array}$$

6. $\mathbb{R}\times\mathbb{R}\to\mathbb{R}:\ (x,y)\mapsto x\pm y, xy$ are continuous.

   We need to prove $x_n\to x$ and $y_n\to y$ in $\mathbb{R} \implies \begin{cases} x_n\pm y_n\to x\pm y\\ x_ny_n\to xy\end{cases}$

   $\mathbb{R}\setminus\{0\}\to\mathbb{R}:\ x\mapsto \frac{1}{x}$ is continuous

7. sum and product of two continuous real valued function on X are continuous

$$f,g:X\xrightarrow{f,g}\mathbb{R}\text{ continuous}\implies \underset{x}{X}\ \underset{\longmapsto}{\xrightarrow{f,g}}\ \underset{(f(x),g(x))}{\mathbb{R}\times\mathbb{R}}\xrightarrow{+}\mathbb{R}$$

$$f:X\to\mathbb{R}\implies \frac{1}{f}:\underbrace{X\setminus f^{-1}(0)}_{\text{open set in }X}\to\mathbb{R}\text{ is continuous}$$

   $\{0\}$ is closed in $\mathbb{R}$, so $f^{-1}(0)$ is closed in X by continuity of f

   Therefore any polynomial in continuous real valued functions on X is continuous.

8. **Special Case:**
   - $\mathbb{R}^n\xrightarrow{T}\mathbb{R}^m$ linear map is continuous where $(x_1,x_2,\cdots,x_n)\longmapsto(a_{11}x_1+\cdots+a_{1n}x_n, a_{21}x_1+\cdots+a_{2n}x_n,\cdots,a_{m1}x_1+\cdots+a_{mn}x_n)$

     $$\text{Matrix of } T=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}$$
   - $\underset{A}{M_{n\times n}(\mathbb{R})}\ \underset{\longmapsto}{\to}\ \underset{det(A)}{\mathbb{R}}$ is continuous

     $\frac{1}{\det}:GL_n(\mathbb{R})\to\mathbb{R}$

     Here $M_{n\times n}$ is a vector space of dimension n^2 in which $GL_n(\mathbb{R})=\{A\mid\det(A)\neq0\}$ is an open set.
   - $\underset{A\longmapsto A^{-1}}{GL_n(\mathbb{R})\to GL_n(\mathbb{R})}$ is continuous.

9. Any norm (f) on $\mathbb{R}^n$ is uniformly continuous w.r.t usual topology on $\mathbb{R}^n$ i.e. $f:\mathbb{R}^n\to\mathbb{R}$ is continuous w.r.t usual norms ($\|\cdot\|=p0$norm for $p=1,2,\infty$) on $\mathbb{R}^n(\|\cdot\|)$ and $\mathbb{R}(|\cdot|)$

Theorem 3.2.2

Any norm (f) on $\mathbb{R}^n$ is uniformly continuous w.r.t usual topology on $\mathbb{R}^n$ i.e. $\forall\ \epsilon>0\ \forall\ x,y\in\mathbb{R}^n\ \exists\ \delta>0$ s.t. $\|x-y\|<\delta\implies|f(x)-f(y)|<\epsilon$

Proof:

$$\left.\begin{aligned} f(x)&\le f(y)+f(x-y)\\ &\qquad\qquad\|\\ f(y)&\le f(x)+f(y-x)\end{aligned}\right\}|f(x)-f(y)|\le f(x-y)$$

Let $x=\sum x_ie_i$ and $y=\sum y_ie_i$ where $\{e_i\}$ is the standard basis of $\mathbb{R}^n$.

$$f(x-y)=f\left(\sum(x_i-y_i)e_i\right)\le\sum f\left((x_i-y_i)e_i\right)=|x_i-y_i|f(e_i)$$

Notice $\sum|x_i-y_i|=\|x-y\|_1$. Let $M=\max\{f(e_i)\}$ Then

$$|f(x)-f(y)|\le f(x-y)\le M\|x-y\|_1$$

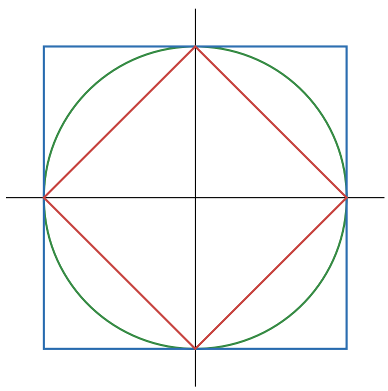
Thus $\|x-y\|<\frac{\epsilon}{M}\implies|f(x)-f(y)|<\epsilon$ □

Chapter 4

Equivalence of Norms

We back to Normed Linear Space for a little while.

In $\mathbb{R}^n$, $u = (u_1, u_2, \cdots, u_n)$ where each $u_i \in \mathbb{R}$. we have $p-$norm: $\|u\|_p = \left(\sum_i |u_i|^p\right)^{\frac{1}{p}}$ where $1 \leq p \leq \infty$.
Balls in $\mathbb{R}^2$ w.r.t. $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$.



Observe: A set V in $\mathbb{R}^2$ is

$$\begin{aligned} &\text{open w.r.t. } \|\cdot\|_1 &&\iff && V = \bigcup_{u\in V} \text{Box in } V \text{ centered box} \\ &\text{open w.r.t. } \|\cdot\|_2 &&\iff && V = \bigcup_{u\in V} \text{Diamond in } V \text{ centered box} \\ &\text{open w.r.t. } \|\cdot\|_\infty &&\iff && V = \bigcup_{u\in V} \text{Circle in } V \text{ centered box} \end{aligned}$$

Definition 4.1: Equivalence of Norms

Suppose $\|\cdot\|$, $\|\cdot\|'$ are two norms in vector space V, We say that the two norms are equivalent if there are constants $\alpha, \beta > 0$ s.t.

$$\alpha\|x\|' \leq \|x\| \leq \beta\|x\|'$$

Example 4.0.1 (Norm Equivalence)

1. $\boldsymbol{p = \infty}$ **and** $\boldsymbol{p = 1}$
   $\|x\|_\infty = \max\{|x_i| \mid 1 \leq i \leq n\} \leq \|x\|_1 = \sum_i |x_i|$
   $\|x\|_\infty \geq$ each $|x_i| \implies n\|x\|_\infty \geq \|x\|_1$
   Hence
   $$\|x\|_\infty| \leq \|x\|_1 \leq n\|x\|_\infty \text{ and } \frac{1}{n}\|x\|_1 \leq \|x\|_\infty \leq \|x\|_1$$

2. $\boldsymbol{p = \infty}$ **and** $\boldsymbol{p = 2}$

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$$

Theorem 4.1

All norms on a finite dimensional vector space are equivalent

Proof: Proved in Theorem 5.2.7 □

Theorem 4.2

Suppose $\|\cdot\|$ and $\|\cdot\|'$ are equivalent on a vector space V. Then

(i) $\{x_n\} \to x$ w.r.t. $\|\cdot\| \iff \{x_n\} \to x$ w.r.t $\|\cdot\|'$

(ii) $S \subset V$ is open w.r.t $\|\cdot\| \iff S$ is open w.r.t $\|\cdot\|'$

Proof: For both proofs if we just prove one direction the we are done actually since we can just replace the words to prove for opposite direction,

(i) **If Part:-**

Since $\|\cdot\|, \|\cdot\|'$ are equivalent we have $\exists\ \alpha, \beta$ such that $\alpha\|x\|' \leq \|x\| \leq \beta\|x\|'$. So if we show $\alpha\left\|x_n - x\right\| < \|x_n - x\| < \alpha\epsilon$ we are done.

Let $\{x_n\} \to x$ w.r.t $\|\cdot\|$ i.e. $\forall\ \epsilon > 0\ \exists\ N$ s.t. $\forall\ n > N\ \ \|x_n - x\| < \alpha\epsilon$. Hence we have $\alpha\left\|x_n - x\right\|' < \alpha\epsilon$. Hence $\forall\ \epsilon > 0\ \exists\ N$ such that $\forall\ n > N\ \ \left\|x_n - x\right\|' < \epsilon$

□

(ii) **Only If Part:-**

V is open w.r.t $\|\cdot\| \iff \bigcup\limits_{x \in V} B_r(x)$ and V' is open w.r.t $\|\cdot\|' \iff \bigcup\limits_{x \in V} B'_r(x)$

Now we have

$$B_r(x) = \{y \mid \|y - x\| < r\} \text{ and } B'_r(x) = \{y \mid \|y - x\|' < s\}$$

Hence by equivalence of the norms for any v

$$\alpha\|v\|' \leq \|v\| \leq \beta\|v\|'$$

Since $\|v\| < r$ we have

$$\|v\|' \leq \frac{r}{\beta} \implies B'_{\frac{r}{\beta}}(x) \subset B_r(x)$$

□

Corollary 4.1

$p = 1$ and $p = \infty$ on $\mathbb{R}^n$ (and $\mathbb{C}^n$) give the same topology as $p = 2$ norm

Corollary 4.2

Let x_m be a square in $\mathbb{R}^n$. $\overline{x_m} = (x_{m_1}, x_{m_2}, \cdots, x_{m_n})$. Then $\{\overline{x_m}\} \to x = (x_1, x_2, \cdots, x_n)$ w.r.t $\|\cdot\|_2 \iff \{x_{m_i}\} \to x_i$ in $\mathbb{R}$ for each i.

Note:-

We can check this w.r.t $\|\cdot\|_\infty$

$$\begin{aligned}
\overline{x_m} \to \overline{x} \text{ w.r.t } \|\cdot\|_\infty \quad &\iff \quad \forall\ \epsilon > 0\ \exists\ N \text{ s.t. } \forall\ m > N\ \max\{|x_{m_i} - x_i| \mid 1 \leq i \leq n\} \\
&\iff \quad \text{each } |x_{m_i} - x_i| < \epsilon\ \forall\ i \\
&\iff \quad \lim_{n\to\infty} x_{m_i} = x_i\ \forall\ i
\end{aligned}$$

Chapter 5

Compactness

5.1 Sequentially Compact

Definition 5.1.1: Sequentially Compact

Let (X, d) be a metric space. X is called sequentially compact if every sequence in X has a convergent subsequence. (Often applied to a subset S of X)

Note:-

For S to be sequentially compact the limit of subsequence must be in S

Definition 5.1.2: Boundedness

A subset S of (X, d) is bounded if $S \subset B_r(x)$ for some $x \in X$ and $r > 0$

Note:-

Boundedness depends on the metric but if two metrics are "equivalent" analogous to norms)

Theorem 5.1.1

A subset K of $\mathbb{R}^n$ is sequentially compact $\iff$ K is closed and bounded

Proof: Proof in steps

1. **A closed interval $[a, b]$ in $\mathbb{R}$ is sequentially compact**

   Proof: Given a sequence $x_1, x_2, \cdots$ in $\mathbb{R}$ in $[a, b]$ we can extract a monotonic subsequence as follows:

   We call x_i to be a peak if $x_i > x_j \; \forall \; j > i$. Now there are two cases. If number of peaks is infinite then the next peak comes after the previous one so smaller than the previous one. So its a strictly decreasing sequence. If number of peaks are finite then at some point we cant find a peak with this property that means no matter which term i peak there is at least one term after that which is greater than or equal to that term. y_1 =a term after the last peak. and y_{i+1} =a term after y_i such that $y_{i+1} \geq y_i$. Hence $y_1, y_2, \cdots$ is a weakly increasing sequence.

   When $\{x_n\}$ contained in $[a, b]$ by boundedness of the monotonic subsequence, it converges to its sup/inf and the limit is in $[a, b]$ $\square$

2. $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$ **is sequentially compact (w.r.t** $p-$**norm for** $p = 1, 2, \infty$. **Later for any norm)**

   Proof: Recall a sequence $\{x_m\} \to x$ in $\mathbb{R}^n \iff$ The sequence converges in each coordinate i.e. $x_{m_i} \to x_i$

   Take a sequence in the given box. Extract a subsequence whose entries in 1st slot converge (necessarily to x_i in $[a_1, b_1]$ by step 1 From this sequence, extract a further subsequence whose entries in second slot converge to $x_2 \in [a_2, b_2]$. Continue ☐

3. **Every closed subset of a sequentially compact set is sequentially compact**

   Proof: Exercise ☐

   This will show each closed and bounded subset of the Euclidean Space $\mathbb{R}^n$ is sequentially compact. (because such a set will be contained in a box)

4. **If K is sequentially compact then K is closed and bounded**

   Proof: If K is not closed then some limit point x of K will not be in K. Then there is a sequence $\{y_m\}$ in K converges to $x \notin K$ violating sequential compactness of K.

   If K is not bounded take $\{x_m\} \in K$ with $\|x_m\| \geq n$ then $\{x_m\}$ can not be convergent ☐

☐

Note:-

Step 4 works for any metric space. Then we need to have a ball instead of norm

Theorem 5.1.2

If K is a sequentially compact of a metric space X, then K is closed and bounded

Proof: Same argument as step 4 use x_m such that $d(x_m, x) \geq m$ ☐

Question 14

If K is closed and bounded in $(X, d) \implies K$ is sequentially compact

Solution: No. Any counter-example. Define a metric on real number which induces same topology as the normal topology in such a way that there is a closed and bounded set that is not compact.

Question 15

1. If V, W are normed linear spaces can we define a norm on $V \times W$?
2. If V, W are metric spaces can we define a metric on $V \times W$?
3. If V, W are topological spaces can we define a topology on $V \times W$?

5.2 Open Cover and Compactness

Definition 5.2.1: Open Cover

Let $\{V_\alpha\}_{\alpha \in I}$ be a family of subsets of metric space X we say that $\{V_\alpha\}_{\alpha \in I}$ is a cover of X if $\bigcup_\alpha V_\alpha = X$ and we say that $\{V_\alpha\}_{\alpha \in I}$ is an open cover if each V_α is open (in X)

Definition 5.2.2: Compact

X is called compact if each open cover of X has a finite subcover i.e. $\{V_{\alpha_1}, V_{\alpha_2}, \cdots, V_{\alpha_n}\} \subset \{V_\alpha\}_{\alpha \in I}$ with $V_{\alpha_1} \cup V_{\alpha_2} \cup \cdots \cup V_{\alpha_n} = X$

Note:-

1. This definition makes sense for any topological space X.

   If X is a metric space then it is a fact that X is compact $\iff$ X is sequentially compact. This is not true for general topological spaces. Both implications fail.

2. Reformulation of compactness for subset K of X in terms of open sets of X

$$
\begin{aligned}
K \text{ is compact} \quad &\iff \quad \text{Every cover of } K \text{ bt open sets of } K \text{ has a finite subcover.}\\
&\text{As open sets of } K \text{ are precisely (open sets of } X) \cap K. \text{ We have the following}\\
K \text{ is compact} \quad &\iff \quad \text{For any family } \{V_\alpha \cap K\}_{\alpha \in I} \text{ where } V_\alpha \text{ are open in } X \text{ whose union is } K,\\
&\qquad \text{there is a finite subcover.}\\
&\iff \quad \text{For any family } \{V_\alpha \cap K\}_{\alpha \in I} \text{ of open sets in } X \text{ such that } \bigcup_{\alpha \in I} V_\alpha \supset K,\\
&\qquad \text{there must be a finite subfamily } V_{\alpha_1}, V_{\alpha_2}, \cdots, V_{\alpha_n} \text{ with } \bigcup_{i=1}^{n} V_{\alpha_i} \supset K
\end{aligned}
$$

If i take this definition of compactness of a subset K of metric space X then K is compact as subset of $X \iff K$ is compact as a subset of it itself

Theorem 5.2.1 Haine Borel Theorem

$$K \subset \mathbb{R}^n \text{ is compact} \quad \iff \quad K \text{ is closed and boundeded}$$

(w.r.t $p = 1, 2$ or ∞ norm as they are equivalent.)

Proof: **Only If Part:-**

Proof in steps

1. Closed interval $[a, b]$ is compact in $\mathbb{R}$. ***Proof:*** Theorem 5.2.4
2. Closed box $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ is compact in $\mathbb{R}^n$. ***Proof:*** Theorem 5.2.6
3. A closed subset of a compact set is compact. ***Proof:*** Theorem 5.2.3

These steps would give the backward direction of Haine Borel Theorem i.e. suppose K is closed and bounded in $\mathbb{R}^n \implies K \in [-M.M]^n \implies$ compact by 2

If Part:-

Bounded: First we have to show that K is compact $\implies K$ is bounded an i.e. $K \subset B_r(x)$ in (X, d) for some $x \in X$, $r > 0$

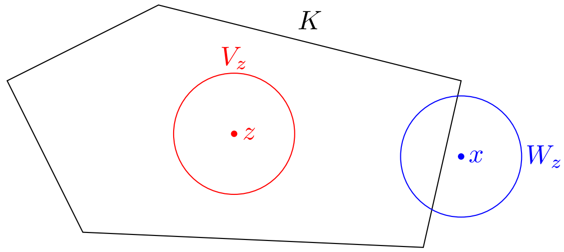
Consider open cover $\{B_n(x)\}_{n \in \mathbb{Z}^+}$ of X and hence of K. This must have a finite subcover $B_{n_1}(x), B_{n_2}(x), \cdots, B_{n_k}(x)$. Take $r = \max\{n_1, n_2, \cdots, n_k\}$ Hence

$$K \text{ is compact} \implies K \text{ is closed}$$

Closed: We will show that $X \setminus K$ is open. Pick $x \notin K$. Enough to construct an open neighborhood $U_x \ni x$ such that $U_x \cap x = \phi$

Take $z \in K$. Let $c = d(x, z)$ then

$$\underbrace{B_{\frac{c}{3}}(x)}_{W_z} \quad \cap \quad \underbrace{B_{\frac{c}{3}}(z)}_{V_z} = \phi \quad \text{by triangle inequality}$$



Now $\bigcup_{z\in K} V_z \supset K$. So $\{V_z\}$ is an open cover of K. By compactness we have $V_{z_1} \cup V_{z_2} \cup \cdots \cup V_{z_n} \supset K$. As $W_z \cap V_z = \phi\ \forall z \in K$. We have $\underbrace{(W_{z_1} \cup W_{z_2} \cup \cdots \cup W_{z_n})}_{\substack{\text{Finite intersection of} \\ \text{open neighborhoods of } x, \\ \text{so call this } U_x}} \cap K = \phi$ □

Key fact that made this work: For $x \neq z$ in X, we could find open neighborhoods of V and W (of x and z respectively) such that $V \cap W = \phi$. Topological spaces that satisfy this property are called Housdorff.

What we proved is the following

Theorem 5.2.2

For a Housdorff Topological space X any compact subset K is closed and bounded

Theorem 5.2.3 Haine Borel Theorem - If Part: Step 3

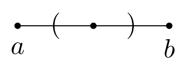
C is a closed subset of compact set $X \implies C$ is compact.

Proof: Take any open cover $\{V_\alpha\}_{\alpha\in I}$ of C by open sets in X i.e $\bigcup_\alpha V_\alpha \supset C$. Now $\{V_\alpha\}_{\alpha\in I} \cup \{X \setminus C\}$ is an open cover of x. We have a finite subcover by compactness of X. The same subcover (after dropping $X \setminus C$ if necessary) works for C. □

Wrong Concept 5.1: Closed interval $[a, b]$ is compact in $\mathbb{R}$

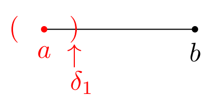
Suppose $\{V_\alpha\}_{\alpha\in I}$ is an open cover of $[a, b]$ by open sets in $\mathbb{R}$.

Hence every one of the points in the interval is covered by one of the V_α. Hence there is some interval contained in the V_α



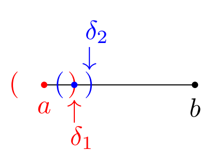
So i could just ignore the V_α and say for each point in the interval we can get an open interval that is part of a V_α. So how can i find a subcover. I could simply travel from one end to the other.

So i start with a so a must be contained in some open interval



Not only that i have covered up a small segment of the closed interval, upto a point, $a+\delta_1$. Say $[a, a+\delta_1) \subset V_1$.

Let $a + \delta_1$ is contained in some open interval which is contained in V_2 upto the point $a + \delta_2$



Now continue.

What is wrong with this ?

We could have smaller and smaller intervals. For example length of first interval can be $\frac{1}{3}$, length of second interval can be $\frac{1}{9}$, length of third interval can be $\frac{1}{27}$ and so on. So its a geometric progression and it will sum less than 1. So i just may not get there in finite number of steps.

Question 16

Suppose X is a topological space that is compact and 5.2 (Take x ro be a compact metric space if you like). Prove that given disjoint compact subsets K and L, there are disjoint open sets U and V with $K \subset U$ and $L \subset V$(First do it for $K =$ single point)

In the above exercise we could have replaced the word compact with another word which is closed because X is given to be compact so any closed set will be compact and in a Housdorff space compact subset is also closed.

Note:-

Cauchy Sequence in Metric space need not converge. For example $(0,1)$ and take the sequence $\frac{1}{n}$. It wants to converge to 0 but 0 is not there.

Theorem 5.2.4 Haine Borel Theorem - If Part: Step 1

$[0,1]$ is compact in $\mathbb{R}$

Proof: Let $\{V_\alpha\}_{\alpha\in I}$ be a family of open sets in $\mathbb{R}$ covering $[0,1]$.

Let $S = \{a \in [0,1] \mid [0,a]$ can be covered by a finite number of V_α's$\}$. Our goal is to prove $1 \in S$.

Let $0 \leq x < y \leq 1$. So $[0,x] \subset [0,y]$. This $y \in S \implies x \in S$ i.e $x \notin S \implies y \notin S$. Now S is nonempty because $0 \in S$ and S is bounded. Let $u =$ lub of S. Clearly $0 \leq u \leq 1$. Hence it is enough to show $u = 1$ and $u \in S$.

$0 \in$ some open set V_α. Hence $\exists\ \epsilon > 0$ $B_\epsilon(0) \subset V_\alpha$. Hence $\forall$ point $x \in [0,\epsilon)$ $x \in S$

For $a \in [0,u)$, $a \in S$ (otherwise a itself would be an upper bound for S). As $\{V_\alpha\}_{\alpha\in I}$ cover $[0,1]$, $u \in V_\beta$. So $\exists\ \epsilon > 0$ such that $(u-\epsilon, u+\epsilon) \subset V_\beta$ As $u - \epsilon \in S$ we have $V_{\alpha_1} \sup V_{\alpha_2} \sup \cdots V_{\alpha_k} \supset [0, u-\epsilon]$ Then $V_{\alpha_\beta} \cup V_{\alpha_1} \cup V_{\alpha_2} \cup \cdots V_{\alpha_k} \supset \left[0, u + \frac{\epsilon}{2}\right]$. So $u = 1$ because otherwise some $u + \delta \in S$ contradicting that u is an upper bound. ☐

Question 17

Can the strategy from the last time be made to work ti actually extract a finite subcover of a given cover.

Theorem 5.2.5

Suppose $X \xrightarrow{f} Y$ continuous and $K \subset X$ is compact. Then $f(K)$ is compact

Proof: Let $\{V_\alpha\}_{\alpha\in I}$ be an open cover of $f(k)$ by open sets V_α of Y. So

$$\bigcup_\alpha V_\alpha \supset f(K) \implies f^{-1}\left(\bigcup_\alpha V_\alpha\right) = \bigcup_\alpha f^{-1}\left(V_\alpha\right) \supset f^{-1}(f(K)) \supset K$$

Thus $\left\{f^{-1}(V_\alpha)\right\}_{\alpha\in I}$ is an open (because of continuity Theorem 3.2.1) cover of K.

Extract a finite subcover

$$\begin{aligned} & f^{-1}\left(V_{\alpha_1}\right) \cup f^{-1}\left(V_{\alpha_2}\right) \cup \cdots f^{-1}\left(V_{\alpha_m}\right) \supset K \\ \implies & f\left(f^{-1}\left(V_{\alpha_2}\right) \cup \cdots f^{-1}\left(V_{\alpha_m}\right)\right) \supset f(K) \\ \implies & \bigcup_{i=1}^m f\left(f^{-1}\left(V_{\alpha_i}\right)\right) \supset f(K) \end{aligned}$$

As $V_{\alpha_i} \supset f\left(f^{-1}\left(V_{\alpha_i}\right)\right)$ we have $\bigcup\limits_{i=1}^{m} V_{\alpha_i} \supset f(K)$ $\square$

Question 18

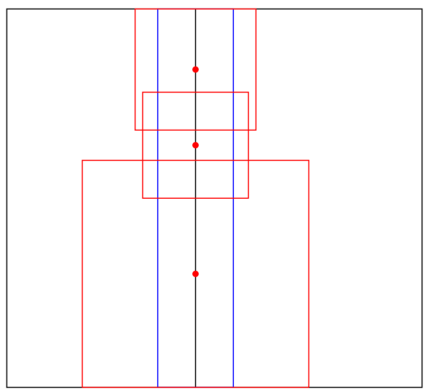
f(Sequentially compact K) is sequentially compact

Theorem 5.2.6 Haine Borel Theorem - If Part: Step 2

$K = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ is compact in $\mathbb{R}^n$

Proof: Induction on n. $n = 1$ we already proved in Theorem 5.2.4.Let $\mathcal{F} = \{V_\alpha\}_{\alpha \in I}$ be a cover of K by open sets in $\mathbb{R}^n$. Fix $u \in [a_1, b_1]$ and consider $\{u\} \times \underbrace{[a_2, b_2] \times \cdots \times [a_n, b_n]}_{\substack{=C \text{ is compact} \\ \text{by induction on } n}}$ Hence $\{u\} \times C$ is compact because $\mathbb{R}^{n-1} \to \mathbb{R}^n$ which maps $(y_2, y \cdots, y_n) \mapsto (u, y_2, \cdots, y_n)$ or $f(C) = \{u\} \times C$ is continuous.

For each $p = (u, y_2, \cdots, y_n)$ in $\{u\} \times C$ pick an open neighborhood $V_p \in \mathcal{F}$. Hence $V_p \supset (x - \epsilon, x + \epsilon) \times \underbrace{(y_2 - \epsilon, y_2 + \epsilon) \times \cdots \times (y_n - \epsilon, y_n + \epsilon)}_{W_p}$ for some $\epsilon = \epsilon_p$ depending on p



By compactness of $\{u\} \times C$, extract a finite subcover of the cover $\{W_p\}$. Hence $W_{p_1} \cup W_{p_2} \cup \times \cup W_{p_k} \supset \{u\} \times C$. Since its a union of open sets we have in fact $W_{p_1} \cup W_{p_2} \cup \times \cup W_{p_k} \supset (u - \epsilon, u + \epsilon) \times C$ where $\epsilon = \min\{\epsilon_{p_1}, \epsilon_{p_2}, \cdots, \epsilon_{p_k}\}$. Let $\mathcal{F}_u = \{V_{p_1}, V_{p_2} < \cdots, V_{p_k}\}$. So

$$V_{p_1} \cup V_{p_2} \cup \cdots \cup V_{p_k} \supset W_{p_1} \cup W_{p_2} \cup \cdots \cup W_{p_k} \supset (u - \epsilon, u + \epsilon) \times C$$

i.e. this finite subcover $\mathcal{F}_u$ cover not just the slice but a tube around it.

Now as u varies in $[a_1, b_1]$, $(u - \epsilon_u, u + \epsilon_u)$ gives an open cover. Extract a finite subcover $(u_1 - \epsilon_{u_1}, u_1 + \epsilon_{u_1}), (u_2 - \epsilon_{u_2}, u_2 + \epsilon_{u_2}), \cdots, (u_l + \epsilon_{u_l}, u_l + \epsilon_{u_l})$. Then $\mathcal{F}_{u_1} \cup \mathcal{F}_{u_2} \cup \cdots \cup \mathcal{F}_{u_l}$ is a finite subcover of $[a_1, b_1] \times C = K$ $\square$

Question 19

Why the map $\mathbb{R}^{n-1} \to \mathbb{R}^n$ which maps $(y_2, y \cdots, y_n) \mapsto (u, y_2, \cdots, y_n)$ or $f(C) = \{u\} \times C$ is continuous ?

Question 20

X, Y are topological spaces. $K \subset X$ and $Y \subset Y$ are compact subsets. Then $K \times L$ is compact subset of $X \times Y$ where Open sets of $X \times Y$ are $\bigcup$(Open set of X)$\times$(Open set in Y)

Theorem 5.2.7

All norms on $\mathbb{R}^n$ are equivalent

Proof: Enough to show any norm $f \sim \|\cdot\|$

$$\text{i.e } \alpha\|u\| \leq f(u) \leq \beta\|u\| \forall\ u$$
$$\text{i.e } \alpha \leq \frac{f(u)}{\|u\|} \leq \beta\ \forall\ u \forall\ u \neq 0$$

Note that $\frac{f(x)}{\|x\|} = f\left(\frac{x}{\|x\|}\right) = f(u)$ where $u = \frac{x}{\|x\|}$, so $\|u\| = 1$. Hence it is enough to show that

$$\alpha \leq f(u) \leq \beta$$

for any u with $\|u\| = 1$

Let $S = \{u \mid \|u\| = 1\}$ is the unit sphere in $\mathbb{R}^n$, which is closed and bounded

$$\begin{aligned}
S \text{ is closed and bounded} \quad &\Longrightarrow \quad S \text{ is compact} \\
&\Longrightarrow \quad f(S) \text{ is compact in } \mathbb{R} \\
&\Longrightarrow \quad f(S) \text{ is closed and bounded in } \mathbb{R} \\
&\Longrightarrow \quad f(S) \text{ has largest element in } \beta \text{ and smallest} \\
&\qquad\quad\ \text{element } \alpha \text{ such that } \alpha \leq f(S) \leq \beta
\end{aligned}$$

$\square$

Chapter 6

Differentiation

Derivative of f at $a \in \mathbb{R}$ is

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

To take this limit f should be defined in some $(a-\epsilon, a+\epsilon)$ i.e. $f:$ neighborhood of $a \to \mathbb{R}$

Goal: Definition of $f'(a)$ for $a \in$(Some open U in $\mathbb{R}^m$)$\xrightarrow{f} \mathbb{R}^n$, $a, h \in \mathbb{R}^m$

$f(a+h) - f(a)$ makes sense in $\mathbb{R}^n$ but can't divide by h, which is a vector in $\mathbb{R}^m$. If $m = 1$ can use the same definition. $f:$ Open U in a $\mathbb{R} \to \mathbb{R}^n$ which maps $a \mapsto (f_1(a), f_2(a), \cdots, f_n(a))$. If $n = 1$ i.e. $\mathbb{R}^m \supset U \xrightarrow{f} \mathbb{R}$ we have partial derivatives.

Example 6.0.1 (Derivative of $f: \mathbb{R}^m \to \mathbb{R}$)

$f(x, y, z) = x^4 \sin(yz)$. Here

$$\begin{aligned}\frac{\partial f}{\partial x} &= 4x^3 \sin(yz) \\ &= \lim_{h \to 0} \frac{f(x+h, y, z) - f(x, y, z)}{h}\end{aligned}$$

Hence

$$\begin{aligned}\left.\frac{\partial f}{\partial x}\right|_{p=(r,s,t)} &= \lim_{h \to 0} \frac{f(r+h, s, t) - f(r, s, t)}{h} \\ &= \lim_{h \to 0} \frac{f(p + he_1) - f(p)}{h} \qquad [p = re_1 + se_2 + te_3 \text{ using standard basis } e_1, e_2, e_3]\end{aligned}$$

Its a real number if the limit exists

6.1 Partial Derivatives

Definition 6.1.1: Partial Derivative of $f: \mathbb{R}^n \supset U \to \mathbb{R}$

For f :(Open U in $\mathbb{R}^m$)$\to \mathbb{R}^n$, define "i–th partial derivative of f at $a \in U$" to be $\left(\text{Notation } \left.\frac{\partial f}{\partial x_i}\right|_a, \frac{\partial f}{\partial x_i}(a), D_i f(a)\right)$ $\lim\limits_{h \to a} \frac{f(a+he_i)-f(a)}{h}$ $\qquad (i = 1, 2, \cdots, m)$

Note that this limit (if exists) is in $\mathbb{R}^n$.

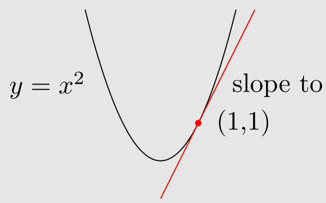
$$\left.\begin{array}{l}\text{If } f = (f_1, f_2, \cdots, f_n)\ (f_i \text{ real} \\ \text{values in function } U \to \mathbb{R})\end{array}\right\} \frac{\partial f}{\partial x_i}(a) = \left(\frac{\partial f_1}{\partial x_i}(a), \frac{\partial f_2}{\partial x_i}(a), \cdots, \frac{\partial f_n}{\partial x_i}(a)\right)$$

So for $f : U \to \mathbb{R}^m$, $a \in U$ we get $\frac{\partial f_j}{\partial x_i}$ where $1 \leq j \leq n$ and $1 \leq i \leq m$. We can arrange these in a matrix of dimensions $n \times m$

$$\begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(a) & \cdots & \dfrac{\partial f_1}{\partial x_m}(a) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_n}{\partial x_1}(a) & \cdots & \dfrac{\partial f_n}{\partial x_m}(a) \end{bmatrix}$$

Note:-

$f'(a)$ can be defined as a linear map $\mathbb{R}^n \to \mathbb{R}^m$
In the old situation $f : \mathbb{R} \to \mathbb{R}$, $f'(a) \in \mathbb{R}$ is a 1×1 matrix, as such it encodes a linear map $\underset{x \mapsto f'(a)x}{\mathbb{R} \to \mathbb{R}}$



The tangent line can be though of as the graph of linear map $\underset{t \mapsto 10t}{\mathbb{R} \to \mathbb{R}}$

6.2 Differentiation

$f'(a)$ is a number such that $\lim\limits_{h \to 0} \frac{f(a+h)-f(a)-f'(a)h}{h} = 0$. Inspired by this for $a \in U(\text{Open in } \mathbb{R}^m) \xrightarrow{f} \mathbb{R}^n$, we define $f'(a)$ is a linear map $\mathbb{R}^m \to \mathbb{R}^n$ such that

$$\lim_{h \to 0} \frac{\|f(a+h) - f(a) - f'(a)h\|}{\|h\|} = 0 \text{ in } \mathbb{R}$$

(h is a small vector $\in \mathbb{R}$)

Definition 6.2.1: Differentiation of $f : \mathbb{R}^m \supset U \to \mathbb{R}^n$

U open set in $\mathbb{R}^m$, $f : U \to \mathbb{R}^n$, $a \in U$ given. We say that f is differentiable at a if there is a linear map $T : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$\lim_{h \to 0} \frac{\|f(a+h) - f(a) - T(h)\|}{\|h\|} = 0$$

i.e $\forall\, \epsilon > 0$, $\exists\, \delta > 0$ s.t.

$$\|h\| < \delta \implies \frac{\|f(a+h) - f(a) - T(h)\|}{\|h\|} < \epsilon$$

We call such a linear map T the derivative of f at a, denoted by $f'(a), D(f(a))$

Note that $f'(a)h$ = Value of linear map $f'(a)$ applied to a vector h

Note:-

If the above limit is 0 w.r.t any norm on $\mathbb{R}^M$ (respectively $\mathbb{R}^n$) then the same limit is 0 w.r.t any other norm on $\mathbb{R}^m$ (respectively $\mathbb{R}^n$) because all norms are equivalent

Theorem 6.2.1

Derivative is unique i.e. if $a \in U(\text{Open in } \mathbb{R}^m) \xrightarrow{f} \mathbb{R}^n$, and

$$\lim_{h \to 0} \frac{\|f(a+h) - f(a) - T(h)\|}{\|h\|} = 0 = \lim_{h \to 0} \frac{\|f(a+h) - f(a) - S(h)\|}{\|h\|}$$

then $T = S$ i.e $T(v) = S(v)\ \forall\, v \in \mathbb{R}^m$

Proof: Let $R = S - T$. Want to show $R(v) = 0 \ \forall \ v \in \mathbb{R}^m$

$$\frac{\|R(h)\|}{\|h\|} = \frac{\|S(h) - T(h)\|}{\|h\|} = \frac{\|(f(a+h) - f(a) - T(h)) - (f(a+h) - f(a) - S(h))\|}{\|h\|}$$
$$\leq \frac{\|(f(a+h) - f(a) - T(h))\|}{\|h\|} + \frac{\|(f(a+h) - f(a) - S(h))\|}{\|h\|}$$

Taking $\lim\limits_{h \to 0}$, we get $\lim\limits_{h \to 0} \frac{\|R(h))\|}{\|h\|} = 0$.

Fix any nonzero v, $\lim\limits_{\lambda \to 0} \lambda v = 0$. Take $h = \lambda v$ ($\lambda \neq 0$)

$$\frac{\|R(h)\|}{\|h\|} = \frac{|\lambda|\|R(v)\|}{|\lambda|\|v\|} = \frac{\|R(v)\|}{\|v\|}$$

Hence

$$0 = \lim_{h \to 0} \frac{\|R(h)\|}{\|h\|} = \lim_{h \to 0} \frac{\|R(v)\|}{\|v\|} \implies \|R(v)\| = 0 \implies Rv = 0$$

□

Question 21

If $f : \underset{v \mapsto Av}{\mathbb{R}^m \to \mathbb{R}^n}$ is a linear map then what is $f' : \mathbb{R}^m \to \mathbb{R}^n$

Solution: See that $f'(a) = f$. (Immediate from definition)

□

Question 22

For an affine map $\underset{v \mapsto Av + c}{\mathbb{R}^m \xrightarrow{g} \mathbb{R}^n}$ for some $c \in \mathbb{R}^n$. Calculate $g'(a)$.

Solution: $g'(a) =$ the map $h \mapsto Ah$

□

Theorem 6.2.2 Matrix of $f'(a)$

Prove that the matrix of $f'(a)$ w.r.t standard basis of $\mathbb{R}^m$ and $\mathbb{R}^n$ is the Jacobian Matrix

Proof: jth column of matrix of $T = T(e_j) \in \mathbb{R}^n$

$$\lim_{\lambda \to 0} \frac{\|f(a + \lambda e_j) - f(a) - T(\lambda e_j)\|}{\|\lambda e_j\|} \text{ by definition of } T = f'(a)$$
$$= \lim_{\lambda \to 0} \frac{\|f(a + \lambda e_j) - f(a) - \lambda T(e_j)\|}{|\lambda|} \begin{cases} T(\lambda e_j) = \lambda T(e_j) \text{ by linearity} \\ \|\lambda e_j\| = |\lambda|\|e_j\| = |\lambda| \end{cases}$$
$$= \lim_{\lambda \to 0} \left\| \frac{f(a + \lambda e_j) - f(a)}{|\lambda|} - \frac{\lambda T(e_j)}{|\lambda|} \right\|$$

Hence for $\lambda > 0$

$$\lim_{\lambda \to 0} \frac{f(a + \lambda e_j) - f(a)}{|\lambda|} - T(e_j) = 0$$

Let $f = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}$. Hence

$$T(e_j) = \lim_{\lambda \to 0} \frac{f(a+\lambda e_j) - f(a)}{\lambda}$$
$$= \lim_{\lambda \to 0} \frac{\begin{bmatrix} f_1(a+\lambda e_j) \\ f_2(a+\lambda e_j) \\ \vdots \\ f_n(a+\lambda e_j) \end{bmatrix} - \begin{bmatrix} f_1(a) \\ f_2(a) \\ \vdots \\ f_n(a) \end{bmatrix}}{\lambda} = \begin{bmatrix} \frac{\partial f_1}{\partial x_j}(a) \\ \frac{\partial f_2}{\partial x_j}(a) \\ \vdots \\ \frac{\partial f_n}{\partial x_j}(a) \end{bmatrix}$$

Matrix of $f'(a)$= Jacobian Matrix

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_m}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(a) & \cdots & \frac{\partial f_n}{\partial x_m}(a) \end{bmatrix}$$

We have proved if $f'(a)$ exists, then all partial derivatives $\frac{\partial f_i}{\partial x_j}$ exists at $x = a$ and make up the matrix of $f'(a)$ $\square$

If all $\frac{\partial f_i}{\partial x_j}$ exists at $x = a$, does not imply f is differentiable at $x = a$?

Question 23

Under Which conditions if all $\frac{\partial f_i}{\partial x_j}$ exists at $x = a$, it implies that f is differentiable at $x = a$?

Theorem 6.2.3

If $f'(a)$ exists then f is continuous at $x = a$

Proof: If $f'(a)$ exists then f is continuous at $x = a \iff \lim_{h \to 0} f(a+h) = f(a) \iff \lim_{h \to 0} \|f(a+h) - f(a)\| = 0$

$$\|f(a+h) - f(a) + T(h) - T(h)\| \le \|f(a+h) - f(a) - T(h)\| + \|T(h)\|$$

Now $\lim_{h \to 0} \frac{\|f(a+h) - f(a) - T(h)\|}{\|h\|} \|h\| = 0 \cdot 0 = 0$ and

$$\lim_{h \to 0} \|T(h)\| = 0 \text{ because } \begin{cases} T \text{ is continuous (being linear) so} \\ T(h) \to T(0) = 0 \end{cases}$$

$\square$

Chapter 7

Examples on Multivariable Differentiation

Example 7.1 (Example where all partial derivatives exist and function is continuous but f' does not exists.)

$$f(x,y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

(i) Is f continuous at origin ?

(ii) Do $\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}$ exist at origin ? elsewhere ?

Solution:

(i) Want $|f(x,y) - f(0,0)| \to 0$ as $(x,y) \to (0,0)$

$$\left| \frac{xy}{\sqrt{x^2+y^2}} \right| \leq \sqrt{\frac{x^2+y^2}{2}} \to 0$$

as $(x,y) \to (0,0)$

(ii)

$$\frac{\partial f}{\partial x} = \frac{y^3}{(x^2+y^2)^{\frac{3}{2}}} \qquad \frac{\partial f}{\partial y} = \frac{x^3}{(x^2+y^2)^{\frac{3}{2}}}$$

Now

$$\left.\frac{\partial f}{\partial x}\right|_{(0,0)} = \lim_{h \to 0} \frac{f(h,0) - f(0,0)}{h} = \lim_{h \to 0} \frac{0-0}{h} = 0$$

Similarly $\left.\frac{\partial f}{\partial y}\right|_{(0,0)} = 0$. So if $f'(0)$ exists then it will be the matrix $\begin{bmatrix} 0 & 0 \end{bmatrix}$. So it will be the zero operator $\implies D_v f(\text{origin}) = 0$ for any direction for any vector v. Let's test for $v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$D_v f(\text{origin}) = \lim_{t \to 0} \frac{f(0+tv) - f(0,0)}{t} = \lim_{t \to 0} \frac{f(t,t)}{t} = \lim_{t \to 0} \frac{t^2}{t\sqrt{2t^2}} \neq 0$$

Thus f is not differentiable at origin. Therefore at least one of the partial derivatives must be discontinuous at origin (here by symmetry both are discontinuous). $\frac{\partial f}{\partial x} = 0$ at origin but $= 1$ at $y-$axis.

Example 7.2 (Example where f' exists but not continuous)

Recall one-variable example $g(x) = \begin{cases} x^2 \sin\frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$. Define $f(x,y) = g(\sqrt{x^2+y^2})$

(i) Is f continuous ?

(ii) Is f differentiable ?

(iii) Is f' continuous at origin ?

Solution:

(i) Because f is composition of two continuous functions. f is continuous.

(ii) Need to check at origin only

Example 7.3

$f(x,y) = \begin{cases} \frac{x^2 y}{x^6+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$

(i) Is f continuous at origin ?

(ii) Calculate the directional derivatives for unit vectors $u = (\cos\theta, \sin\theta)$

(iii) Is f differentiable at origin ?

Solution:

(i)

$$f(x, x^3) = \frac{x^5}{2x^6} = \frac{1}{2x}$$

It has no limit as $x \to 0$. Hence f is not continuous at origin.

(ii)

$$\begin{aligned} D_u f(0) &= \lim_{h\to 0} \frac{f(0+hu) - f(0)}{h} \\ &= \lim_{h\to 0} \frac{f(h\cos\theta, h\sin\theta)}{h} \\ &= \lim_{h\to 0} \frac{1}{h} \frac{h^3\cos^2\theta\sin\theta}{h^6\cos^6\theta + h^2\sin^2\theta} \\ &= \lim_{h\to 0} \frac{\cos^4\theta\sin\theta}{h^4\cos^6\theta + \sin^2\sin\theta} = \frac{\cos^2\theta}{\sin\theta} \qquad \text{when } \sin\theta \neq 0 \end{aligned}$$

When $\sin\theta = 0$, $f = 0$ on $x-$axis. So $D_u f(0) = 0$ for $\theta = 0, \pi, \ldots$. SO $D_u f()$ exists for all u

(iii) If $f'(0)$ exists then it's matrix would be $\begin{bmatrix} & 0 \end{bmatrix}$. But then all directional derivatives would have to be zero because $D_u f(a) = f'(a)v$ which is not possible

Chapter 8

Chain Rule of Differentiation and Operator Norm

8.1 Operator Name

V, W are vector spaces. $\mathcal{L}(V, W) =$ Set of linear maps $V \to W$ is a vector space via $(A+B)(v) = A(v) + B(v)$ and $A(\lambda v) = \lambda A(v)$

If $V = \mathbb{R}^m$ and $\mathbb{R}^n$, $\dim(\mathbb{R}^m, \mathbb{R}^n) = mn$. We can identity $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ with $n \times m$ matrices.

$$\|A\|_{\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)} = \|A\| = \sup_{\|u\|=1} \|A(u)\|$$

This gives a norm because $\|A\| \geq 0$ and $\|A\| = 0 \implies A = 0$ and $\|\lambda A\| = |\lambda| \|A\|$. As $(A+B)(u) = A(u) + B(u)$ we have $\|(A+B)(u)\| \leq$
$A(u)\| + \|B(u)\|$ in W and hemce $\|A+B\| \leq \|A\| + \|B\|$.

Question 24

Why this is well defined ?

Solution: The set $S = \{u \mid \|u\| = 1\}$ is closed and bounded in V, therefore compact. A being linear is continuous. $\therefore A(S)$ is a compact subset of $W \implies A(S)$ is bounded.

$\square$

Basic Properties:-

1. $\|Av\| \leq \|A\|\|v\|$ i.e. $\|Av\|_W \leq \|A\|_{\mathcal{L}} \|v\|_V$

   Proof: If $v = 0$ then we are done. If $v \neq 0$, $u = \frac{v}{\|v\|}$ so $\|u\| = 1$. Hence

$$\|A\| \geq \|Au\| = \left\| A\left(\frac{v}{\|v\|}\right) \right\| = \frac{\|Av\|}{\|v\|}$$

   $\square$

2. $\|A(v)\| \leq M\|v\|$ for all $v \implies \|A\| \leq M$ in fact $\inf\{M \mid \|A(v)\| \leq M\|v\| \ \forall\ v\}$

   Proof: Suppose $\|A(v)\| \leq M\|v\|\ \forall\ v$. In particular $\forall\ v$ with $\|v\| = 1$. So $\|A(v)\| \leq M$.
   Rest exercise: If $L <$ inf of the set show $\exists\ u$ of norm=1 with $\|A(v)\| > L$. $\square$

3. $U \xrightarrow{A} V \xrightarrow{B} W$ linear maps between finite dimensional vector spaces then $\|BA\| \leq \|B\|\|A\|$

   Proof: Take u with $\|u\| = 1$. Then

$$\|BA(u)\| \leq \|B\|\|A(u)\| \leq \|B\|\|A\|\|u\| = \|B\|\|A\|$$

   Now take sup over u. $\square$

Note:-

$\underset{\mathcal{L}(U,V)\oplus\mathcal{L}(V,W)}{A,B} \underset{\to}{\mapsto} \underset{\mathcal{L}(U,W)}{BA}$ is continuous because each slot of matrix of BA is obtained by adding/multiplying entries of A and B.

Question 25

Show $A_n \to A$ in $\mathcal{L}(U,V)$, $B_n \to B$ in $\mathcal{L}(V,W)$ then $B_nA_n \to BA$ in $\mathcal{L}(U,W)$

8.2 Chain Rule

Theorem 8.2.1

Let

$$\begin{array}{ccccc} & & \mathbb{R}^n & & \\ & & \cup & & \\ \mathbb{R}^m \supset U & \xrightarrow{f} & V & \xrightarrow{g} & \mathbb{R}^k \\ \Uplus & & \Uplus & & \\ a & \longmapsto & b & & \end{array}$$

f is differentiable at a. and g is differentiable at $b = f(a)$. Then $g \circ f$ is differentiable at a and

$$(g \circ f)'(a) = \underbrace{g'(f(a))f'(a)}_{\substack{\text{Multiplication}\\ \text{of matrices}}}$$

Proof: **1-Variable Case**

$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$

$$\begin{aligned} \frac{dz}{dx} = \lim_{\Delta x \to 0} \frac{\Delta z}{\Delta x} &= \lim_{\Delta x \to 0} \frac{\Delta z}{\Delta y}\frac{\Delta y}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Delta z}{\Delta y} \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \\ &= \lim_{\Delta y \to 0} \frac{\Delta z}{\Delta y} \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \qquad [\text{As } \Delta x \to 0,\ \Delta y \to 0] \end{aligned}$$

Multi Variable Case

Note that $\mathbb{R}^m \xrightarrow{f'(a)} \mathbb{R}^n \xrightarrow{g'(f(a))} \mathbb{R}^k$

If $T = f'(a)$ then

$$\lim_{h \to 0} \frac{\|f(a+h) - f(a) - T(h)\|}{\|h\|} = 0$$

and $S = g'(f(a)) = g'(b)$ then

$$\lim_{h \to 0} \frac{\|g(b+k) - g(b) - S(k)\|}{\|k\|} = 0$$

Let

$$\alpha(h) = f(a+h) - f(a) - T(h) \qquad \epsilon(h) = \frac{\|\alpha(h)\|}{\|h\|} \to 0 \text{ as } h \to 0$$

$$\beta(k) = g(b+k) - g(b) - S(k) \qquad \eta(k) = \begin{cases} \frac{\|\beta(k)\|}{\|k\|} \to 0 \text{ as } k \to 0 \\ 0 \text{ when } k = 0 \end{cases}$$

η is continuous at $k = 0$. Now note that $\eta : V - b \to \mathbb{R}$ because we ae always taking $b + k$ for η. We want to show that

$$\lim_{h\to 0} \frac{\|g(f(a+h)) - g(f(a)) - ST(h)\|}{\|h\|} = 0 \iff \lim_{k\to 0} \frac{\|g(b+k) - g(b) - ST(h)\|}{\|h\|} = 0$$

where $f(a+h) = b+k \iff k = f(a+h) - f(a)$. We have taken a specific value of k depending on h. So now k is a function of h. Hence $T(h) = f(a+h) - f(a) - \alpha(h) = k - \alpha(h)$

$$\begin{aligned}
&g(b+k) - g(b) - ST(h)\\
=&g(b+k) - g(b) - S(k - \alpha(h))\\
=&g(b+k) - g(b) - S(k) + S(\alpha(h))
\end{aligned}$$

Therefore

$$\frac{\|g(b+k) - g(b) - ST(h)\|}{\|h\|} \leq \frac{\|g(b+k) - g(b) - S(k)\|}{\|h\|} + \frac{\|S(\alpha(h))\|}{\|h\|}$$

want to bound each of these separately

$$\frac{\|S(\alpha(h))\|}{\|h\|} \leq \|S\| \frac{\|\alpha(h)\|}{\|h\|} \to 0$$

Now how to bound the first term. In the first term $\frac{\|\beta(k)\|}{\|h\|} = \eta(k) \frac{\|k\|}{\|h\|}$. Now

$$\begin{aligned}
&k = T(h) + \alpha(h)\\
\implies &\|k\| \leq \|T(h)\| + \|\alpha(h)\|\\
\implies &\frac{\|k\|}{\|h\|} \leq \frac{\|T(h)\|}{\|h\|} + \frac{\|\alpha(h)\|}{\|h\|} \leq \frac{\|T\|\|h\|}{\|h\|} + \frac{\|\alpha(h)\|}{\|h\|} = \|T\| + \frac{\|\alpha(h)\|}{\|h\|}
\end{aligned}$$

Hence

$$\frac{\|\beta(k)\|}{\|k\|} = \eta(k) \frac{\|k\|}{\|h\|} \leq \eta(k) \left[\|T\| + \frac{\|\alpha(h)\|}{\|h\|} \right]$$

As $h \to 0$ $\|T\| + \frac{\|\alpha(h)\|}{\|h\|} \to \|T\| + 0$ which is finite. And as $h \to 0$, $k \to 0 \implies \eta(k) \to 0$ because η is continuous at 0. $\square$

8.3 Special Case of Chain Rule: When $m = k = 1$

Open interval in $\mathbb{R} \xrightarrow{\gamma} \mathbb{R}^n \xrightarrow{g} \mathbb{R}$. γ = parameterized curve in $\mathbb{R}^n$

$$(g \circ \gamma)'(t) = g'(\gamma(t)) \cdot \gamma'(t)$$

$\mathbb{R} \xrightarrow{\gamma} \mathbb{R}^n$ maps $t \to \begin{bmatrix} \gamma_1(t) \\ \vdots \\ \gamma_n(t) \end{bmatrix}$ hence $\gamma'(t) = \begin{bmatrix} \gamma_1'(t) \\ \vdots \\ \gamma_n'(t) \end{bmatrix}$

Now $g'(y) = \left[\left.\frac{\partial g}{\partial x_1}\right|_y \cdots \left.\frac{\partial g}{\partial x_n}\right|_y \right]$. Hence

$$\begin{aligned}
(g \circ \gamma)'(t) &= \left[\left.\frac{\partial g}{\partial x_1}\right|_y \cdots \left.\frac{\partial g}{\partial x_n}\right|_y \right] \begin{bmatrix} \gamma_1'(t) \\ \vdots \\ \gamma_n'(t) \end{bmatrix}\\
&= \begin{bmatrix} \left.\frac{\partial g}{\partial x_1}\right|_y \\ \vdots \\ \left.\frac{\partial g}{\partial x_n}\right|_y \end{bmatrix} \cdot \begin{bmatrix} \gamma_1'(t) \\ \vdots \\ \gamma_n'(t) \end{bmatrix} \qquad \text{[Usual dot product of vectors in } \mathbb{R}^n]
\end{aligned}$$

Call $\begin{bmatrix} \left.\frac{\partial g}{\partial x_1}\right|_y \\ \vdots \\ \left.\frac{\partial g}{\partial x_n}\right|_y \end{bmatrix} = \nabla g(\gamma(t)) =$ Gradient of g at the point $\gamma(t)$. Hence $(g \circ \gamma)'(t) = \nabla g(\gamma(t)) \cdot \gamma'(t)$

Question 26

Fix $u, v \in \mathbb{R}^n$ and take parametrized curve $\gamma(t) = u + tv$. What does the above equation give at $t = 0$ (for a given function g)

Chapter 9

Mean Value Theorem

We will use Euclidean norm on $\mathbb{R}^n$ and have Cauchy-Schwarz Inequality $|v \cdot w| \leq \|v\|\|w\|$.

Theorem 9.1 1-Variable MVT

If $f : [a, b] \to \mathbb{R}$ continuous and f' exists on (a, b), then $\exists\ c \in (a, b)$ s.t

$$f(a) - f(b) = f'(c)(b - a)$$

Proof: 1-variable MVT

$\Uparrow$ Via Rolle's Theorem, using $f'(\text{extremum}) = 0$

Extreme Value Theorem

$\Uparrow$ $[a, b]$ is compact, f is continuous $\implies$ $f([a, b])$ is compact in $\mathbb{R}$ $\implies$ closed and bounded

Heine Borel Theorem $\square$

Question 27

First consider $f : [a, b] \to \mathbb{R}^n$ continuous and f' exists on (a, b). Is there a $c \in (a, b)$ s.t.

$$\|f(b) - f(a)\| \stackrel{?}{=} \|f'(c)\|(b - a)$$

Solution: No. For example $f : [0, 2\pi] \to \mathbb{R}^2$ which maps $t \mapsto (\sin t, \cos t)$. Then $f'(t) = (\cos t, -\sin t)$ and $\|f'(t)\| = 1$. $f(2\pi) - f(0) = (0, 0)$

$\square$

Theorem 9.2 MVT of Real-Valued Functions

Let f be a continuous function $[a, b] \to \mathbb{R}^n$ and $f'(c)$ exists $\forall\ c \in (a, b)$. Then $\exists\ c \in (a, b)$ s.t

$$\|f(b) - f(a)\| \leq (b - a)\|f'(c)\|$$

(Here the norm is Euclidean norm. For the inequality which norm we take does matter.)

Proof: Clever use if 1−Variable MVT. We want to bound norm of $f(b) - f(a) = z$.

Idea: Dot with z and then use Cauchy Schwarz

$$\begin{tikzcd}
{[a,b]} \arrow[r, "f"] & \mathbb{R}^n \arrow[r, "g=\langle -, z\rangle"] & \mathbb{R} \\
t \arrow[r, mapsto] \arrow[rr, bend right, "\phi(t)"'] & f(t) \arrow[r, mapsto] & \langle f(t), z\rangle
\end{tikzcd}$$

Notice $g : \mathbb{R}^n \to \mathbb{R}$ where g maps $x \longmapsto \langle x, z\rangle = x_1z_1 + \cdots + x_nz_n$. Hence g is differentiable and $g'(x) = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix}$. Now we can apply MVT to ϕ.

$$\phi(b) - \phi(a) = (b-a)\phi'(c)$$

$$\begin{aligned} LHS = \phi(b) &= \phi(a) \\ &= \langle f(b), z\rangle - \langle f(a), z\rangle \\ &= \langle f(b) - f(a), z\rangle = \langle z, z\rangle = \|z\|^2 \end{aligned}$$

And

$$\phi'(c) = (g \circ f)'(c) = g'(f(c)) \circ f'(c) = \langle z, f'(c)\rangle$$

Therefore

$$\|z\|^2 = \langle z, f'(c)\rangle \leq (b-a)\|z\|\|f'(c)\|$$

Nothing to prove if $\|z\| = 0$ and else cancel $\|z\|$ from both sides. $\square$

Theorem 9.3 General Multivariable MVT

Let $\mathbb{R}^m \supset$ Convex Open $U \xrightarrow{f} \mathbb{R}^n$, f differentiable on U and $\|f'(x)\| < M\ \forall\ x \in U$. Then $\forall\ a, b \in U$

$$\|f(b) - f(a)\| \leq M\|b-a\|$$

Proof: Given $a, b \in U$ holds $\gamma : [0,1] \to U$ which maps $t \to a + t(b-a)$ [This is valid by convexity]. Apply MVT to

$$\begin{tikzcd} {[0,1]} \arrow[r, "\gamma"] \arrow[rr, bend right, "g"'] & U \arrow[r, "f"] & \mathbb{R}^n \end{tikzcd}$$

We get $\in (0,1)$ such that $\|g(1) - g(0)\| \leq \|g'(t)\|$ i.e

$$\|f(b) - f(a)\| \leq \| \underbrace{f'(\gamma(t)) \circ \gamma'(t)}_{\text{Matrix Vector Multiplication}} \| \underbrace{\leq}_{\downarrow \text{ Justify}} M\|b-a\|$$

$\square$

Suppose U is convex in $\mathbb{R}^m$, $f : U \to \mathbb{R}^n$ and $f'(a) = 0\ \forall\ a \in U$. Then $f =$ Constant because

$$\|f(b) - f(a)\| \leq 0\|b-a\|$$

$\forall\ a, b \in U$ by MVT. What happens if U is open but not convex. If U is connected the conclusion is again true.

Definition 9.1: Connected Set in $\mathbb{R}^n$

A set S in $\mathbb{R}^n$ is (path)connected if $\forall\ a, b \in S\ \exists$ continuous function

$$\begin{aligned} \gamma :\ [0,1] &\longrightarrow S \\ 0 &\longmapsto a \\ 1 &\longmapsto b \end{aligned}$$

Theorem 9.4

If S is connected open set in $\mathbb{R}^n$ and $f : S \to \mathbb{R}^n$ is differentiable on U with $f'(a) = 0\ \forall\ a \in U$ then $f(a) =$ Constant.

Proof: $\forall\ x \in \gamma([0,1])$ find $B_r(x) \subset S$. $\gamma([0,1])$ is compact. So $\exists$ finite subcover of $\gamma([0,1])$ by balls around $\gamma(t_1), \gamma(t_2), \ldots, \gamma(t_N)$.

Order these alls so that $a \in$ first ball and $b \in$ last ball, any two consecutive balls overlap. This gives piecewise linear path from $a \to b$. Use MVT for each segment. $\square$

Chapter 10

Higher Derivatives

10.1 Class C^1 Functions

Open U in $\mathbb{R}^m \xrightarrow{f} \mathbb{R}^n$. $D(f(a))$ is a linear map $\mathbb{R}^m \to \mathbb{R}^n$ i.e $D(f(a)) \in \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$. If f is differentiable at each $a \in U$, then we get a function $Df : U \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ which maps $a \longmapsto D(f(a)) = f'(a)$. We can ask about continuity and differentiability of this map Df.

We want to consider $C^1(U)$ functions which are all functions that are differentiable at each $a \in U$ and Df is continuous i.e $C^1(U) =$ Set of continuously differentiable functions

Definition 10.1.1: C^1 Functions

U open in $\mathbb{R}^m \xrightarrow{f} \mathbb{R}^n$. Suppose $f'(a)$ exists $\forall\ a \in U$ then we get a function

$$\begin{aligned} f : U &\longrightarrow \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n) \backsimeq \mathbb{R}^{mn} \\ a &\longmapsto f'(a) \end{aligned}$$

which maps $a \longmapsto f'(a)$. We say $f \in C^1(U)$, "f is continuously differentiable" if $f'(a)$ exists for each a and f' is a continuous function

Theorem 10.1.1

A function $f : U$ Open in $\mathbb{R}^m \to \mathbb{R}^n$ is $C^1(U)$ $\iff$ $\left.\dfrac{\partial f_i}{\partial x_j}\right|_a$ exists at each $a \in U$ and are continuous functions

Proof: **If Part:-**

1. $a \in U$ Open in $\mathbb{R}^m \xrightarrow{f} \mathbb{R}^n$ s.t $f(a) = \begin{bmatrix} f_1(a) \\ \vdots \\ f_n(a) \end{bmatrix}$, $\lim\limits_{h \to 0} \dfrac{f(a+h) - f(a) - Th}{\|h\|} = 0$Matrix of T w.r.t standard basis of $\mathbb{R}^m$ and $\mathbb{R}^n$ is

$$T = \begin{bmatrix} \left.\dfrac{\partial f}{\partial x_1}\right|_a & \cdots & \left.\dfrac{\partial f}{\partial x_m}\right|_a \end{bmatrix} = \begin{bmatrix} \left.\dfrac{\partial f_1}{\partial x_1}\right|_a & \cdots & \left.\dfrac{\partial f_1}{\partial x_m}\right|_a \\ \vdots & \ddots & \vdots \\ \left.\dfrac{\partial f_n}{\partial x_1}\right|_a & \cdots & \left.\dfrac{\partial f_n}{\partial x_m}\right|_a \end{bmatrix}$$

2. $\lim\limits_{x \to v} f(x) = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \iff \lim\limits_{x \to v} f_I(x) = b_i$ for each $i = 1, 2, \ldots, n$

By 1 and 2 the proof of forward direction is obvious

Only If Part:-

If we prove that $f'(a)$ exists for each $a \in U$ then f' is automatically continuous because by 1 the matrix of $f'(a)$ must be the Jacobian Matrix and we are given that all entries of this matrix namely the functions $\dfrac{\partial f_i}{\partial x_j}$ are continuous so apply 2

Another reduction: We may assume that $n = 1$ because this case in general, it follows immediately that for $f = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$,

$$f'(a) = \begin{bmatrix} f_1(a) \\ \vdots \\ f_n(a) \end{bmatrix} = \begin{bmatrix} \left.\dfrac{\partial f}{\partial x_1}\right|_a & \cdots & \left.\dfrac{\partial f}{\partial x_m}\right|_a \end{bmatrix}$$

Hence

$$\frac{f(a+h) - f(a) - Th}{\|h\|} = \frac{1}{\|h\|}\left(\begin{bmatrix} f_1(a+h) \\ \vdots \\ f_n(a+h) \end{bmatrix} - \begin{bmatrix} f_1(a) \\ \vdots \\ f_n(a) \end{bmatrix} - \begin{bmatrix} f_1'(a)h \\ \vdots \\ f_n'(a)h \end{bmatrix} \right)$$

$\lim\limits_{h \to 0}$ of this $= 0$ because in each slot the limits is 0 by $n = 1$ case which we have assumed, and will prove now.

> **Note:-**
>
> Proof of the fact that in case of $n = 1$ if $\dfrac{\partial f_i}{\partial x_j}$ $(j = 1, 2, \ldots, m)$ are continuous functions $U \to \mathbb{R}$ then $f'(a)$ exists for each $a \in U$

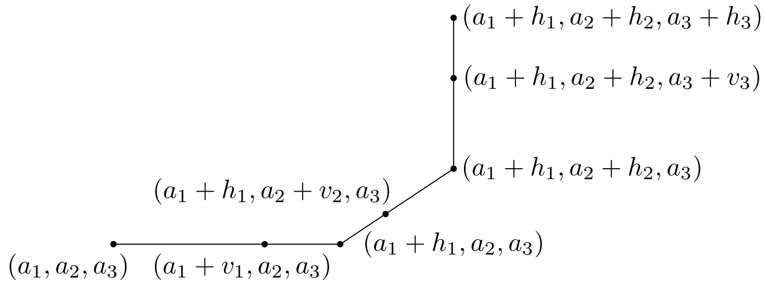
We want to show $f'(a) = \begin{bmatrix} \left.\frac{\partial f}{\partial x_1}\right|_a & \cdots & \left.\frac{\partial f}{\partial x_m}\right|_a \end{bmatrix}$ i.e $\lim\limits_{h \to 0} \dfrac{f(a+h) - f(a) - Th}{\|h\|} = 0$. We want to bound the numerator. Fix $a = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}^T$. Let $h = \begin{bmatrix} h_1 & \cdots & h_m \end{bmatrix}^T$. Now choose $r > 0$ such that $B_r(a) \subset U$ and restrict h such that $\|h\| < r$.

$$Th = \sum_j \left.\frac{\partial f}{\partial x_j}\right|_a h_j = \left\langle \underbrace{\begin{bmatrix} \left.\frac{\partial f}{\partial x_1}\right|_a \\ \vdots \\ \left.\frac{\partial f}{\partial x_m}\right|_a \end{bmatrix}}_{\substack{\| \\ q}}, \begin{bmatrix} h_1 \\ \vdots \\ h_m \end{bmatrix} \right\rangle$$

And $f(a+h) - f(a) = f(a_1, h_1, \ldots, a_m + h_m) - f(a_1, \ldots, a_n)$

Idea: Bound this in terms of partial derivatives using the mean value theorem (ordinary 1-variable version, which is applicable because each $\frac{\partial f}{\partial x_j}$ is continuous)

$$\begin{aligned} f(a+h) - f(a) &= f(a_1 + h_1, a_2, \ldots, a_m) - f(a) \\ &\quad + f(a_1 + h_1, a_2 + h_2, \ldots, a_m) - f(a_1 + h_1, a_2, \ldots, a_m) \\ &\qquad\vdots \qquad\qquad\qquad\qquad\qquad \vdots \\ &\quad + f(a_1 + h_1, a_2 + h_2, \ldots, a_m + h_m) - f(a_1 + h_1, a_2 + h_2, \ldots, a_{m-1} + h_{m-1}, a_m) \\ &\overset{\text{MVT}}{=} \left.\frac{\partial f}{\partial x_1}\right|_{v_1} h_1 + \left.\frac{\partial f}{\partial x_2}\right|_{v_2} h_2 + \cdots + \left.\frac{\partial f}{\partial x_m}\right|_{v_m} h_m = \left\langle \underbrace{\begin{bmatrix} \left.\frac{\partial f}{\partial x_1}\right|_{v_1} \\ \vdots \\ \left.\frac{\partial f}{\partial x_m}\right|_{v_m} \end{bmatrix}}_{\substack{\| \\ p}}, \begin{bmatrix} h_1 \\ \vdots \\ h_m \end{bmatrix} \right\rangle \end{aligned}$$



Notice that $v_1, v_2, \ldots, v_m$ are functions of h. Putting together we get

$$\begin{aligned}\frac{f(a+h)-f(a)-Th}{\|h\|} &= \frac{1}{\|h\|}\left\langle h, \begin{bmatrix} \left.\frac{\partial f}{\partial x_1}\right|_{v_1} - \left.\frac{\partial f}{\partial x_1}\right|_a \\ \vdots \\ \left.\frac{\partial f}{\partial x_m}\right|_{v_m} - \left.\frac{\partial f}{\partial x_1}\right|_a \end{bmatrix} \right\rangle \\ &= \frac{1}{\|h\|}\|h\|\|p-q\| = \|p-q\|\end{aligned}$$

Showing $\lim_{h\to 0}\|p-q\| = 0$ is enough to complete the proof

$$\lim_{h\to 0}\|p-q\| = 0 \impliedby p-q \to 0 \text{ as } h \to 0 \impliedby \left.\frac{\partial f}{\partial x_i}\right|_{v_i} - \left.\frac{\partial f}{\partial x_i}\right|_a \to 0 \text{ as } h \to 0$$

which is true because $\frac{\partial f}{\partial x_i}$ is a continuous function. More formally choose $\|h\| < \delta$ s.t

$$\left\|\frac{\partial f}{\partial x_i}(b) - \frac{\partial f}{\partial x_1}(a)\right\| < \frac{\epsilon}{m} \;\forall\; b \in B_\delta(a)$$

That ensures $\|p-q\| < \epsilon$ by triangle inequality.

Question 28

$$f(x,y) = \begin{cases} \frac{xy^2}{xx^2+y^4} & (x,y) \neq (0,0) \\ 0 & \text{else} \end{cases}$$

1. Calculate all directional derivatives in particular $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$
2. Does $f'(a,b)$ exists at all $a, b \in \mathbb{R}$
3. Is f continuous everywhere

□

10.2 Higher Derivatives and Class C^k functions

Now we want to define class C^k of functions for $k \geq 0$. [Like in 1-Variable case. Usefull for Taylor's theorem].

Now second derivative of $f : U \to \mathbb{R}^n$ at $a \in U \subset \mathbb{R}^m$ = derivative at a of $f' : U \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n) \backsimeq \mathbb{R}^{mn}$. $\therefore f''(a)$ or $D^2f(a) : \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$. Matrix of $f''(a)$ has mn rows and m columns and equals to $\frac{\partial f'_{pq}}{\partial x_j}$ where $p = 1, \ldots, n$, $q = 1, \ldots, m$ and $j = 1, \ldots, m$

Question 29

Do $\frac{\partial}{\partial x_l}$ and $\frac{\partial}{\partial x_k}$ commute ?

Solution: No but actually yes under some good conditions. We discussed here

$\square$

If $f''(a)$ exists for all $a \in U$, then we get a function f'' or $D^2 f : U \to \mathcal{L}(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n))$ which maps $a \mapsto f''(a)$. Dimension of the RHS is $m^2 n$. Now we can ask about continuity and differentiability of f''

Definition 10.2.1: C^k Functions

Open $U \subset \mathbb{R}^m$, $f : U \to \mathbb{R}^n$. Then

$$\begin{array}{lcl} f \text{ is } C^0 \text{ if} & & f \text{ is continuous} \\ f \text{ is } C^1 \text{ if} & f'(a) \text{ exists } \forall\, a \in U \text{ and} & f' \text{ is continuous} \\ f \text{ is } C^2 \text{ if} & f''(a) \text{ exists } \forall\, a \in U \text{ and} & f'' \text{ is continuous} \\ \vdots & \vdots & \vdots \\ f \text{ is } C^k \text{ if} & f^{(k)}(a) \text{ exists } \forall\, a \in U \text{ and} & f^{(k)} \text{ is continuous} \end{array}$$

Note:-

How to understand $\mathcal{L}(V, \mathcal{L}(U, W))$ where U, V, W are vector spaces. Just set theoretically

$$\begin{array}{ccccccc} \text{Maps}(& A & , \text{Maps}(B , C)) & \backsimeq & \text{Maps}(& A \times B , C &) \\ & \downarrow & & & & \downarrow & \\ & f(a) & : \quad B \longrightarrow C & \Longleftrightarrow & & \phi \longmapsto \phi(a,b) & \\ & & \quad b \longmapsto \phi(a,b) & & & & \end{array}$$

Under this dictionary, maps in $\mathcal{L}(V, \mathcal{L}(U, W))$ must correspond to some special kind of maps $V \times U \to W$.

Hence we can say $\mathcal{L}(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n))$ is equivalent to the space of maps $\mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n$ which is space of bilinear maps from $\mathbb{R}^m \times \mathbb{R}^m$ to $\mathbb{R}^n$

Component functions of $f' : U \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ are precisely $\frac{\partial f_i}{\partial x_j}$ where $i = 1, \dots, n$ and $j = 1, \dots, m$. So matrix of $f''(a)$ w.r.t standard basis of $\mathbb{R}^m$ and $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ will consist of numbers $\left(\frac{\partial}{\partial x_k}\left(\frac{\partial}{\partial x_j} f_i\right)\right)(a)$. If $f''(a)$ exists then these are generated to exist. If $f''(a)$ exists at each $a \in U$ then we have the function

$$\begin{aligned} f'' : U &\longrightarrow \mathcal{L}(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)) \backsimeq \text{Space of bilinear maps } \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n \\ a &\longmapsto f''(a) \end{aligned}$$

$$f'' \text{ is } C^2(U) \overset{\text{definition}}{\Longleftrightarrow} f'' \text{ is continuous on } U \iff \frac{\partial}{\partial x_k}\frac{\partial}{\partial x_j} f_i \text{ are continuous functions } U \to \mathbb{R}$$

Theorem 10.2.1

Let U be open in $\mathbb{R}^2$ and $f : U \to \mathbb{R}$. $(a, b) \in \mathbb{R}$ and $U \supset Q(h, k) = [a, a+h] \times [b, b+k]$. Define

$$\begin{aligned} \Delta(h,k) &= f(a+h, b+k) - f(a, b+k) - f(a+h, b) + f(a, b) \\ &= [f(a+h, b+k) - f(a+h, b)] - [f(a, b+k) - f(a, b)] \end{aligned}$$

Then $\exists\ (s, t) \in$ interior of the rectangle $Q(h, k)$ such that

$$\Delta(f, Q) = hk \left(\frac{\partial}{\partial y}\frac{\partial}{\partial x}\right) f(s, t) \coloneqq D_{21} f(s, t)$$

Proof. $U(x) = f(x, b+k) - f(x, b)$. Hence

$$\Delta(f, Q) = U(a+h) - U(a)$$

By MVT we have $s \in (a, a+h)$ such that

$$\Delta(f,Q) = hU'(s) = h\left[\frac{\partial f}{\partial x}(s, b+k) - \frac{\partial f}{\partial x}(s,b)\right]$$

Apply MVT again and we get $t \in (b, b+k)$ such that

$$\frac{\partial f}{\partial x}(s, b+k) - \frac{\partial f}{\partial x}(s,b) = k\left(\frac{\partial}{\partial y}\frac{\partial}{\partial x}\right) f(s,t)$$

And hence

$$\Delta(f,Q) = hk\left(\frac{\partial}{\partial y}\frac{\partial}{\partial x}\right) f(s,t)$$

□

Theorem 10.2.2
Suppose for f, $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial}{\partial y}\frac{\partial}{\partial x} f$ exist everywhere and $D_{21} f = \frac{\partial}{\partial y}\frac{\partial}{\partial x} f$ is continuous at (a,b). Then $D_{12} f(a,b) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} f\Big|_{a,b}$ exists and

$$D_{21} f(a,b) = D_{12} f(a,b)$$

Proof. Let $\epsilon > 0$ then continuity of D_{21} means that $\exists\ \delta > 0$ such that $\forall\ h,k$ with $\max(|h|,|k|) < \delta$

$$|D_{21} f(x,y) - D_{21} f(a,b)| < \epsilon$$

$\forall\ x,y \in Q(h,k)$

Take h,k as above and use the Theorem 10.2.1 to find (s,t) such that $\Delta(f,Q) = hkD_{21} f(s,t)$. So

$$\left|\frac{\Delta(f,Q)}{hk} - D_{21} f(a,b)\right| < \epsilon \text{ i.e. } \left|\frac{1}{h}\left(\frac{f(a+h,b+k) - f(a+h,b)}{k} - \frac{f(a,b+k) - f(a,b)}{k}\right) - D_{21} f(a,b)\right| < \epsilon$$

Take limits as $k \to 0$

$$\left|\frac{1}{h}\left(\frac{\partial f}{\partial y} f(a+h,b) - \frac{\partial f}{\partial y} f(a,b)\right) - D_{21} f(a,b)\right| < \epsilon$$

As we take limit $h \to 0$ the quantity $\frac{1}{h}\left(\frac{\partial f}{\partial y} f(a+h,b) - \frac{\partial f}{\partial y} f(a,b)\right)$ actually exists and is equal to $D_{21} f(a,b)$ i.e. $\frac{\partial}{\partial x}\left(\frac{\partial}{\partial y} f(a,b)\right) = D_{12} f(a,b)$ exists and is equal to $D_{21} f(a,b)$ □

Corollary 10.2.1
If f is $C^2(U)$ then $D_{21} f = D_{12} f$ at each point of U

Theorem 10.2.3
Let $U \subset \mathbb{R}^m$ and $f : U \to \mathbb{R}^n$ a C^k map i.e. $k-$th total derivative $f^{(k)}$ exists and is continuous on U then

$$D_{i_1 i_2 \cdots i_k} f = D_{i_{\sigma(1)} i_{\sigma(2)} \cdots i_{\sigma(k)}} f$$

e.g $D_{24714} f = D_{42417}(f)$

Proof. May take $m = 1$ and work with component real valued functions for $k > 2$ keep all but two variables fixed and use earlier result for requisite partial derivative of f. Any permutation can be realized as a sequence of transpositions □

Chapter 11

Multivariable Taylor Theorem

In one variable

$$f(a+h) = f(a) + f'(a)\frac{h}{1!} + f''(a)\frac{h^2}{2!} + \cdots + f^{(n-1)}(a)\frac{h^{n-1}}{(n-1)!} + f^{(n)}(c)\frac{h^n}{n!}$$

for a c between $a, a+h$.

Theorem 11.1 Multivariable Taylor Theorem

Let $U \subset \mathbb{R}^n$ open and $f : U \to \mathbb{R}^m$ a C^m map $(m \geq 1)$. Given $a \in U$, for any h in some neighborhood W of origin, O we have $W + a \subset U$

$$f(a+h) = f(a) + f'(a)\tfrac{h}{1!} + f''(a)\tfrac{h^2}{2!} + \cdots + f^{(m-1)}(a)\frac{h^{m-1}}{(m-1)!} + f^{(m)}(c)\frac{h^m}{m!}$$

$$\begin{bmatrix} D_1 f(a) & D_2 f(a) & \cdots & D_n f(a) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix} \qquad \frac{\Sigma \text{ terms like } D_{ij} f(a) h_i h_j}{2!}$$

Hence

$$f(a+h) = \sum_{k=0}^{m-1} \sum_{s_1+s_2+\cdots+s_n=k} \frac{(D_1^{s_1} \cdots D_n^{s_n} f)(a)}{s_1! s_2! \cdots s_n!} h_1^{s_1} h_2^{s_2} \cdots h_n^{s_n} + \underbrace{r(h)}_{\text{remainder term}}$$

where $r(h)$ is of the form

$$\sum_{s_1+s_2+\cdots+s_n=m} \frac{(D_1^{s_1} \cdots D_n^{s_n} f)(a+\theta h)}{s_1! s_2! \cdots s_n!} h_1^{s_1} h_2^{s_2} \cdots h_n^{s_n}$$

where $\theta \in (0,1)$

Note:-

$\frac{r(h)}{\|h\|^{m-1}} \to 0$ as $h \to 0$

In one-variable $f : [a,b] \to \mathbb{R}$. Then $f^{(0)}, f^{(1)}, \ldots, f^{(m-1)}$ exists in $[a,b]$ and $f^{(m)}$ exists in (a,b). Suppose $s, t \in [a,b]$. Then there exists θ exactly between s and t such that

$$f(t) = \underbrace{f(s) + f'(s)(t-s) + \cdots + \frac{f^{(m-1)}(s)}{(m-1)!}(t-s)^{m-1}}_{p(t)} + \frac{f^{(m)}(\theta)}{(m)!}(t-s)^{m-1}$$

. Then

$$p(s) = f(s), p'(s) = f'(s), p''(s) = f''(s), \dots, p^{(m-1)}(s) = f^{(m-1)}(s) \text{ and } p^{(m)}(x) = 0 \text{ identically}$$

So for $g(x) = f(x) - p(x)$

$$g(s) = g'(s) = \cdots = g^{(m-1)}(s) = 0$$

.

Idea: Use MVT on $g, g', \dots, g^{(m-1)}$ on $[s,t]$

If $g(t) = 0$ then with $g(s) = 0$ we get (by Rolle's theorem) θ_1 between s and t such that $g'(\theta_1) = 0$. Now $g'(\theta_1) = 0$ and $g'(s) = 0 \implies$ we get θ_2 between θ_1 and s such that $g''(\theta_2) = 0$. Now $g''(\theta_2) = 0$ and $g''(s) = 0 \implies$ we get θ_3 between θ_2 and s such that $g'''(\theta_3) = 0$ and so on.. till we get θ_m with $g^{(m)}(\theta_m) = 0$. Take $\theta = \theta_m$.

But is $g(t) = 0$? $g(t) = f(t) - p(t)$ need not be zero.

Idea: We can adjust g by constant $M(x-s)^m$ without affecting $g(s) = g'(s) = \cdots = g^{(m-1)}(s) = 0$ and we also want to apply the Rolle's theorem. Adjust constant M to make $g(t) = 0$

New $g(x) = f(x) - p(x) - M(x-s)^m$ such that $g(t) = f(t) - p(t) - M(t-s)^m = 0$. Hence

$$M = \frac{f(t) - p(t)}{(t-s)^m}$$

We get $g^{(m)}(\theta) = f^{(m)}(\theta) - 0 - m!M = 0$. S

$$M = \frac{f^{(m)}(\theta)}{m!}$$

Equate these two expressions for M and solve for $f(\theta)$ to get the result.

Question 30

Carry out proof of multivariable taylor' theorem following the strategy sketched in the class, specially using the chain rule to calculate $\frac{d^n}{dt^n} f(a + th)$

Question 31

In 'some sense', the one-variable Taylor's Theorem for $f(a+th)$ stays valid in multivariable case.

It is enough to proof for $m = 1$. We have $a \in U \subseteq \mathbb{R}^n \xrightarrow{f} \mathbb{R}$, f is C^m. Then there is a neighborhood W of origin in $\mathbb{R}^n$ such that for any $h \in W$ we have $a + h \in U$ and

$$f(a+h) = f(a) + f'(a)\tfrac{h}{1!} + f''(a)\tfrac{h^2}{2!} + \cdots + f^{(m-1)}(a)\frac{h^{m-1}}{(m-1)!} + r(h)$$

where $r(h) = \frac{f^{(m)}(a+\theta h)}{m!} h^m$ for some $\theta \in (0,1)$ but need to make sense of this.

Proof. Use one-variable taylor's theorem for the composite

$$[0,1] \longrightarrow a + W \subset U \xrightarrow{\quad f \quad} \mathbb{R}$$

$$t \longmapsto a + th \longmapsto f(a+th) = g(t)$$

g is C^m because the map $t \mapsto a + th$ is C^∞Hence

$$f(a+h) = g(1) = \sum_{k=0}^{m-1} \frac{g^{(k)}(0)}{k!} + \frac{g^{(m)}(\theta)}{m!}$$

for some $\theta \in (0,1)$. Thus we will be done by showing

$$g^{(k)}(t) = \sum_{s_1+s_2+\cdots+s_n=k} \frac{k!}{s_1!s_2!\cdots s_n!} D_1^{s_1}\cdots D_n^{s_n} f(a+th) h_1^{s_1} h_2^{s_2}\cdots h_n^{s_n}$$
$$= \sum_{1\le i_1,\ldots,i_k\le n} D_{i_1}\cdots D_{i_k} f(a+th) h_{i_1} h_{i_2}\cdots h_{i_n}$$

Using chain rule for $k=1$

$$\begin{aligned} g'(t) &= \frac{d}{dt} f(a+th) \\ &= f'(a+th)\frac{d}{dt}(a+th) \\ &= f'(a+th)h \\ &= \sum_{i=1}^{n} D_i f(a+th) h_i \end{aligned}$$

For $k=2$

$$g''(t) + \frac{d}{dt} g'(t) = \frac{d}{dt}\sum_{i=1}^{n} D_i f(a+th) h_i = \sum_{i=1}^{n} \frac{d}{dt} D_i f(a+th) h_i$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} D_j D_i f(a+th) h_i h_j = \sum_{1\le i,j\le n} D_j D_i f(a+th) h_i$$

Continue like this $\square$

Addendum to Taylor's Formula: Bounding the error term

For $a \in U \subseteq \mathbb{R}^n$ and f of class C^m from U to $\mathbb{R}$ we know that for $h \in$ some ball B around origin, we have $a + B \subset U$ and

$$f(a+h) = \sum_{k=0}^{m-1} \sum_{s_1+s_2+\cdots+s_n=k} \frac{(D_1^{s_1}\cdots D_n^{s_n} f)(a)}{s_1!s_2!\cdots s_n!} h_1^{s_1} h_2^{s_2}\cdots h_n^{s_n} + r(h)$$

where $r(h)$ is of the form

$$\sum_{s_1+s_2+\cdots+s_n=m} \frac{(D_1^{s_1}\cdots D_n^{s_n} f)(a+\theta h)}{s_1!s_2!\cdots s_n!} h_1^{s_1} h_2^{s_2}\cdots h_n^{s_n}$$

where $\theta \in (0,1)$

Now because $a+\overline{B}$ is compact and $D_1^{s_1}\cdots D_n^{s_n} f$ is continuous on U, we can find a constant c such that for any $s_1,\ldots,s_n$ with $\sum_{i=)}^{n} = m$

$$\left| \frac{D_1^{s_1}\cdots D_n^{s_n} f(a+x)}{s_1!s_2!\cdots s_n!} \right| < c$$

for each $h \in \overline{B}$ Also $|h_i| \le \|h\|$. Therefore

$$|r(h)| < \sum_{s_1+\cdots+s_n=m} c\|h\|^m = k\|h\|^m$$

and therefore $\frac{r(h)}{\|h\|^{m-1}} \to 0$ as $h \to 0$

Chapter 12

Maximum and Minimum of Multivariable Functions

For a C^3 function (in a neighborhood of a in $\mathbb{R}$), by Taylor's Theorem

$$f(a+h) = f(a) + f'(a)h + \frac{1}{2}f''(a)h^2 + \underbrace{\frac{1}{6}f'''\begin{pmatrix}\text{some point}\\ \text{between}\\ a \ \text{ and } \ a+h\end{pmatrix}h^3}_{\substack{\text{Remainder term } r(h)\\ \frac{r(h)}{h^2}\to 0 \text{ as } h\to 0}}$$

Suppose $f'(a) = 0$ "a is a critical point of f". Then

$$\frac{f(a+h)-f(a)}{h^2} = \frac{1}{2}f''(a) + \frac{r(h)}{h^2}$$

If $f''(a) > 0$ then f has a local minimum at a because choose $\delta > 0$ such that $|h| < \delta$, $\left|\frac{r(h)}{h^2}\right| < \frac{1}{2}f''(a)$. Then $RHS > 0 \; \forall \; h$ such that $|h| < \delta$ and so for $h \in (-\delta, \delta)$, $f(a+h) > f(a)$ i.e. $f(a)$ is minimum value of f in the neighborhood $(a-\delta, a+\delta)$. Similarly $f''(a) < 0$ then f has a local maximum at a.

We want to find an analogy of this for multivariable case

$f :$ (open U in $\mathbb{R}^n$) $\to \mathbb{R}$ a C^3 function. Then for $h \in$ some open neighborhood W of origin, $a + h \in U$

$$f(a+h) = f(a) + f'(a)h + \frac{1}{2}f''(a)(h,h) + \underbrace{\frac{1}{6}f'''\begin{pmatrix}\text{some point}\\ \text{between}\\ a \ \text{ and } \ a+h\end{pmatrix}(h,h,h)}_{\substack{\text{Remainder term } r(h)\\ \frac{r(h)}{h^2}\to 0 \text{ as } h\to 0}}$$

$$= f(a) + \begin{bmatrix} D_1 & \cdots & D_n \end{bmatrix}\begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} + \frac{1}{2}\sum_{i,j} D_iD_jf(a)h_ih_j + r(h)$$

$$= f(a) + \begin{bmatrix} D_1 & \dots & D_n \end{bmatrix}\begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} + \frac{1}{2}\begin{bmatrix} h_1 & \cdots & h_n \end{bmatrix}\left[D_iD_jf(a)\right]\begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} + r(h)$$

$$= f(a) + \nabla f(a)\cdot h + \frac{1}{2}h^T \underbrace{\left[D_iD_jf(a)\right]}_{\substack{\text{Hessian Matrix}\\ \text{of } f \text{ at } a}} h + r(h)$$

Definition 12.1: Hessian Matrix of f

Let $f : (\text{open } U \text{ in } \mathbb{R}^n) \to \mathbb{R}$ such that $\begin{cases} f \text{ is } C^1 \iff \dfrac{\partial f}{\partial x_i} \text{ are not continuous on } U \\ f'' \text{ exists at } a \end{cases}$ So components of f'' are $D_iD_jf(a)$. Hessian of f at a = Square matrix $[D_iD_jf(a)]$

When f is C^2, Hessian matrix is Symmetric Matrix

Definition 12.2: Critical Point

Let f be a C^1 function, Open U in $\mathbb{R}^n \xrightarrow{f} \mathbb{R}$. $a \in U$ is called critical point if $f'(a) = 0 \iff \nabla f(a) = 0$

If f has local maximum at a, then along any line through a the same must be hold, so all directional derivative =0 at a.

Definition 12.3: Non-degenerate Point

If f is C^2 then a critical point a is called non-degenerate if the Hessian, $Hf(a)$ is non-singular i.e. $\det(Hf(a)) \neq 0$

Claim 12.1

Symmetric Matrix A is positive (semi)definite $\iff$ $\forall$ nonzero vector $x \in \mathbb{R}^n$, $x^TAx > 0$ (resp. ≥ 0)

Proof. **If Part:**

$x = \sum\limits_i c_iv_i$. Where v_i is the eigen-basis. Then

$$x^TAx = \left(\sum_i c_iv_i\right)^T A\left(\sum_j c_jv_j\right) = \left(\sum_i c_iv_i\right)^T\left(\sum_j \lambda_jc_jv_j\right) = \sum_i \lambda_ic_i^2 > 0 \qquad [v_i^Tv_j = \delta_{ij}]$$

Only If Part:

Use $x^TAx > 0$ for $x = v_i$ eigenvector < 0, $v_i^TAv_i = v_i\lambda_iv_i = \lambda_i$ $\square$

Note:-

Determinant of positive definite matrix > 0 and Determinant of negative definite matrix has sign $(-1)^n$

Theorem 12.1

Let $f : (\text{open } U \text{ in } \mathbb{R}^n) \to \mathbb{R}$. Suppose f has a local maximum or minimum at a then

1. If $f'(a)$ exists then $f'(a) = 0$ i.e. a is a critical point.
2. Suppose in addition to that $f''(a)$ exists then if f has local maximum at a, then $f''(a) \leq 0$ and if f has local minimum at a, then $f''(a) \geq 0$

Proof. 1. For $n = 1$ let we have local minimum at a. Then for small $|h|$

$$\left.\begin{array}{ll} \frac{f(a+h)-f(a)}{h} \geq 0 & \text{for } h > 0 \\ \frac{f(a+h)-f(a)}{h} \leq 0 & \text{for } h < 0 \end{array}\right\} \text{Thus imply respectively that } f'(a) \text{ must be } \geq 0 \text{ and } \leq 0$$

For $n > 1$ use $n = 1$ in every direction i.e. for function $f|_{a+tv}$ for $t \in$ open interval to conclude $D_vf(a) = 0$ $\forall$ directions. So $f'(a) = 0$

$\square$

② For $n = 1$

$$f''(a) = \lim_{h\to 0} \frac{f'(a+h) - f'(a)}{h} = \lim_{h\to 0} \frac{f'(a+h)}{h}$$

Observation: If f has local maximum at a then for $0 < |h| < \delta$, $f(a+h) \geq f(a)$. So by MVT there is k between 0 and h such that

$$\frac{f(a+h) - f(a)}{h} = f'(a+k)$$

Using the observation $f''(a) = \lim_{h\to 0} \frac{f'(a+k)}{h} \geq 0$

For $n > 1$ applying this to each $f|_{a+tv}$ $\forall$ direction vectors v we get all $D_v^2 f(a) \geq 0$. In terms of Hessian let $v = \sum c_i e_i \implies D_v f = \sum c_i D_i f \implies D^2 f(a) = \sum_{i,j} c_j c_i D_j D_i f(a)$ in a neighborhood of a.

$$D_v^2 f(a) = \begin{bmatrix} c_1 & \cdots & c_n \end{bmatrix} H f(a) \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

$\square$

Theorem 12.2

If $f : (\text{open } U \text{ in } \mathbb{R}^n) \to \mathbb{R}$ is a C^3 function and a is a non-generate critical point of f then

$$\begin{array}{rcl}
f \text{ has a local minimum at } a & \iff & H \text{ is positive definite} \\
 & \iff & \text{All eigenvalues of } H \text{ are positive} \\
f \text{ has a local maximum at } a & \iff & H \text{ is negative definite} \\
 & \iff & \text{All eigenvalues of } H \text{ are negative} \\
f \text{ has saddle-point otherwise} & & H \text{ is indefinite}
\end{array}$$

Proof. **If Part:**

We already proved the if direction in Theorem 12.1

Only If Part:

By Taylor's theorem

$$f(a+x) - f(a) = \cancelto{0}{f'(a)}x + \frac{1}{2}x^T H x + r(x)$$

with as $\|x\| \to 0$, $\frac{r(x)}{\|x\|^2} \to 0$. Let's assume that H is positive definite. So far $x \neq 0$ and $x^T H x > 0$. The function $x \to x^T H x$ is continuous, so on the compact set $\{u \mid \|u\| = 1\}$ it is bounded and achieves its infimum μ. So $\mu > 0$ So

$$\frac{x^T H x}{\|x\|^2} \geq \mu \;\forall\; x \neq 0 \implies \left(\frac{x}{\|x\|}\right)^T H \left(\frac{x}{\|x\|}\right)$$

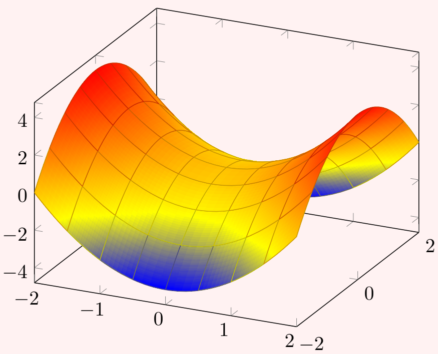
Since $\frac{r(x)}{\|x\|^2} \to 0$ as $\|x\| \to 0$, we can find $\delta > 0$ such that $\frac{|r(x)|}{\|x\|^2} < \frac{\mu}{2}$ when $\|x\| < \delta$. Thus for $\|x\| < \delta$ we have $f(a+x) - f(a) \geq 0$ i.e. f has a local minimum at a $\square$

Definition 12.4: Saddle Point

At a nondegenrate critical point a, H has both

$$\begin{gathered} a \text{ positive eigenvalue, say } \lambda_1 \text{ with eigen vector } u_1 \\ a \text{ negative eigenvalue, say } \lambda_2 \text{ with eigen vector } u_2 \end{gathered}$$

This means $D^2_{u_1} f(a) > 0$, so in the u_1 direction f has local minimum and $D^2_{u_2} f(a) < 0$, so in the u_2 direction f has local maximum



Example 12.1

Many times functions are C^∞ whenever defined so all of the above applies.

- $f(x,y) = c$, constant. All derivatives are zero, H is zero.
- $f(x,y) = ax + by + c$ linear, $(a,b) \neq (0,0)$. No critical points.
- $f(x,y) =$ quadratic.

  General case $(x_1, x_2, \dots, x_n) = x \in \mathbb{R}^n$

$$\begin{aligned} \Phi(x) &= \sum_{i=1}^n a_{ii} x_i^2 + \sum_{1 \le i < j \le n} 2a_{ij} x_i x_j + \sum_{i=1}^n p_i x_i + r \\ &= x^T A x + px + r \\ &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} p_1 & \cdots & p_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + r \quad [\text{where } a_{ij} = a_{ji}] \end{aligned}$$

  Hence $D_i \Phi(x) = \sum_{j=1}^n a_{ij} x_j + p_i$, $D\Phi(x) = 2Ax + p$. Critical points: x such that $2Ax_p = 0$

  If $2A = H$ is nonsingular then there is an unique critical point, namely $x = -H^{-1}p$. Then this point is local minimum is H is positive definite, local maximum id H is negative definite and saddle point otherwise

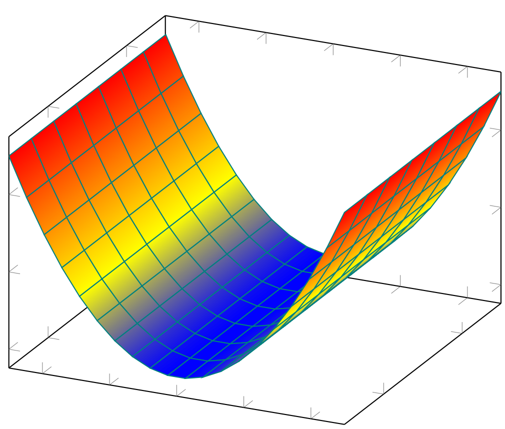
Chapter 13

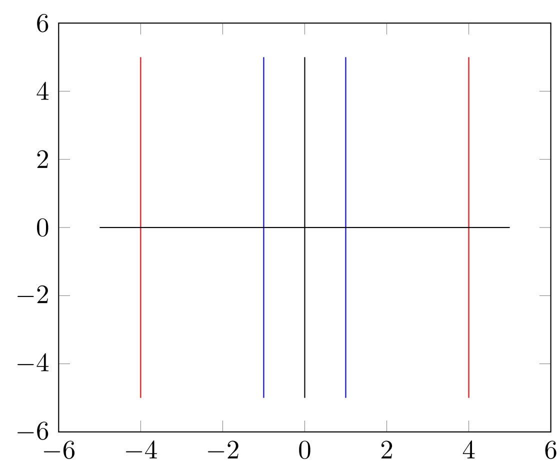
Examples of Functions and Analyze Critical Points

Graph of $\Phi(x) = \Phi(x_1, \ldots, x_n)$ is in $\mathbb{R}^{n+1}$. We can visualize it in $\mathbb{R}^n$ by drawing level sets, namely plot $\Phi(x_1, \ldots, x_n) = c$ for various values of constant c in $\mathbb{R}$

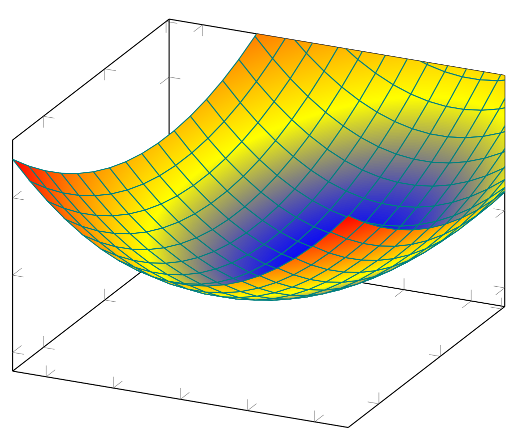
Examples

① $f(x, y) = x^2$

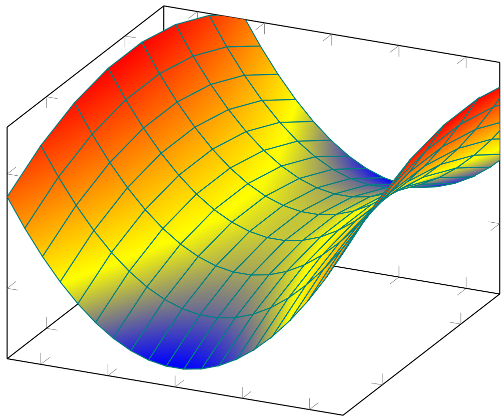


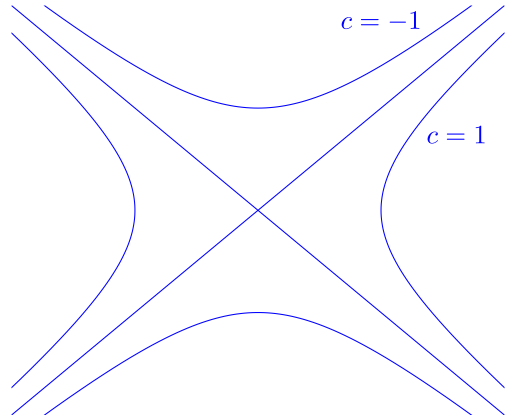


② $f(x, y) = x^2 + y^2$. Level Sets = Circles centered at $(0, 0)$



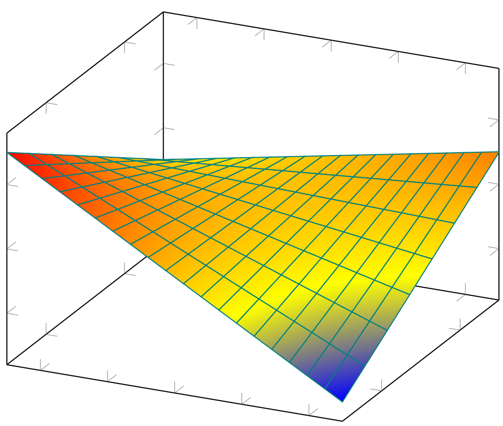
③ $f(x,y) = x^2 - y^2$. Level Sets $c = 0 \implies x = \pm y$, $c = 1 \implies x^2 - y^2 = 1$, $c = -1 \implies x^2 - y^2 = -1$

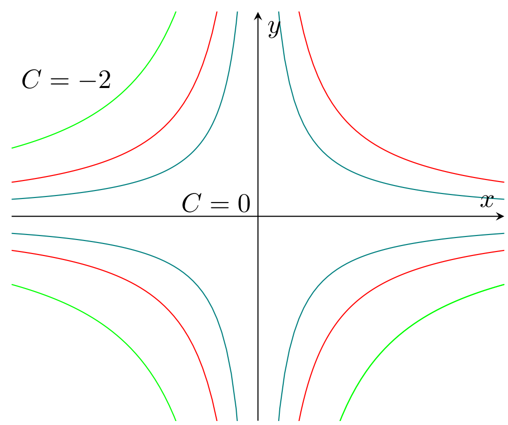




④ $f(x,y) = xy$

$u = \frac{x+y}{\sqrt{2}}, v = \frac{x-y}{\sqrt{2}}$. Then $x = \frac{u+v}{\sqrt{2}}, y = \frac{u-v}{\sqrt{2}}$ and $f(x,y) = \frac{u^2-v^2}{2}$. Here $A = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Hence eigenvectors are $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$





We should understand graphs of 'Quadratic Hypersurfaces' $\Phi(x) = 0$, where $\Phi(x)$ is a quadratic polynomial in n variables.

'Standard Form' is $\lambda_1 x_2^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2 +$ Constant. We will see that by a shift of origin and orthogonal change of coordinates, we can express any general quadratic Φ to the Standard Form

① Getting Rid of Linear Part

$$\begin{aligned} &\lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2 + p_1 x_1 + \cdots + p_n x_n + \text{ constant} \\ =&\lambda_1 (x_1 - a_1)^2 + \cdots + \lambda_n (x_n - a_n)^2 + \text{ another constant} \quad [-2\lambda_i a_i = p_i \implies a_i = -\frac{p_i}{2\lambda_i}, \text{ assuming } \lambda_i \neq 0] \end{aligned}$$

② In general we express x in terms of new basis consisting of orthonormal eigenvectors of A.

Nationalizing a matrix A, $\Gamma^{-1} A \Gamma = D$-diagonal matrix where columns of Γ = eigen basis corresponding to matrix A. Here Γ is orthogonal matrix $\Gamma\Gamma^T = \Gamma^T\Gamma = I$ and we have $\Gamma^T A \Gamma = D \implies A = \Gamma D \Gamma^T$. Now

$$\Phi(x) = x^T A x + pX + r$$

Let x^* = coordinate vector of x in terms of new basis consisting of columns of Γ

$$\begin{aligned} x^* &= \Gamma^{-1}x = \Gamma^T x \text{ we use this to formulate } \Phi \\ &= (x^T\Gamma)D(\Gamma^T x) + p\Gamma(\Gamma^T x) + r = \Phi(x) \\ &= \underbrace{x^{*T}Dx^*}_{\substack{\downarrow \\ \text{standard} \\ \text{form}}} + \underbrace{p\Gamma x^*}_{\substack{\downarrow \\ \text{linear} \\ \text{form}}} + r = \Psi(x^*) \end{aligned}$$

Use step 1 to eliminate the linear term

Now we will look into some more examples.

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

$$A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \text{ and } H = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

H is positive definite because diagonal entries are positive and determinant $= 3 > 0$. So the unique critical point $(0,0)$ is a local minima

> **Note:-**
>
> 2×2 symmetric matrix $\begin{bmatrix} a & c \\ c & b \end{bmatrix}$ is positive definite $\iff \begin{cases} a, b > 0 \\ ab - c^2 > 0 \end{cases}$

2 $\Phi(x) = 2x^2 + 3y^2 - 4xy - 12x - 14y + 21 = \begin{bmatrix} x \\ y \end{bmatrix}^T A \begin{bmatrix} x \\ y \end{bmatrix} + p \begin{bmatrix} x \\ y \end{bmatrix} + r$

$$A = \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix} \text{ and } H = \begin{bmatrix} 4 & -4 \\ -4 & 6 \end{bmatrix} \text{ and } p = \begin{bmatrix} -12 & 14 \end{bmatrix}$$

H is positive definite as diagonal entries are positive and determinant $= 8 > 0$. The critical point is the solution of the equation

$$H \begin{bmatrix} x \\ y \end{bmatrix} = -\begin{bmatrix} -12 \\ 13 \end{bmatrix} \iff \begin{bmatrix} 4 & -4 \\ -4 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = -\begin{bmatrix} -12 \\ 14 \end{bmatrix}$$

Hence $x = 2$, $y = -1$. Therefore minimum value $\Phi(2,-1) = 2$

> **Note:-**
>
> Another way: Complete the squares
>
> $$\begin{aligned} \Phi(x) &= 2(x-2)^2 + 4(y+1)^2 - 4(x-2)(y+1) + 2 \\ &= 2u^2 + 3v^2 - 4uv + 2 \end{aligned}$$

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

$$f'(x,y) = \begin{bmatrix} 3x^2 - 3 & 3y^2 - 3 \end{bmatrix}, \qquad \nabla f = \begin{bmatrix} 3x^2 - 3 \\ 3y^2 - 3 \end{bmatrix}$$

Critical points are (x,y) such that $f'(x,y) = 0$ i.e. $\begin{cases} 3x^2 - 3 = 0 \\ 3y^2 - 3 = 0 \end{cases}$. There are 4 critical points $= (\pm 1, \pm 1)$

$$\text{Hessian } H = \begin{bmatrix} 6x & 0 \\ 0 & 6x \end{bmatrix}$$

$$(1,1,) \to \text{local min}, (-1,-1) \to \text{local max}, (\pm 1, \mp 1) \to \text{saddle points}$$

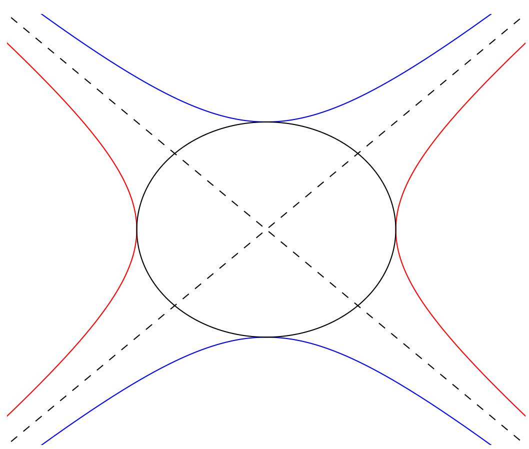
Note:-

For $x^3 - y^2 + 3x - 3y$ there are no critical points

Chapter 14

Constrained Optimizations and Lagrange Multipliers

Example: Optimize $f(x, y)y^2 - x^2$ subject to the constraint $h(x, y) = x^2 + y^2 = 1$



In other words we want to find extrema of $f|_M$ where M is the level curve for h at level 1 i.e. $M = h^{-1}(1)$

Form the way level sets of f interact with M, here we see that we have maximum at $(0, \pm 1)$ and minimum at $(\pm 1, 0)$

It also appears that the constrained graph and the level curve of the objective function f are $\underset{\text{What it means}}{\underset{\downarrow}{\text{tangential}}}$ to each other

We will define Tangent Space to a level set of a C^1 function at a point p on M

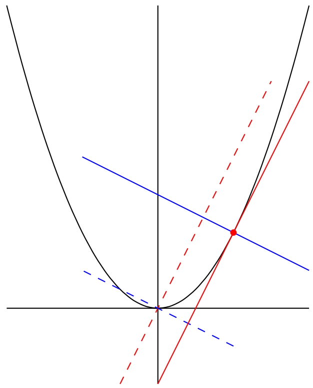
14.1 Tangent Space

Definition 14.1.1: Tangent Space

Tangent Space to a hypersurface $M = f^{-1}(c)$ in $\mathbb{R}^n$ where $f : (\text{Open } U \subset \mathbb{R}^n) \to \mathbb{R}$ is a C^1 function and $c \in \mathbb{R}$ at a point $p \in M$ is a subspace of $\mathbb{R}^n$ defined to be

$$T_pM = \ker(f'(p)) = \{v \in \mathbb{R}^n \mid f'(p)(v) = 0\} = \{v \in \mathbb{R}^n \mid \nabla f(p) \cdot v = 0\}$$

Geometric tangent space considering to our mental image $= T_pM + p =$ Shift T_pM by vector p. Likewise define Normal Space to be the set of vectors orthogonal to T_pM i.e. $T_pM^{\perp}$



Eg. $f(x,y) = y - x^2$, $M = f^{-1}(0)$. $p = (3,9) \in M$.
Here $f'(p) = \begin{bmatrix} -2x & 1 \end{bmatrix}_{(3,9)} = \begin{bmatrix} -6 & 1 \end{bmatrix}$

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto -6x + y$$

$T_pM = \ker(f'(p)) = \{(x,y) \mid y = 6x\} =$ red line.
$N_pM =$ line $y = -\frac{1}{6}x =$ blue line
Geometric tangent space $= T_pM + p$ and geometric normal space $= N_pM + p$

14.2 Lagrange Multiplier

Let U be open in $\mathbb{R}^n$. $f : U \to \mathbb{R}$ objective function and $h : U \to \mathbb{R}$ constraint function. Want to find extrema of f restricted to the level set $M = \{x \in U \mid h(x) = c\} = h^{-1}(c)$ for $c \in \mathbb{R}$

$f|_M$ has local maxima at $p \in M$ means for some $W \subset U$, $f(p) \geq f(x)\ \forall\ x \in W \cap M$

Definition 14.2.1: C^1 Path and Velocity Vector

A C^1 path centered at $p \in U$ in $U \subset \mathbb{R}^n$ is a C^1 map $\gamma : (-\varepsilon, \varepsilon) \to U$ where $0 \mapsto p$. We call $\gamma'(0) =$ velocity of γ at 0

Theorem 14.2.1 Lagrange Multiplier
Let U be open in $\mathbb{R}^n$. $f : U \to \mathbb{R}$, $h : U \to \mathbb{R}$. Let f, h are C^1 functions. Let $M = h^{-1}(c)$. If $h'(p) \neq 0$ and $f|_M$ has a local extrema at $p \in M$ then $\exists! \lambda \in \mathbb{R}$ such that

$$\nabla f(p) = \lambda \nabla h(p)$$

Proof. Consider paths on level set $M = h^{-1}(c)$ i.e.

$$\begin{tikzcd} \gamma : & (-\varepsilon, \varepsilon) \arrow[r] \arrow[dr] & M \arrow[d, phantom, "\cap"] & = h^{-1}(c) \\ & & U \arrow[r, "h"] & \mathbb{R} \end{tikzcd}$$

Then $h = \gamma(t) = c\ \forall\ t \in (-\varepsilon, \varepsilon)$. Hence by Chain Rule

$$h'(p)\gamma'(0) = \nabla h(p) \cdot \gamma'(0) = 0$$

i.e. $\{$velocity vectors of all paths γ on M centered at $p\} \subset T_pM$
Key Fact: When $h'(p) \neq 0$ we have equality! Proof of this fact uses Implicit Function Theorem

Now let's recall the objective function f and recall that p is assured to be a local max/min. If γ is a C^1 curve on M then in particular $f|_{\text{image}(\gamma)}$ also has a max/min at p. Therefore

$$0 = (f \circ \gamma)'(o) = \nabla f(p) \cdot \gamma'(0)$$

i.e. ∇f is orthogonal to velocity vectors to all curves centered at p.

By claim $\nabla f(p) \perp T_pM$, we already say $\nabla h(p) \perp T_pM$. We know $\nabla h(p) \neq 0$ by assumption. Hence $\exists! \lambda$ such that $\nabla f(p) = \lambda \nabla h(p)$ $\square$

14.3 Some Examples for Applications

(i) $f(x,y) = y^2 - h^2$ and $h(x,y) = x^2 + y^2$, $c = 1$. Therefore $M = h^{-1}(1) =$ Unit Circle

Suppose $p = \begin{bmatrix} a \\ b \end{bmatrix}$ is an extremum of $f|_M$

$$\nabla f(p) = \begin{bmatrix} -2x \\ 2y \end{bmatrix}_{(a,b)} = \begin{bmatrix} -2a \\ 2b \end{bmatrix} \qquad \nabla h(p) = \begin{bmatrix} 2x \\ 2y \end{bmatrix}_{(a,b)} = \begin{bmatrix} 2a \\ 2b \end{bmatrix}$$

We know that $\exists! \lambda \in \mathbb{R}$ such that

$$\begin{bmatrix} -2a \\ 2b \end{bmatrix} = \lambda \begin{bmatrix} 2a \\ 2b \end{bmatrix}$$

This is not possible unless one of a, b os 0. Therefore

$$\begin{aligned} a = 0 \quad &\implies \quad b = \pm 1 \text{ and } \lambda = 1 \\ b = 0 \quad &\implies \quad a = \pm 1 \text{ and } \lambda = -1 \end{aligned}$$

(ii) $f(x, y) = y$ is subject to constraint $h(x, y) = y - g(x) = 0$ where $g : \mathbb{R} \to \mathbb{R}$ is some C^1 function. This is equivalent to finding extrema of $y = g(x)$ as in school

Suppose $p = \begin{bmatrix} a \\ b \end{bmatrix}$ gives an extremum

$$\nabla f(p) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \lambda \nabla h(p) = \lambda \begin{bmatrix} -g'(a) \\ 1 \end{bmatrix}$$

i.e. $1 = \lambda \implies 0 = -\lambda g'(a) \implies g'(a) = 0$ as expected

(iii) $f(x, y) = x^2$ subject to $h(x, y) = y = 0$

$$\nabla f = \begin{bmatrix} 2x \\ 0 \end{bmatrix} = \lambda \nabla h = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \implies \lambda = 0, x = 0$$

Note:-

If we instead take $h(x, y) = y^2$, then we get $(, xy) = (0, 0)$ but λ arbitrary

(iv) $f(x, y) = xy$ subject to $h(x, y) = \frac{x^2}{9} + \frac{y^2}{4} = 1$

$$\nabla f = \begin{bmatrix} y \\ x \end{bmatrix} = \lambda \nabla h = \lambda \begin{bmatrix} \frac{2x}{9} \\ \frac{y}{2} \end{bmatrix}$$

Therefore

$$y = \frac{2x}{9}\lambda, \quad x = \frac{y}{2}\lambda, \quad \frac{x^2}{9} + \frac{y^2}{4} = 1$$

$\lambda = \pm 3$. Find extrema. As constraint = ellipse, a compact set, evaluating f as candidates is enough to find max and min.

(v) Find the points on the sphere $x^2 + y^2 + z^2 = 9$ closest/furthest from $(a, b, c) \to$ arbitrary point in $\mathbb{R}^3$

$f(x, y, z) = (x - a)^2 + (y - b)^2 + (z - c)^2$ and $h(x, y, z) = x^2 + y^2 + z^2 = 9$. Complete this and see that geometrically obvious solution emerge

Next we will prove Inverse Function Theorem and Implicit Function Theorem and come back to justify the claim. In fact we will then be able to prove the general version of Lagrange Multiplier Method i.e. with multiple constraints

14.4 Generalized Lagrange Multiplier

Theorem 14.4.1 Generalized Lagrange Multiplier

U open $\subset \mathbb{R}^n = \mathbb{R}^{d+m}$ want to find extrema of objective function $f : U \to \mathbb{R}$ subject to constraint $h = c$

> for a C^1 function: $U \to \mathbb{R}^m$ where $c \in \mathbb{R}^m$ i.e. we want to find extreme of $f|_{M=h^1(c)}$

Key Assumption: $\forall\, x \in M$, $h'(x)$ is surjective i.e. $h'(x) : \mathbb{R}^{d+m} \to \mathbb{R}^m$. (So $\ker(h'(x))$ has $\dim d$. Recall we called $\ker(h'(x)) = T_x M$)
Suppose $f|_M$ has a local extremum at $p \in M$ Then $\exists!$ real numbers $\lambda_1, \lambda_2, \dots, \lambda_m$ such that

$$\nabla f(p) = \lambda_1 \nabla h_1(p) + \cdots + \lambda_n \nabla h_n(p)$$

where $h(p) = \begin{bmatrix} h_1(p) & \cdots & h_m(p) \end{bmatrix}^T \in \mathbb{R}^m$

Proof. we will show that

1. $\nabla f(p) \perp T_p M$
2. Any vector $\perp T_p M$ is a linear combination of $\nabla h_i(p)$

These are the steps.

1. Let $v \in T_p M = \ker(f'(p))$. By HW4 Problem v can be represented by some curve based at p i.e. we can find a C^1 curve $\gamma : (-\varepsilon, \varepsilon) \to M \subset U$ where $0 \mapsto p$ such that $\gamma'(0) = v$.

   As we have an extremum of $f|M$ at p it is also an extreme point for $(-\varepsilon, \varepsilon) \xrightarrow{\gamma} M \xrightarrow{f} \mathbb{R}$. So by 1-Variable Calculus $(f \circ \gamma)'(0) = 0$ i.e. $f'(p)\gamma'(0) = 0$ i.e. $\nabla f(p) \cdot v = 0$

2. For every curve γ as above $h \circ \gamma =$ constant. Therefore $h'(p)\gamma'(0) = 0$ i.e. $\nabla h'(p) \cdot v = 0$

   $$h'(p) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(p) & \cdots & \frac{\partial h_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1}(p) & \cdots & \frac{\partial h_m}{\partial x_n}(p) \end{bmatrix} = \begin{bmatrix} \nabla h_1(p)^T \\ \vdots \\ \nabla h_m(p)^T \end{bmatrix} = \begin{bmatrix} \nabla h_1(p) & \cdots & \nabla h_m(p) \end{bmatrix}^T$$

   So

   $$\nabla h_1(p) \cdot v = 0, \dots, \nabla h_m(p) \cdot v = 0$$

   Therefore $\underbrace{\nabla h_i(p)}_{\substack{m \text{ linearly} \\ \text{independent} \\ \text{vectors}}} \perp \underbrace{T_p M}_{\substack{\dim n-m \\ =d}}$. Everything is in $\mathbb{R}^n = \mathbb{R}^{m+d}$. $\therefore$ $(T_p M)^\perp$ has $\nabla h_1(p), \dots, \nabla h_m(p)$ as a basis. i.e. 2 is proved

□

Chapter 15

Inverse Function Theorem

Definition 15.1: Homeomorphism

A bijective continuous function whose inverse is also continuous is called homeomorphism

Theorem 15.1 Inverse Function Theorem

Suppose U be an open set in $\mathbb{R}^n$. $f : U \to \mathbb{R}^n$ be a C^1 function. $f'(a) = A$ is invertible . Then

1. f is injective in some neighborhood of a

2. There are open sets $V \subset U, a \in V$ and $W \subset \mathbb{R}^n, f(a) \in W$ such that f is a bijection $V \underset{g}{\overset{f}{\rightleftarrows}} W$ whose inverse, g is also continuous i.e. a local homeomorphism i.e. at the given point a there exists a neighborhood at which f is homeomorphism

3. f^{-1} is also differentiable on W i.e. for any $f(u) \in W$

$$Df^{-}(f(u)) = Df(u)^{-1}$$

Note:-

1. **Crucial** that $\dim U$ and target are the same

2. There are appropriate versions of the theorem when $f'(a)$ is injective / surjective / arbitrary (when $f'(a)$ is surjective it is the Implicit Function Theorem) those versions can be proved using the theorem

Proof.

- $n = 1$ is easy. Directly using MVT

- We may assume that $a = 0$ and $f(a) = 0$ (replace f by $f(u+a) - f(a)$) and $f'(a) = $ Identity (replace f by $f'(a)^{-1}f(a)$) check that the result for given f follows easily from result for this normalized f

  Normalization makes formulation / calculation in the proof a bit simpler but may assume a bit the natural main ideas, so we won't normalize.

1. Injectivity of f on a ball B around a of small radius ε. We will choose ε later.

   Best linear approximation for $f(x)$ near a is $f(x) + f'(a)(x - a)$. If f were = this function, the theorem is easy so let's examine the remainder

$$\begin{aligned} r(x) &= f(x) - f(a) - f'(a)(x - a) \\ r'(x) &= f'(x) - f'(a) \end{aligned}$$

   We can make $f'(x) - f'(a)$ small in some ball B around a by continuity of f' at a.

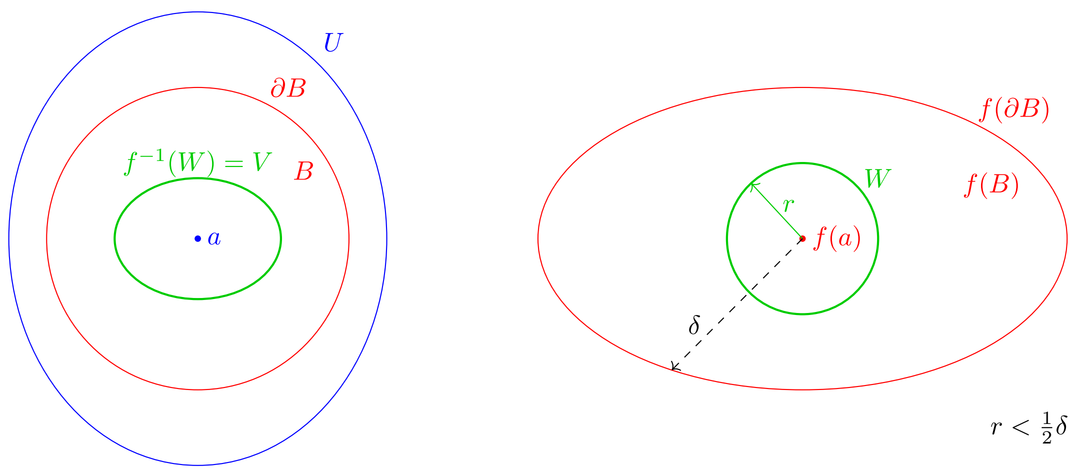
$$r(x_1) - r(x_2) = f(x_1) - f(x_2) - f'(a)(x_1 - x_2)$$

   Choose a good open ball B centered at a with all of the following properties:

(i) Ensure that $\forall\, x \in B\ \|f'(x) - f'(a)\| < \varepsilon$

(ii) $U \xrightarrow{f} L(\mathbb{R}^n) \xrightarrow{\det} \mathbb{R}$ is continuous at a and $\det f'(a) \neq 0$ so can choose B such that $\forall\, x \in B$, $\det f'(x) \neq 0$ and hence $f(x)$ is invertible.

(iii) Shrink B further if necessary to ensure $\overline{B} \subset U$ (useful later to minimize a continuous function on this compact set.)



$\forall\, x \in B$, $f'(x)$ is invertible. $\|r'(x)\| = \|f'(x) - f'(a)\| < \varepsilon$. By MVT applied on $r(x)$ on the convex set B, we get for any $x_1, x_2 \in B$

$$\|f(x_1) - f(x_2) - f'(a)(x_1 - x_2)\| = \|r(x_1) - r(x_2)\| \leq \varepsilon\|x_1 - x_2\|$$

Now recall $\|p - q\| \geq \|p\| - \|q\|$. Hence

$$\|f(x_1) - f(x_2) - f'(a)(x_1 - x_2)\| \geq \|f'(a)(x_1 - x_2)\| - \|f(x_1) - f(x_2)\|$$

Upshot: $\|f(x_1) - f(x_2)\| \geq \|f'(a)(x_1 - x_2)\| - \varepsilon\|x_1 - x_2\| \geq \overset{\substack{\text{Needed}\\ \downarrow}}{\cdots}$

Note:-

At this point if we had normalized $f'(a) =$ Identity then we would have gotten

$$\|f(x_1) - f(x_2)\| \geq (1 - \varepsilon)\|x_1 - x_2\|$$

Taking $\varepsilon < 1$ fives the injectivity of f

In our case we need to find lower bound on $\|f'(a)(x_1 - x_2)\|$. Minimize $\{\|f'(a)u\| \mid \|u\| = 1\}$. $f'(a)$ is continuous and the set of all unit vectors is compact. This set has a minimum, minimum$=m > 0$ as it is invertible so $f'(\text{non zero vector}) \neq 0$.

Now take $\varepsilon < m$ and then in the resulting ball B we have

$$\|f(x_1) - f(x_2)\| \geq (m - \varepsilon)\|x_1 - x_2\| \tag{15.1}$$

This gives the injectivity of f on B. So we have the bijection $B \underset{f^{-1}=g}{\overset{f}{\rightleftarrows}} f(B)$. (15.1) is saying that any $y_1 = f(x_1), y_2 = f(x_2)$ in $f(B)$, i.e. $g(y_1) = x_1, g(y_2) = x_2$

$$\|g(y_1) - g(y_2)\| \leq \frac{1}{m - \varepsilon}\|y_1 - y_2\|$$

i.e. g is uniformly continuous.

② We have bijection of f between B and $f(B)$, V is supposed to be open but we have taken open ball, so its open. Inverse of f (g) is continuous so what left is $f(B)$ open

To show that f is a local Homeomorphism it is enough to find an open ball W around $f(a)$ with $W \subset f(B)$. Then we simply take $V = f^{-1}(W)$ which is open by continuity of f and clearly $V \underset{g}{\overset{f}{\rightleftarrows}} W$ are bijections just restrict f, g from $B, f(B)$ respectively.

How to construct W? What radius to take around W? Stay away from $f(\partial B)$. $\delta = \min\{\|f(x) - f(a)\| \mid x \in \partial B\} > 0$. Choose radius of W to be $\frac{1}{2}\delta$. We will be done if we show $W \subset f(B)$ i.e. given any $c \in W$ $\exists$ $x^* \in B$ such that $f(x^*) = c$ (x^* is necessarily unique, by injectivity).

Idea: Consider the differentiable function $\begin{cases} B \to \mathbb{R}_{\geq 0} \\ x \mapsto \|f(x) - c\|^2 \end{cases}$

Note that $\|c -$ any point on $f(\partial B)\| > r$ by triangle inequality where as $\|c - f(a)\| < r$ (as c is inside $W =$ ball of radius $r < \frac{1}{2}\delta$ around $f(a)$). Hence $\|f(x) - c\|^2$ will take its minimum value at some point say $x^* \in B$. Now $f = (f_1, f_2, \ldots, f_n)$ and $c = (c_1, \ldots, c_n)$

$$\mu(x) = \|f(x) - c\|^2 = \sum_{i=1}^{n} (f_i(x) - c_i)^2$$

Derivative of this function is 0 at x^*

$$B \xrightarrow{\quad f \quad} \mathbb{R}^n \xrightarrow{\qquad \mu \qquad} \mathbb{R}$$

$$x \longmapsto f(x) = y \longmapsto \|y - c\|^2$$

Hence

$$\underbrace{\mu'(f(x^*))}_{\uparrow} \circ f'(x^*) = 0$$

$$\begin{bmatrix} 2(f_1(x^*) - c_1) & \cdots & 2(f_n(x^*) - c_n) \end{bmatrix} \text{ (Matrix of } f'(x^*) \text{ – Invertible)}$$

Therefore $\begin{bmatrix} 2(f_1(x^*) - c_1) & \cdots & 2(f_n(x^*) - c_n) \end{bmatrix}$ must be 0 i.e. $f_i(x^*) = c_i$ i.e. $f(x^*) = c$. So we showed that each $c \in W$ is in the image of f. Now take $V = f^{-1}(W)$ and we have the Homeomorphism.

③ Differentiability of $f^{-1} = g$ at any point $y \in W$.

$$\begin{tikzcd} x \arrow[r, shift left, "f"] \arrow[d, bend right, "\text{add } h"'] & y \arrow[l, shift left, "g"] \arrow[d, bend left, "\text{add } k"] \\ x + h \arrow[r, shift left, "f"] & y + k \in W \arrow[l, shift left, "g"] \end{tikzcd}$$

Take small $k \in \mathbb{R}^n$ and let $h = g(y + k) - g(y)$ and $k = f(x + h) - f(x)$. Each of h and k determines the other uniquely. In particular $h \neq 0 \iff k \neq 0$ (by bijectivity). $h \to 0 \iff k \to 0$ (by continuity of f and g). $\alpha(h) = f(x + h) - f(x) - Th = k - Th$ where $T = f'(x)$. Then we have $\frac{\|\alpha(h)\|}{\|h\|} \to 0$ as $\|h\| \to 0$. We want to show $g'(y) = T^{-1}$

Let $\beta(k) = g(y+k) - g(y) - T^{-1}k = h - T^{-1}k$. We will show that as $k \to 0$, $\frac{\|\beta(k)\|}{\|k\|} \to 0$

$$\begin{aligned}\frac{\|\beta(k)\|}{\|k\|} = \frac{\|h - T^{-1}k\|}{\|k\|} = \frac{\|T^{-1}(Th-k)\|}{\|k\|} &\le \frac{\|T^{-1}\|}{\|k\|}\|Th-k\| \\ &= \frac{\|T^{-1}\|}{\|k\|}\|\alpha(h)\| \\ &= \frac{\|T^{-1}\|}{\|k\|}\frac{\|\alpha(h)\|}{\|h\|}\|h\| \\ &= \|T^{-1}\|\frac{\|h\|}{\|k\|}\frac{\|\alpha(h)\|}{\|h\|}\end{aligned}$$

We know by $\frac{\|h\|}{\|k\|} < \frac{1}{m-\varepsilon}$ by (15.1). $\frac{\|\alpha(h)\|}{\|h\|} \to 0$ as $k \to 0$ because then $h \to 0$.

□

Note:-

- For another proof of surjectivity onto W, see Rudin's use of contraction property
- There is a more general result which assumed only invertibility of $f'(x)$ for $x \in U$ but not continuity of f' everywhere. (See exposition on Terence Tao's Blog: https://terrytao.wordpress.com/tag/inverse-function-theorem/)
- f need not be globally invertible!

  Example = See Problem 17 from Rudin

  $f(x,y) = (e^x \cos y, e^x \sin y)$. Then

  $$f'(x,y) = \begin{bmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{bmatrix} \xrightarrow{\det} (e^x)^2(\cos^2 x + \sin^2 y) = e^{2x} > 0$$

  Thus f is locally invertible everywhere with C^1 inverse.
- f is not globally one-one $f(x,y) = f(x, y+2\pi)$. Do the rest.

Corollary 15.1

If f is a C^1 map from open U in $\mathbb{R}^n$ to $\mathbb{R}^n$ and $f'(x)$ is invertible $\forall\, x \in U$ then

1. f is an open map
2. f is locally invertible with each such inverse a C^1 dunction (because matrix of $(f^{-1})'$ = inverse of matrix of f' and entries of $A^{-1} = \frac{1}{\det A}$ (polynomials in entries of A) in particular $A \to A^{-1}$ is continuous)

Chapter 16

Implicit Function Theorem

Notation: For $n > m$ let $n = m + d$. Write points of $\mathbb{R}^n = \mathbb{R}^d \times \mathbb{R}^m$ as (x, y) where $x \in \mathbb{R}^d$, $y \in \mathbb{R}^m$

Theorem 16.1 Implicit Function Theorem

Let U open in $\mathbb{R}^{d+m}$. $\Phi : U \to \mathbb{R}^m$ is a C^1 map such that $\Phi'(p)$ is surjective (which means columns of the $m \times (d+m)$ matrix of $\Phi'(p)$ span $\mathbb{R}^m$). WLOG suppose the last m columns of $\Phi'(p)$ are linearly independent and hence span $\mathbb{R}^m$ i.e. the $m \times m$ matrix "$\left.\frac{\partial \Phi}{\partial y}\right|_p = \begin{bmatrix} D_{d+1}\Phi(p) & \cdots & D_{d+m}\Phi(p) \end{bmatrix}$" is invertible. Then

1. $\exists$ a neighorhood W of a in $\mathbb{R}^d$ and a unique C^1 map $W \xrightarrow{f} \mathbb{R}^m$ such that $\left.\begin{array}{l} f(a) = b, (x, f(x)) \in U \ \forall\ x \in W \text{ and} \\ \Phi(x, f(x)) = c \ \forall\ x \in W \end{array}\right\}$ i.e. f is an implicit solution to the equation $\Phi(x, y) = c$
2. One can calculate $f'(x)$ by "Implicit Differentiation"

To understand this, first examine two cases:

- When Φ is a linear map given by a matrix A. Here we are solving the equation $A\begin{bmatrix} x \\ y \end{bmatrix} = c$
- $d = m = 1$ i.e. $n = 2$ $\Phi(x, y) = x^2 + y^2 - 1$, solving $\Phi(x, y) = 0 = c$. When $\left.\frac{\partial \Phi}{\partial y}\right|_{p=(a,b)} \neq 0$ we can locally solve for y in terms of x near p.

$$D\Phi = \left.\begin{bmatrix} \frac{\partial \Phi}{\partial x} & \frac{\partial \Phi}{\partial y} \end{bmatrix}\right|_{(a,b)} = \begin{bmatrix} 2a & 2b \end{bmatrix}$$

$2b = 0$ at $(\pm 1, 0)$

Proof. We will choose W later. Define

$$U \xrightarrow{\psi} \mathbb{R}^{d+m}$$

$$(x, y) \longmapsto (x, \Phi(x, y))$$

Note ψ' has the matrix $\begin{bmatrix} I & O \\ \frac{\partial \Phi}{\partial x} & \frac{\partial \Phi}{\partial y} \end{bmatrix}$. This is nonsingular in a neighborhood of p. So by Inverse Function Theorem ψ is invertible with C^1 inverse in a neighborhood V of p

$$\begin{array}{rcl} V & \longleftrightarrow & \psi(V) \\ (a, b) & \longmapsto & (a, c) \\ (x, y) & \longmapsto & (x, \Phi(x, y)) \\ (u, \alpha(u, v)) & \longleftarrow & (u, v) \end{array}$$

Definition of $\alpha(u,v)$ defined on $\psi(V)$. This tells us $\alpha(a,c)=b$. Whenever $\Phi(x,y)=c$ i.e.

$$(x,y) \xrightarrow{\Phi} (x,c) \xrightarrow{\psi^{-1}} (x,\alpha(x,c)) = (x,y)$$

i.e. $y=\alpha(x,c)$ and $\Phi(x,\alpha(x,c))=c$

So we are forced to define $f(x)=\alpha(x,c)$. But what should be the domain of this function f i.e. what should we take W to be.

$$(a,c) \in \psi(V) \text{ is open} \supset \left(\begin{array}{c} \text{open ball } W \\ \text{around } a \text{ in } \mathbb{R}^d \end{array} \right) \times \{c\}$$

Now for any $x \in W$ we know $(x,c) \in \psi(V)$ i.e. $(x,\alpha(x,c)) \in V$ so we define $f : W \to \mathbb{R}^m$ where $f(x)=\alpha(x,c)$ and we have derived the function. Now ϕ^{-1} is C^1 and α is component of ϕ^{-1} so all components of ϕ^{-1} is also C^1. hence f is C^1

Uniqueness of f is not true in general for arbitrary W. $\Phi(x,y)=x^2+y^2,\ c=1$. In $W=W_1 \sqcup W_2$

$$f(x) = \begin{cases} \sqrt{1-x^2} & x \in W_1 \qquad \text{[is forced]} \\ \sqrt{1-x^2} \text{ or } -\sqrt{1-x^2} & x \in W_2 \end{cases}$$

. We have choice for f on W_2.

If W is connected, f will be unique. Eg. take W to be a ball. Suppose g is another solution to $\Phi(x,y)=c$ i.e. $\Phi(x,g(x))=c$ for $x \in W$ and $g(a)=b$. Then consider the set $S=\{x \in W \mid f(x)=g(x)\}$. Show that this set is both closed (easy $S=(f-g)^{-1}(0)$) and open.

Calculate derivative of f using the fact that $\psi \circ \psi^{-1} =$ Identity and Chain Rule. □

Example 16.0.1 (Application of Implicit Function Theorem)

(i) Linear map $\Phi : \mathbb{R}^{d+m} \to \mathbb{R}^m$ given by matrix A. Given $A\begin{bmatrix} a \\ b \end{bmatrix} = c$. Want to solve $A\begin{bmatrix} x \\ y \end{bmatrix} = c$

$A=[P \mid Q]$ where P is $m \times d$ and Q is $m \times m$ and Q is invertible. i.e.

$$[P \mid Q]\begin{bmatrix} x \\ y \end{bmatrix} = c \iff [Q^{-1}P \mid I]\begin{bmatrix} x \\ y \end{bmatrix} = Q^{-1}c \iff Q^{-1}Px + y = Q^{-1}c$$
$$\iff y = Q^{-1}c - Q^{-1}Px$$

(ii) We can solve for y in terms of x near any (a,b) on the unit circle when $\left.\frac{\partial \Phi}{\partial y}\right|_{(a,b)} \neq 0$. [This is mate when $b \neq 0$ i.e. at all points except $(\pm 1, 0)$].

$$D\Phi|_{(a,b)} = \begin{bmatrix} 2a & 2b \end{bmatrix}$$

We can see directly

$$\left. \begin{array}{ll} \text{when } b>0 & y=\sqrt{1-x^2} \\ \text{when } b<0 & y=-\sqrt{1-x^2} \end{array} \right\} \text{near (a,b) in fact } \forall\ x \in (-1,1)$$

Similarly we can solve for x in terms of y when $\left.\frac{\partial \Phi}{\partial x}\right|_{(a,b)} = 2a \neq 0$ This is true when $a \neq 0$

Remark: Implicit Function Theorem gives a sufficient condition to be able to locally solve a system of linear equations

$$\left.\begin{aligned} \Phi_1(x_1,\dots,x_d,y_1,\dots,y_m) &= c_1 \\ \Phi_2(x_1,\dots,x_d,y_1,\dots,y_m) &= c_1 \\ \vdots \qquad \vdots \qquad &\vdots \\ \Phi_m(x_1,\dots,x_d,y_1,\dots,y_m) &= c_1 \end{aligned}\right\} \begin{array}{l}\text{for } y_i\text{'s in terms of } x_i\text{'s} \\ \text{locally near a given solution} \\ y=b \text{ and } x=a\end{array}$$

Note:-

The condition of invertibility of submatrix of Φ is not necessary. Eg. $\Phi(x,y) = y - x^3$ near $(0,0)$

$$D\Phi|_{(0,0)} = \begin{bmatrix}-3x^2 & 1\end{bmatrix}\Big|_{(0,0)} = \begin{bmatrix}0,1\end{bmatrix} \qquad \left.\frac{\partial\Phi}{\partial x}\right|(0,0) = 0$$

but still we can solve for x in terms of y: $x = \sqrt[3]{y}$

Chapter 17

Complex Differentiation

Suppose U open in $\mathbb{C} = \mathbb{R}^2$, $U \xrightarrow{f} \mathbb{C}$ a differentiable map i.e. Df as an $\mathbb{R}$ linear operator $\mathbb{R}^2 \to \mathbb{R}^2$ is defined.

Note:-

There is one thing that makes $\mathbb{C}$ differ from $\mathbb{R}^2$ i.e. $\mathbb{C}$ forms a field.

Definition 17.1: Complex Differentiation

$f : U \to \mathbb{C}$ is called complex differentiable at $z_0 \in U$ if

$$\lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h} \qquad \text{exists}$$

Thus the limit equals to $f'(z_0)$. h is a complex number and this is a division in the field $\mathbb{C}$. '$h \to 0$' means $\|h\| \to 0$

Definition 17.2: Holomorphic Function

We say that f is holomorphic on U if $f'(z)$ exists $\forall\ z \in U$. Then $f' : U \to \mathbb{C}$ is the derivative of f where $z \mapsto f'(z)$

Theorem 17.1

f is holomorphic $\iff$ so is f'

Cauchy Riemann Conditions

Suppose $f'(z_0)$ exists where $z_0 = a + ib \in$ open $U \subset \mathbb{C}$. $f(z) = u + iv$ where $u : U \to \mathbb{R}$ and $v : U \to \mathbb{R}$. First take $h = t \in \mathbb{R}$.

$$\begin{aligned} f'(z_0) &= \lim_{t \to 0} \frac{f(a + t + ib) - f(a + ib)}{t} \\ &= \lim_{t \to 0} \frac{u(a + t, b) - u(a, b)}{t} + i \lim_{t \to 0} \frac{v(a + t, b) - v(a, b)}{t} \\ &= \left.\frac{\partial u}{\partial x}\right|_{z_0} + i \left.\frac{\partial v}{\partial x}\right|_{z_0} \end{aligned} \tag{17.1}$$

Now take $h = it$, $t \in \mathbb{R}$

$$\begin{aligned}
f'(z_0) &= \lim_{t\to 0} \frac{f(a+ib+it) - f(a+ib)}{it} \\
&= \lim_{t\to 0} \frac{u(a,b+t) - u(a,b)}{it} + i \lim_{t\to 0} \frac{v(a,b+t) - v(a,b)}{it} \\
&= \frac{1}{i} \left[\left.\frac{\partial u}{\partial y}\right|_{z_0} + i \left.\frac{\partial v}{\partial y}\right|_{z_0} \right] \\
&= \left.\frac{\partial v}{\partial y}\right|_{z_0} - i \left.\frac{\partial u}{\partial y}\right|_{z_0} \qquad (17.2)
\end{aligned}$$

Equating (17.1) and (17.2) f is complex differentiable at z_0 and

$$\boxed{\left.\frac{\partial u}{\partial x}\right|_{z_0} = \left.\frac{\partial v}{\partial y}\right|_{z_0} \qquad \left.\frac{\partial v}{\partial x}\right|_{z_0} = -\left.\frac{\partial u}{\partial y}\right|_{z_0}}$$

In fact more is true.

Claim 17.1

For open $U \subset \mathbb{C}$ if $U \xrightarrow{f} \mathbb{C}$ is complex differentiable at $z_0 \in U$ then f is also real differentiable as a function $U \to \mathbb{R}^2$ and its Jacobin is (letting $f = (u,v)$, $u, v : U \to \mathbb{R}$)

$$Jf(z_0) = \left.\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}\right|_{z_0} = \left.\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ -\frac{\partial u}{\partial y} & \frac{\partial u}{\partial x} \end{bmatrix}\right|_{z_0} = \left.\begin{bmatrix} \frac{\partial v}{\partial y} & -\frac{\partial v}{\partial x} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}\right|_{z_0}$$