Contents

Unit 1 Topology and Real Analysis

Week 1 Normed Vector Spaces

1.1 Normed Vector Spaces

Definition. Normed Vector Space (NVS): A normed vector space is a vector space V over $\mathbb{R}$ equipped with a function $\| \ \| : V \to [0, \infty)$, called the norm on V, such that

1. $\|v\| = 0 \iff v = 0$.
2. For all $\alpha \in \mathbb{R}$ and $v \in V$, $\|\alpha v\| = |\alpha| \cdot \|v\|$.
3. (Triangle Inequality) For all $u, v \in V$, $\|u + v\| \leq \|u\| + \|v\|$.

Note: We often denote a normed vector space by the pair $(V, \| \cdot \|)$, where the first element is the vector space and the second in the norm function. Geometrically,

- $\|v\|$ denotes "the length of v" or "the distance between v and 0".
- $\|v - w\|$ denotes "the distance between v and w".

Note. Motivation of NVS: The field of real analysis is the study of objects relating to the real numbers, e.g. $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{R}^{\mathbb{R}}$, etc. However, it is often preferable to use "nicer" elements such as $\mathbb{Q}$. NVSs allow us to measure distance, therefore to measure error, therefore to make approximations.

Example: $(\mathbb{R}, |\cdot|)$ is an NVS, so is $(\mathbb{R}, \|\cdot\|)$, $\|a\| = 3|a|$.

Definition. p Norm: Let $V = \mathbb{R}^n$. For $p \in \mathbb{Z}_{\geq 1}$ and $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$,

$$\|v\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{1/p}$$

is a norm on $\mathbb{R}^n$, therefore $(\mathbb{R}^n, \|\cdot\|_p)$ is an NVS.

Note. Euclidean Norm: A p norm with $p = 2$, i.e. the square root of the sum of the squares of components, is called the Euclidean norm, and is the usual measure of distance in $\mathbb{R}^n$. Unless stated otherwise, always assume that $\mathbb{R}^n$ is equipped with $\|\cdot\|_2$.

Example. Infinity Norm: When $p = \infty$, we have $\|v\|_\infty = \sup\{|v_1|, \ldots, |v_n|\}$. This is a norm on $\mathbb{R}^n$.

Example. $\mathbb{R}^{\mathbb{N}}$: Note $\mathbb{R}^{\mathbb{N}} = \{(a_n)_{n=1}^{\infty} : a_i \in \mathbb{R}\}$ is the vector space of real sequences. Let $v = (v_1, v_2, \ldots) \in V$ be a sequence, define the p norm

$$\|v\|_p = \left(\sum_{i=1}^{\infty} |v_i|^p \right)^{1/p}$$

and

$$\|v\|_\infty = \sup\{|v_1|, |v_2|, \dots\}$$

While these are not norms, e.g. when the sequence diverges (note the norm must be a non-negative real number), we can find subspaces for which these are norms. E.g.

- $\ell^p = \{v \in \mathbb{R}^\mathbb{N} : \|v\|_p < \infty\}$ then $(\ell^p, \|\cdot\|_p)$ is an NVS.
- $\ell^\infty = \{v \in \mathbb{R}^\mathbb{N} : \|v\|_\infty < \infty\}$ then $(\ell^\infty, \|\cdot\|_\infty)$ is an NVS.

Example. $C([a,b])$**:** Note $C([a,b]) = \{f : [a,b] \to \mathbb{R} : f \text{ is continuous}\}$ is the set of *continuous*, real functions defined on $[a,b]$. Again we define the p norm:

$$f \in C([a,b]), \|f\|_p = \left(\int_a^b |f(x)|^p dx\right)^{1/p}$$

and

$$f \in C([a,b]), \quad \|f\|_\infty = \sup\{|f(x)| : x \in [a,b]\} \overset{\text{EVT}}{=} \max\{|f(x)| : x \in [a,b]\}$$

These each turn $C([a,b])$ into a NVS. The latter norm is called the uniform norm and, unless otherwise stated, we equip $C([a,b])$ with the uniform norm $\|\cdot\|_\infty$.

1.2 Convergence

Notation: By a sequence in V we always means $(a_n)_{n=1}^\infty = (a_1, a_2, \dots)$ where each $a_i \in V$. We will abusively use the shorthand to denote this by $(a_n) \subseteq V$.

Definition. Convergence: Let V be a NVS and $(a_n) \subseteq V$. We say (a_n) converges to $a \in V$, written $a_n \to a$, if for all $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that $\|a_n - a\| < \epsilon$ for all $n \geq N$.

Definition. Divergence: Let V be a NVS and $(a_n) \subseteq V$. We say (a_n) diverges, if for all $a \in V$, $a_n \not\to a$ ((a_n) does not converge to a).

Example: Let $(a_n) \subseteq \ell^\infty$ with $a_n = (1, \frac{1}{2}, \dots, \frac{1}{n}, 0, 0, \dots)$ and $a = (1, \frac{1}{2}, \frac{1}{3}, \dots)$. Then $a_n \to a$. Let $(b_n) \subseteq \ell^\infty$ with $b_n = (\underbrace{1, 1, \dots, 1}_{n}, 0, 0, \dots)$ and $b = (1, 1, 1, \dots)$. Then $b_n \not\to b$.

Definition. Boundedness: Let V be a NVS. $(a_n) \subseteq V$, $A \subseteq V$.

- We say A is bounded if there is an $M > 0$ such that $\|a\| \leq M$ for all $a \in A$.
- We say (a_n) is bounded if $\{a_1, a_2, a_3, \dots\}$ is bounded. I.e. $\exists M > 0, \forall n \in \mathbb{Z}_{\geq 1}, \|a_n\| \leq M$)

Proposition: Let V be an NVS, $(a_n) \subseteq V$. If (a_n) is convergent then (a_n) is bounded.

Proof. Suppose $a_n \to a \in V$. Let N be such that $\|a_n - a\| < 1$ for all $n \geq N$. So for $n \geq N$

$$\|a_n\| = \|a_n - a + a\| \overset{\text{T.I.}}{=} \|a_n - a\| + \|a\| < 1 + \|a\| \in V$$

Taking $M = \max\{\|a_1\|, \dots, \|a_{n-1}\|, 1 + \|a\|\}$, we have $\|a_n\| \leq M$ for all $n \in \mathbb{N}$. □

Proposition. Limit Laws: Let V be an NVS and $(a_n), (b_n) \subseteq V$ with $a_n \to a$ and $b_n \to b$. Then

- $a_n + b_n \to a + b$.
- For $\alpha \in \mathbb{R}$, $\alpha a_n \to \alpha a$.

1.3 Completeness

Definition. Cauchy Sequence: Let V be an NVS and $(a_n) \subseteq V$. (a_n) is Cauchy if for all $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that $\|a_n - a_m\| < \epsilon$ for all $n, m \geq N$.

Proposition: Let V be an NVS. If $(a_n) \subseteq V$ is convergent then (a_n) is Cauchy.

Proof. Let $\epsilon > 0$. There is $N \in \mathbb{N}$ and $a \in V$ such that $\|a_n - a\| < \frac{\epsilon}{2}$ for all $N \geq N$. So for all $n, m \geq N$,

$$\|a_n - a_m\| \leq \|a_n - a\| + \|a_m - a\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

□

Example: Define the subspace of ℓ^∞, $C_{00} = \{(x_n) \in \ell^\infty : \exists N, \forall n \geq N, x_n = 0\}$ to be the vector space of all convergent sequences with 0 tails. Note that $(C_{00}, \|\cdot\|_\infty)$ is a NVS. Define $(a_n) \subseteq C_{00}$ as $a_n = (1, \frac{1}{2}, \ldots, \frac{1}{n}, 0, 0, \ldots)$ and $a = (1, \frac{1}{2}, \frac{1}{3}, \ldots) \notin C_{00}$.

We know $a_n \to a$ in ℓ^∞, therefore $(a_n) \subseteq \ell^\infty$ is Cauchy, however, this implies $(a_n) \subseteq C_{00}$ is Cauchy. We know limits are unique, so since C_{00} is a subspace of ℓ^∞ and $a \notin C_{00}$, $(a_n) \subseteq C_{00}$ diverges. Therefore, convergence implies Cauchy, but the converse is not true.

Definition. Completeness: Let V be a NVS. We say $A \subseteq V$ is complete if whenever $(a_n) \subseteq A$ is Cauchy, then for some $a \in A$, $a_n \to A$.

Definition. Banach Space: Let V be a NVS. We call V a Banach space if V is complete. E.g. $(\mathbb{R}, |\cdot|)$ is a Banach space.

Proposition: $(\mathbb{R}^n, \|\cdot\|_\infty)$ is a Banach space.

Proof. Suppose $(a_k) \subseteq \mathbb{R}^n$ is Cauchy. Suppose $a_k = (a_k^{(1)}, a_k^{(2)}, \ldots, a_k^{(n)})$ where the superscript marks the component of the tuple. Let $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that $\|a_k - a_\ell\|_\infty < \epsilon$ for all $k, \ell \geq N$. Therefore, for each $1 \leq i \leq n$,

$$|a_k^{(i)} - a_\ell^{(i)}| \leq \|a_k - a_\ell\|_\infty < \epsilon.$$

i.e. $(a_k^{(i)})_{k=1}^\infty \subseteq \mathbb{R}$ is a Cauchy sequence. Now $\mathbb{R}$ is known to be complete, hence suppose for all $1 \leq i \leq n$ $a_k^{(i)} \to b_i$.

Now let $b = (b_1, \ldots, b_n)$. Let (a different) $\epsilon > 0$. For each $1 \leq i \leq n$, there exists $N_i \in \mathbb{N}$ such that for all $k \geq N_i$, $|a_k^{(i)} - b_i| < \epsilon$. Now let $N = \max\{N_1, N_2, \ldots, N_n\}$, for all $k \geq N$,

$$\|a_k - b\|_\infty = \max\{|a_k^{(i)} - b_i| : 1 \leq i \leq n\} < \epsilon$$

Hence, (a_k) must converge, therefore $(\mathbb{R}^n, \|\cdot\|_\infty)$ is a Banach space. □

Remark: Let $v = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$ and $1 \le p < \infty$.

1. $\|v\|_p^p = |v_1|^p + \cdots + |v_n|^p \le n\|v\|_\infty^p$.
2. $\|v\|_\infty \le |v_1|^p + \cdots + |v_n|^p = \|v\|_p^p$.
3. $\|v\|_p \le \sqrt[p]{n}\|v\|_\infty$ and $\|v\|_\infty \le \|v\|_p$.

Proposition: $(\mathbb{R}^n, \|\cdot\|_p)$ is a Banach space for $1 \le p < \infty$.

Proof. Suppose $(a_k) \subseteq \mathbb{R}^n$ is Cauchy (with respect to $\|\cdot\|_p$). Therefore, there is an $N \in \mathbb{N}$ such that for all $k, \ell \ge N$,

$$\|a_k - a_\ell\|_\infty \le \|a_k - a_\ell\|_p < \epsilon$$

Therefore (a_k) is Cauchy with respect to $\|\cdot\|_\infty$. This means (a_k) converges to some $a \in \mathbb{R}^n$ with respect to $\|\cdot\|_\infty$ since $(\mathbb{R}^n, \|\cdot\|_\infty)$ is a Banach space. Now let $\epsilon > 0$ and pick an K such that if $k \ge K$, then $\|a_k - a\|_\infty < \frac{\epsilon}{\sqrt[p]{n}}$, so for $k \ge K$

$$\|a_k - a\|_p \le \sqrt[p]{n}\|a_k - a\|_\infty < \epsilon$$

and therefore $(\mathbb{R}^n, \|\cdot\|_p)$ is a Banach space. □

Proposition: ℓ^∞ is a Banach space.

Proof. Recall ℓ^∞ is the space of bounded sequences. Let $(a_n) \subseteq \ell^\infty$ be Cauchy. We write this sequence as $a_n = (a_n^{(1)}, a_n^{(2)}, a_n^{(3)}, \dots)$, where $a_n^{(i)} \in \mathbb{R}$ is the ith element of the nth sequence.

Let $\epsilon > 0$. There is an $N \in \mathbb{N}$ such that $\|a_n - a_m\|_\infty < \epsilon$ for all $n, m \ge N$. For each fixed i,

$$|a_n^{(i)} - a_m^{(i)}| \le \sup\{a_n^{(i)} - a_m^{(i)} : i \in \mathbb{N}\} = \|a_n - a_m\|_\infty < \epsilon$$

Hence each sequence $(a_n^{(i)})$ (ranging on n) is Cauchy. By the completeness of $\mathbb{R}$, there is a $b_i \in \mathbb{R}$ such that $a_n^{(i)} \to b_i$ (as $n \to \infty$) for each $i \in \mathbb{N}$.

Let $b = (b_1, b_2, b_3, \dots) \in \ell^\infty$. Let $\epsilon > 0$. Now for some $N \in \mathbb{N}$,

$$n, m \ge N \implies |a_n^{(i)} - a_m^{(i)}| \le \|a_n - a_m\|_\infty < \epsilon$$

for all $i \in \mathbb{N}$. Now taking $m \to \infty$, we have $|a_n^{(i)} - b_i| \le \epsilon$ for all $i \in \mathbb{N}$, and therefore $\|a_n - b\|_\infty \le \epsilon$ for all $n \ge N$ □

Definition. Subsequence: A subsequence of (a_n) is a sequence $(a_{n_k})_{k=1}^\infty$ such that $n_1 < n_2 < n_3 < \dots$.

Definition. Strongly Cauchy: A sequence $(a_n) \subseteq V$ is said to be strongly-Cauchy if there exists a convergent series $\sum_{n=1}^{\infty} \varepsilon_n$ of positive real numbers such that $||a_{n+1} - a_n|| \le \varepsilon_n$ for all $n \in \mathbb{N}$.

Week 2 Topology

2.1 Closed and Open Sets

Definition. Closed Set: Let V be a NVS. A subset $C \subseteq V$ is said to be closed if whenever $(a_n) \subseteq C$ is such that whenever $a_n \to a \in V$, then $a \in C$.

Definition. Open Set: Let V be a NVS. A subset $U \subseteq V$ is said to be open if $V \setminus U$ is closed.

Definition. Topology: The collection $\{U \subseteq V : U \text{ is open}\}$ is called the topology on V. The study of open and closed sets on a space is called topology.

Example: $\emptyset, V \subseteq V$ are both closed sets ($\emptyset$ is vacuously true), hence taking complements they're both open. $[0,1) \subseteq \mathbb{R}$ is neither open nor closed since a sequence could converge to 1 and it is not open since a sequence could converge to 0.

Definition. Ball: For $r > 0$ and $a \in V$ where V is a NVS. The closed ball of radius r centred at a is $\overline{B}_r(a) = \{x \in V : \|a - x\| \leq r\}$, this is a closed set (in particular, for $V = \mathbb{R}$ $\overline{B}_r(a) = [a-r, a+r]$). The open ball of radius r centred at a is $B_r(a) = \{x \in V : \|a-x\| < r\}$, this is an open set.

Proof. Proof $\overline{B}_r(a)$ is a closed ball. Let $(a_n) \subseteq \overline{B}_r(a)$ such that $a_n \to b \in V$. By definition, $\|a_n - a\| \leq r$ for all $n \in \mathbb{N}$. Since $a_n \to b$, $\|a_n - a\| \to \|b - a\|$. Since $\|a_n - a\| \leq r$ and limits preserve order (inequalities), $\|b - a\| \leq r$, hence $b \in \overline{B}_r(a)$ and so $\overline{B}_r(a)$ is closed. □

Remark: This leads to a similar closed set: $\{x \in V : \|x - a\| \geq r\}$ is closed. Note this $B_r(a)$ is open.

Example: Where $V = \ell^\infty$ is the NVS, $C_0 = \{(x_n) \in \ell^\infty : x_n \to 0\}$ is closed.

Proof. Let $(a_n) \subseteq C_0$ be such that $a_n \to a \in \ell^\infty$. Let $a_n = (a_n^{(1)}, a_n^{(2)}, \dots)$ for all $n \in \mathbb{N}$. Therefore, for all $n \in \mathbb{N}$, $\lim\limits_{k\to\infty} a_n^{(k)} = 0$.

Suppose $a = (b_1, b_2, \dots)$. Let $\epsilon > 0$. We can pick $N_1, N_2 \in \mathbb{N}$ such that, for all $n \geq N_1$, $\|a_n - a\|_\infty < \frac{\epsilon}{2}$ and for all $k \geq N_2$, $|a_{N_1}^{(k)}| < \frac{\epsilon}{2}$. Now for $k \geq N_2$,

$$\begin{aligned}
|b_k| &= |a_{N_1}^{(k)} - b_k - a_{N_1}^{(k)}| \\
&\leq |a_{N_1}^{(k)} - b_k| + |a_{N_1}^{(k)}| \\
&\leq \|a_{N_1} - a\|_\infty + |a_{N_1}^{(k)}| && \text{Since } \|\cdot\|_\infty \text{ is supremum over } k \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon
\end{aligned}$$

Hence, $b_k \to 0$ and so $a = (b_1, b_2, \dots) \in C_0$, thereby making C_0 closed. □

Proposition. Standalone Definition of Openness: The following are equivalent for a NVS V and $U \subseteq V$:

- U is open.
- For all $a \in U$, there exists an $r > 0$ such that $B_r(a) \subseteq U$.

The logic follows that if U is closed at some point then taking a at that point, any positive r contains points excluded in U.

Proof. ($\implies$) Assume U is open, hence $V \setminus U$ is closed. For the sake of contradiction, suppose $\exists a \in U, \nexists r > 0, B_r(a) \subseteq U$. For all $n \in \mathbb{N}$, we can pick $a_n \in B_{\frac{1}{n}}(a)$ with $a_n \notin U$. Note that $\|a_n - a\| < \frac{1}{n}$, hence $a_n \to a$. However $(a_n) \subseteq V \setminus U$ which is closed, hence $a \in V \setminus U$, this is a contradiction.

($\impliedby$) Assume for all $a \in U$, there is an $r > 0$ such that $B_r(a) \subseteq U$. Let $(a_n) \subseteq V \setminus U$ with $a_n \to a \in V$. For the sake of contradiction, suppose $a \in U$. Hence there is an $r > 0$ such that $B_r(a) \subseteq U$. Since $a_n \to a$, there is an $N \in \mathbb{N}$ such that $\|a_N - a\| < r$. Therefore, $a_N \in B_r(a) \subseteq U$ and so $a_N \in U$, a contradiction. □

2.2 Closure and Interior

Proposition: Where V is a NVS, the following are true.

1. If $\{U_\alpha\}_{\alpha \in I}$ is a collection of open sets in V, then $U = \bigcup_{\alpha \in I} U_\alpha$ is open. This is the union of any finite or infinite number of open sets.
2. If $\{C_\alpha\}_{\alpha \in I}$ is a collection of closed sets in V, then $C = \bigcap_{\alpha \in I} C_\alpha$ is closed. This is the intersection of any finite or infinite number of closed sets.
3. If $U_1, \ldots, U_n \subseteq V$ are open, then $U = U_1 \cap \cdots \cap U_n$ is open. This is a finite intersection.
4. If $C_1, \ldots, C_n \subseteq V$ are closed, then $U = U_1 \cup \cdots \cup U_n$ is closed. This is a finite union.

Proof. (1) Let $a \in \bigcup_{\alpha \in I} U_\alpha$. So there exists an $\alpha \in I$ such that $a \in U_\alpha$, however since U_α is open, there is an $r > 0$ such that $B_r(a) \subseteq U_\alpha \subseteq \bigcup_{\alpha \in I} U_\alpha$. Hence $\bigcup_{\alpha \in I} U_\alpha$ is an open set by the equivalent definition of openness.

(2) Notice $\bigcap_{\alpha \in I} C_\alpha$ is closed if and only if $V \setminus \bigcap_{\alpha \in I} C_\alpha = \bigcup_{\alpha \in I} (V \setminus C_\alpha)$ is open (by De Morgan's law). We know by (1) that this is set is open since each of $V \setminus C_\alpha$ is open, hence $\bigcap_{\alpha \in I} C_\alpha$ is closed.

(3) Let $a \in U_1 \cap \cdots \cap U_n$. For all $1 \leq i \leq n$, there is an $r_i > 0$ such that $B_{r_i}(a) \subseteq U_i$. Hence pick $r = \min\{r_1, \ldots, r_n\}$, hence $B_r(a) \subseteq B_{r_1}(a), \ldots, B_{r_n}(a)$ and hence $B_r(a) \subseteq U_1 \cap \cdots \cap U_n$. By the equivalent definition for openness, $U_1 \cap \cdots \cap U_n$ is open.

(4) Notice $C_1 \cup \cdots \cup C_n$ is closed if and only if $V \setminus (C_1 \cup \cdots \cup C_n) = (V \setminus C_1) \cap \cdots \cap (V \setminus C_n)$ is open (De Morgans's law). We know by (3) this is true given each set in the intersection is open. □

Note: Notice the fact that the number of sets is finite in (3) and (4) is important as, for instance, $\bigcap_{n=1}^{\infty}(-\frac{1}{n}, \frac{1}{n}) = \{0\}$ is an intersection of infinitely many open sets, yielding a not open (in fact closed) set. Similarly, $\bigcup_{n=1}^{\infty}[0, 1-\frac{1}{n}] = [0,1)$ is a union of closed sets yielding a not open set.

Definition. Closure: The closure of a set $A \subseteq V$ where V is a NVS is defined to be $\overline{A} := \bigcap_{A\subseteq C, C \text{ is closed}} C$.

Definition. Interior: The interior of a set $A \subseteq V$ where V is a NVS is defined to be $\text{Int}(A) := \bigcup_{U\subseteq A, U \text{ is open}} U$.

Remark: The idea is that $\overline{A}$ is the smallest closed set containing A and $\text{Int}(A)$ is the largest open set contained in A.

Definition. Limit Point: If $A \subseteq V$ where V is a NVS, then a limit point of A is $a \in V$ such that there is a sequence $(a_n) \subseteq A$ with $a_n \to a$.

Definition. Interior Point: If $A \subseteq V$ where V is a NVS, then an interior point of A is $a \in V$ such that there is an $r > 0$ where $B_r(a) \subseteq A$.

Note: Notice this means every element and hence every interior point of a set A is a limit point of A, however not every limit point is an interior point.

Proposition: For $A \subseteq V$ where V is a NVS, (1) $\overline{A} = \{\text{limit points of } A\}$ and (2) $\text{Int}(A) = \{\text{interior points of } A\}$.

Proof. (1) Let $X = \{\text{limit points of } A\}$. Let $(a_n) \subseteq X$ be such that $a_n \to a \in V$. For all $n \in \mathbb{N}$, there is a $b_n \in A$ such that $\|a_n - b_n\| < \frac{1}{n}$ since sequence in A can converge to any point in X. This means

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} b_n - a_n + a_n = \lim_{n\to\infty}(b_n - a_n) + \lim_{n\to\infty} a_n = 0 + a = a$$

since $(b_n) \subseteq A$, and $b_n \to a$, a is a limit point, hence $a \in X$ and so X is closed. X is a closed set containing A, hence $\overline{A} \subseteq X$.

Now let $x \in X$. Hence, there is a sequence $(a_n) \subseteq A$ such that $a_n \to X$. Now let $C \subseteq V$ be a closed set such that $A \subseteq C$. So $(a_n) \subseteq C$ and since C is closed and $a_n \to x$, $x \in C$. Therefore each closed set containing A contains X, i.e. $X \subseteq C$ and so $X \subseteq \bigcap_{A\subseteq C, C \text{ is closed}} C = \overline{A}$ or $\overline{A} = X$

(2) Let $X = \{\text{interior points of } A\}$. Let $a \in \text{Int}(A)$. Notice that $\text{Int}(A) \subseteq A$, hence $a \in A$. Further, notice $\text{Int}(A)$ is open, hence there is an $r > 0$ such that $B_r(a) \subseteq \text{Int}(A) \subseteq A$. This means by definition, a is an interior point of A, hence $\text{Int}(A) \subseteq X$.

Now let $x \in X$. Since x is an interior point of A, there is an $r > 0$ such that $B_r(x) \subseteq A$. Necessarily, $x \in B_r(x)$. Therefore, for each $x \in X$, there is an open subset (ball) of A which contains x. Since $\text{Int}(A)$ is the union of all open subsets of A, $B_r(x) \subseteq \text{Int}(A)$ and hence $x \in \text{Int}(A)$. This implies $X \subseteq \text{Int}(A)$ and therefore $X = \text{Int}(A)$. $\square$

Remark: Notice A is closed if and only if it contains all its limit points, hence A is closed if and only if $A = \overline{A}$. Similarly, by definition of a limit point and the equivalent definition of openness, A is open if and only all its points are interior points, hence A is open if and only if $A = \text{Int}(A)$. Note also $\text{Int}(A) \subseteq A \subseteq \overline{A}$.

2.3 Examples

Remark: Intuitively, the interior of A is all points in A which aren't on a boundary and the closure is all points on a in A or a boundary. For instance, with $A = [0, 1)$, $\text{Int}(A) = (0, 1)$, $\overline{A} = [0, 1]$. See start of module 2.3 for example in $\mathbb{R}^2$.

Example: Let $A = \{(a_n) \in \ell^1 : a_n \in \mathbb{Q}\}$. Then $\overline{A} = \ell^1$. For instance, let $x = (x_1, x_2, \ldots) \in \ell^1$ and let $\epsilon > 0$. By the density of the rationals, for all $n \in \mathbb{N}$, there is a $y_n \in \mathbb{Q}$ such that $|x_n - y_n| < \frac{\epsilon}{2^n}$. Now let $y = (y_1, y_2, \ldots)$. So $\|x - y\|_1 = \sum_{n=1}^{\infty} |x_n - y_n| < \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon$ by the sum of geometric series. Notice y is in fact in A since it is completely rational and given that $x - y$ converges ($\|x - y\|_1 < \infty$) and x converges, we must have y converges. Since hence $\ell^1 \subseteq \overline{A}$ and therefore $\ell^1 = \overline{A}$.

Remark: This gives rise to a method of proof where if we can show find elements in A is arbitrarily close to an element b, then there is a sequence $(a_n) \subseteq A$ with $a_n \to b$.

Example: Let $V = \ell^\infty$ be the NVS. Then $\overline{C_{00}} = C_0$ where C_{00} is the set of sequences which have a tail of zeroes and C_0 is the set of sequences converging to zero.

Clearly $C_{00} \subseteq C_0$ and we know C_0 is closed. So $\overline{C_{00}} \subseteq C_0$. Let $x = (x_1, x_2, \ldots) \in C_0$ and let $\epsilon > 0$. Since $x \to 0$, there is an $N \in \mathbb{N}$ such that $|x_n| < \epsilon$ for all $n \geq N$. Now let $y = (x_1, \ldots, x_{N-1}, 0, 0, \ldots)$. Hence

$$\|x - y\|_\infty = \|(0, \ldots, 0, x_N, x_{N+1}, \ldots\|_\infty = \sup\{|x_k| : k \geq N\} \leq \epsilon$$

2.4 More Properties

Remark: Note that if $A \subseteq B$ then $\overline{A} \subseteq \overline{B}$ and $\text{Int}(A) \subseteq \text{Int}(B)$

Proposition: Where V is a NVS and $A, B \subseteq V$:

1. $\overline{A \cup B} = \overline{A} \cup \overline{B}$.
2. $\text{Int}(A \cap B) = \text{Int}(A) \cap \text{Int}(B)$.
3. $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$.
4. $\text{Int}(A \cup B) \supseteq \text{Int}(A) \cup \text{Int}(B)$.

Proof. (1) Since $\overline{A} \cup \overline{B}$ is closed and $A \cup B \subseteq \overline{A} \cup \overline{B}$, we have $\overline{A \cup B} \subseteq \overline{A} \cup \overline{B}$. Now since $A, B \subseteq A \cup B$, we have that $\overline{A}, \overline{B} \subseteq \overline{A \cup B}$. Therefore, $\overline{A} \cup \overline{B} \subseteq \overline{A \cup B}$.

(2) Notice $A \cap B \subseteq A, B$ so $\text{Int}(A \cap B) \subseteq \text{Int}(A), \text{Int}(B)$ therefore $\text{Int}(A \cap B) \subseteq \text{Int}(A) \cap \text{Int}(B)$. Now notice $\text{Int}(A) \cap \text{Int}(B) \subseteq A, B$ so $\text{Int}(A) \cap \text{Int}(B) \subseteq A \cap B$. Therefore $\text{Int}(A) \cap \text{Int}(B) = \text{Int}(\text{Int}(A) \cap \text{Int}(B)) \subseteq \text{Int}(A \cap B)$ and thus $\text{Int}(A \cap B) = \text{Int}(A) \cap \text{Int}(B)$.

(3) Notice that $A \cap B \subseteq A, B$, so $\overline{A \cap B} \subseteq \overline{A}, \overline{B}$ therefore, $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$.

(4) Notice $A, B \subseteq A \cup B$, so $\text{Int}(A), \text{Int}(B) \subseteq \text{Int}(A \cup B)$ and so $\text{Int}(A) \cup \text{Int}(B) \subseteq \text{Int}(A \cup B)$. □

Example: Here is a counterexample that $\overline{A \cap B} = \overline{A} \cap \overline{B}$: $A = (0,1), B = (1,2)$. So $\overline{A \cap B} = \overline{\emptyset} = \emptyset$ and $\overline{A} \cap \overline{B} = [0,1] \cap [1,2] = \{1\}$.

Here is a counterexample that $\text{Int}(A \cup B) = \text{Int}(A) \cup \text{Int}(B)$: $A = [0,1], B = [1,2]$. So $\text{Int}(A \cup B) = (0,2)$ and $\text{Int}(A) \cup \text{Int}(B) = (0,1) \cup (1,2) = (1,2) \setminus \{1\}$.

Proposition: Let $A \subseteq V$ where V is a NVS. Then (1) $\text{Int}(V \setminus A) = V \setminus \overline{A}$ and (2) $\overline{V \setminus A} = V \setminus \text{Int}(A)$.

Proof. (1) Since $V \setminus \overline{A} \subseteq V \setminus A$ and $V \setminus \overline{A}$ is open, $V \setminus \overline{A} \subseteq \text{Int}(V \setminus A)$ by the largeness of the interior. Now note that $\text{Int}(V \setminus A) \subseteq V \setminus A$, hence $V \setminus \text{Int}(V \setminus A) \supseteq V \setminus (V \setminus A) = A$. So given $V \setminus \text{Int}(V \setminus A)$ is closed, $\overline{A} \subseteq V \setminus \text{Int}(V \setminus)$ by the smallness of the closure. Taking complements, $V \setminus \overline{A} \supseteq \text{Int}(V \setminus A)$.

(2) By (1), $\text{Int}(A) = \text{Int}(V \setminus (V \setminus A)) = V \setminus \overline{(V \setminus A)}$, taking complements we get $V \setminus \text{Int}(A) = \overline{V \setminus A}$. □

Definition. Boundary: For $A \subseteq V$ where V is a NVS, the boundary of A is $\partial(A) := \overline{A} \setminus \text{Int}(A)$.

Proposition: For $A \subseteq V$ where V is a NVS, (1) $\partial(A)$ is closed and (2) A is closed if and only if $\partial(A) \subseteq A$.

Proof. (1) Note $\partial(A) = \overline{A} \setminus \text{Int}(A) = \overline{A} \cap (V \setminus \text{Int}(A))$ and $\overline{A}$ is closed and $V \setminus \text{Int}(A)$ is closed, hence their intersection is closed.

(2) ($\implies$) A is closed, hence $A = \overline{A}$, thus $\partial(A) \subseteq \overline{A} = A$. ($\impliedby$) Suppose $\partial(A) \subseteq A$. Note $\partial(A) = \overline{A} \setminus \text{Int}(A)$, hence $\overline{A} = \partial(A) \cup \text{Int}(A)$. Since $\partial(A) \subseteq A$ and $\text{Int}(A) \subseteq A$, thus $A \subseteq \overline{A} \subseteq A$ so $A = \overline{A}$ and so A is closed. □

Week 3 Compactness

3.1 Compactness 1

Definition. Relatively Open (resp. Closed): Let V be a NVS. Let $A, B \subseteq V$. We say B is relatively open (resp. closed) in A if there is an open (resp. closed) $U \subseteq V$ such that $B = A \cap U$.

Definition. Convergence Preserving: Let V and W be NVS's and let $A \subseteq V$. A function $f : A \to W$ is convergence preserving if for all $(a_n) \subseteq A$ such that $a_n \to a \in A$, $f(a_n) \to f(a)$.

Definition. Compact: Let V be a NVS. We say $C \subseteq V$ is compact if every $(a_n) \subseteq C$ has a subsequence $a_{n_k} \to a \in C$.

Example: Every closed and bounded subset of $\mathbb{R}^n$ is compact. Note A is bounded if and only if there is an $M \in \mathbb{R}$ such that $\forall a \in A, \|a\| \leq M$. Let $(a_k) \subseteq A$. Note (a_k) is bounded. By A2 we know there is a subsequence (a_{k_ℓ}) such that $a_{k_\ell} \to a \in \mathbb{R}^n$. Since A is closed, $a \in A$.

Example: Consider the sequence $(e_n) = ((1, 0, 0, \ldots), (0, 1, 0, \ldots), \ldots)$. Notice $(e_n) \subseteq A$ has no convergent subsequence since the distance between any two points if 1 and hence it cannot be Cauchy, which contradicts the fact that it converges. As a corollary of this example, $\overline{B_1(0)} \subseteq \ell^\infty$ is closed and bounded, but not compact.

Proposition: Let V be a NVS. Let $C \subseteq V$ be compact. Then C is closed and bounded.

Proof. (Closed) Let $(a_n) \subseteq C$ be such that $a_n \to a \in V$. By the compactness of C, there is a subsequence (a_{n_k}) such that $a_{n_k} \to b \in C$. However, we must have $a = b \in C$.

(Bounded) Suppose C is not bounded for contradiction. For all $n \in \mathbb{N}$, we may find $a_n \in C$ such that $\|a_n\| \geq n$. Consider $(a_n) \subseteq C$. Every subsequence of (a_n) is unbounded, hence it diverges. This contradicts the hypothesis. □

Theorem. Heine-Borel: A set $C \subseteq \mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proposition: Let V be a NVS. Let $C \subseteq V$ be compact. If $A \subseteq C$ is closed, then A is compact.

Proof. Let $(a_n) \subseteq A \subseteq C$. There is a sequence $a_{n_k} \to a \in C$. However, A is closed, hence $a \in A$ and so A is compact. □

3.2 Open Covers

Definition. Open Cover: Let V be a NVS. Let $A \subseteq V$. An open cover of A is a collection of open sets $\{U_\alpha : \alpha \in I\}$ such that $A \subseteq \bigcup_{\alpha \in I} U_\alpha$. In the case that $|I| < \infty$ we say the collection a finite open cover.

Definition. Subcover: Let V be a NVS. Let $A \subseteq V$. Let $\{U_\alpha : \alpha \in I\}$ be an open cover of A. A subset of $\{U_\alpha : \alpha \in I\}$ which is an open cover of A is called a subcover of $\{U_\alpha : \alpha \in I\}$.

Example: Let $A = [0, 1] \subset \mathbb{R}$. An open cover of A is $A \subseteq \bigcup_{\alpha \in [0,1] \cap \mathbb{Q}} (\alpha - \frac{1}{4}, \alpha + \frac{1}{4})$. A finite subcover is $A \subseteq (-\frac{1}{4}, \frac{1}{4}) \cup (0, \frac{1}{2}) \cup (\frac{1}{4}, \frac{3}{4}) \cup (\frac{1}{2}, 1) \cup (\frac{3}{4}, \frac{5}{4})$.

Example: Let $V = \mathbb{R}^2$. Let $A = \mathbb{Z} \times \mathbb{Z} = \{(a, b) : a, b \in \mathbb{Z}\}$. Then $A \subseteq \bigcup_{\alpha \in \mathbb{Z}\times\mathbb{Z}} B_{\frac{1}{2}}(\alpha)$. Notice there is no finite subcover as there are infinitely many $\alpha \in \mathbb{Z} \times \mathbb{Z}$ each of which is covered by exactly one ball.

Example: Let $V = \mathbb{R}$, $A = (0, 1]$. Then $A \subseteq \bigcup_{\alpha \in \mathbb{N}} (\frac{1}{n}, 2)$. There is no finite subcover as taking finitely many n will leave points near 0.

Theorem: Let V be a NVS. Let $A \subseteq V$. Then $A \subseteq V$ is compact if and only if every open cover of A has a finite subcover.

3.3 Compactness 2

Lemma. Lebesgue Number Lemma: Let V be a NVS. Let $A \subseteq V$ be compact. Let $A = \bigcup_{\alpha \in I} U_\alpha$ be an open cover of A. There exists an $R > 0$, called the Lebesgue number, such that for all $a \in A$, $B_R(a) \subseteq U_\alpha$ for some $\alpha \in I$.

Proof. Suppose for the sake of contradiction no such $R > 0$ exists. In particular, for all $n \in \mathbb{N}$ there is an $a_n \in A$ such that $B_{\frac{1}{n}}(a_n) \not\subseteq U_\alpha$ for all $\alpha \in I$. Since $(a_n) \subseteq A$ and A is compact, there is a subsequence with $a_{n_k} \to a \in A$.

Now say $a \in U_{\alpha_0}$ for some $\alpha_0 \in I$. We can pick $M \in \mathbb{N}$ such that $B_{\frac{2}{M}}(a) \subseteq U_{\alpha_0}$. Moreover, since $a_{n_k} \to a$, there is an $N \in \mathbb{N}$ such that $a_{n_k} \in B_{\frac{1}{M}}(a)$ for $k \geq N$ (distance less than $\frac{1}{M}$). Then, for $k \geq N$ such that $n_k > M$, take $x \in B_{\frac{1}{M}}(a_{n_k})$. So

$$\|x - a\| = \|x - a_{n_k} + a_{n_k} - a\| \leq \|x - a_{n_k}\| + \|a_{n_k} - a\| < \frac{1}{M} + \frac{1}{M} = \frac{2}{M}$$

Therefore, $x \in B_{\frac{2}{M}}(a)$ and so $B_{\frac{1}{M}}(a_{n_k}) \subseteq B_{\frac{2}{M}}(a) \subseteq U_{\alpha_0}$. Now since $n_k > M$, $B_{\frac{1}{n_k}}(a_{n_k}) \subseteq B_{\frac{1}{M}}(a_{n_k}) \subseteq U_{\alpha_0}$. This is a contradiction by our first assumption. □

Proposition: Let V be a NVS. If $A \subseteq V$ is compact, then every open cover of A has a finite subcover.

Proof. Suppose $A \subseteq V$ is compact. Let $A \subseteq \bigcup_{\alpha \in I} U_\alpha$ be an open cover of A. We may find a Lebesgue number $R > 0$ as in the above lemma. If there is $a_1, \ldots, a_n \in A$ such that $A \subseteq B_R(a_1) \cup \cdots \cup B_R(a_n)$, then we are done.

Otherwise, suppose no such cover exists. Find $a_1 \in A$. Since no covering of open balls of size R exist, there is an $a_2 \in A$ such that $a_2 \notin B_R(a_1)$. Further, there is an $a_3 \in A$ such that $a_3 \notin B_R(a_1) \cup B_R(a_2)$. We keep doing this indefinitely. We have then a sequence $(a_n) \subseteq A$ and by the compactness of A, a has a convergent subsequence. We have that since $a_m \notin B_R(a_n)$ therefore $\|a_m - a_n\| \geq R$ for all $n < m$. This means there is no Cauchy subsequence of (a_n), therefore there is no convergent subsequence. This is a contradiction. □

3.4 Compactness 3

Lemma: Let V be a NVS. Let $A \subseteq V$. Suppose every open cover of A has a finite subcover. If $A \subseteq \bigcup_{\alpha \in I} U_\alpha$, where each U_α is relatively open in A, then there are $\alpha_1, \dots, \alpha_n \in I$ such that $A \subseteq U_{\alpha_1} \cup \cdots \cup U_{\alpha_n}$.

Proof. Suppose $A \subseteq \bigcup_{\alpha \in I} U_\alpha$ where each $U_\alpha = A \cap O_\alpha$ where $O \subseteq V$ is open. So we have $A \subseteq \bigcup_{\alpha \in I} (A \cap O_\alpha) = A \cap (\bigcup_{\alpha \in I} O_\alpha) \subseteq \bigcup_{\alpha \in I} O_\alpha$. Hence, $\bigcup_{\alpha \in I}$ is an open covering of A, however we have that all open coverings have finite subcovers by our hypothesis. Hence, there is a finite subcovering $A \subseteq O_{\alpha_1} \cup \cdots \cup O_{\alpha_n}$ and therefore $A \subseteq A \cap (O_{\alpha_1} \cup \cdots \cup O_{\alpha_n}) = U_{\alpha_1} \cup \cdots \cup U_{\alpha_n}$ where each U is relatively open. □

Proposition: Let V be a NVS. Suppose $A \subseteq V$ is such that every open cover of A has a finite subcover. Then A is compact.

Proof. Suppose A is as described. Consider $(a_n) \subseteq A$. For $k \in \mathbb{N}$, let $C_k = \overline{\{a_n : n \geq k\}} \cap A$. We wish to show that $\bigcap_{k=1}^{\infty} C_k \neq \emptyset$. Notice each C_k is relatively closed in A. Hence every $U_k = A \setminus C_k$ is relatively open in A.

Now suppose for the sake of contradiction that $\bigcap_{k=1}^{\infty} C_k = \emptyset$, then by De Morgan's law

$$A = A \setminus \emptyset = A \setminus \left(\bigcap_{k=1}^{\infty} C_k \right) = \bigcup_{k=1}^{\infty} (A \setminus C_k) = \bigcup_{k=1}^{\infty} U_k$$

By our above lemma, there is $i_1, \dots, i_\ell$ such that $A \subseteq U_{i_1} \cup \cdots \cup U_{i_\ell}$. Now since by definition $C_1 \supseteq C_2 \supseteq C_3 \supseteq \cdots$, we have $U_1 \subseteq U_2 \subseteq U_3 \subseteq \cdots$. Therefore, $A \subseteq U_{i_\ell} \subseteq A$, hence $A = U_{i_\ell}$. We have then $C_{i_\ell} = A \setminus U_{i_\ell} = A \setminus A = \emptyset$. However, we necessarily have $a_{i_\ell} \in C_{i_\ell} = \emptyset$ which is a contradiction. Therefore, there is an $a \in \bigcap_{k=1}^{\infty} C_k$. In particular, since a is in the tail of all sequences, we may find $n_1 < n_2 < \cdots$ such that $\|a_{n_k} - a\| < \frac{1}{k}$ for all $k \in \mathbb{N}$. Hence, $(a_{n_k}) \subseteq A$ has $a_{n_K} \to a \in A$. □

Week 4 Limits and Continuity

4.1 Limits

Definition. Limit: Let V, W be NVSs. Let $A \subseteq V$ and $f : A \to W$. We say the limit of $f(x)$ as x approaches $a \in V$ is $w \in W$ if: (1) $a \in \overline{A \setminus \{a\}}$ and (2) for all $\epsilon > 0$ there is a $\delta > 0$ such that $x \in A$ with $0 < \|x - a\| < \delta$ then $\|f(x) - w\| < \epsilon$. We write $\lim_{x \to a} f(x) = w$. Note this w is unique.

Definition. Isolated Point: Let V be a NVS. Let $A \subseteq V$. If $a \notin \overline{A \setminus \{a\}}$ then we say $a \in V$ is an isolated point with respect to A.

Note: Note that if a is isolated, then there is an $r > 0$ such that $B_r(a) \cap A = \{a\}$ or $B_r(a) \cap A = \emptyset$. To see this, suppose no such r exists. Then for all $n \in \mathbb{N}$, there is an $a_n \neq a$ such that $a_n \in B_{\frac{1}{n}}(a)$. This implies for all $\epsilon > 0$, picking $N = \frac{1}{\epsilon}$, there is an $a \neq a_n \in B_{\frac{1}{n}}(a)$ for all $n \geq N$ and thus $\|a - a_n\| < \frac{1}{n} < \frac{1}{N} = \epsilon$. Thus $a_n \to a$, hence a is a limit point of $A \setminus \{a\}$, or $a \in \overline{A \setminus \{a\}}$.

Remark: We need a not to be a limit point in our definition of continuity since if it were then there would be an $r > 0$ such that there is no $x \in A$ with $0 < \|x - a\| < r$.

Proposition. Limits Preserve Order: Let V be a NVS. Let $A \subseteq V$ and $a \in \overline{A \setminus \{a\}}$. Let $f, g : A \to \mathbb{R}$. If $\lim_{x \to a} f(x)$ and $\lim_{x \to a} g(x)$ exist and $f(x) \leq g(x)$ for all $x \in A$, then $\lim_{x \to a} f(x) \leq \lim_{x \to a} g(x)$.

Proposition. Squeeze Theorem: Let V be a NVS. Let $A \subseteq V$ and $a \in \overline{A \setminus \{a\}}$. Let $f, g, h : A \to \mathbb{R}$. If $f(x) \leq g(x) \leq h(x)$ and $\lim_{x \to a} f(x) = \lim_{x \to a} h(x) = L$ exists, then $\lim_{x \to a} g(x) = L$ as well.

Remark: We will use all limit laws freely as their proofs are similar to the real case.

Example: To find $\lim_{(x,y,z) \to (0,0,0)} \frac{xy^2 + x^2z + xyz}{\sqrt{x^2 + y^2 + z^2}}$, notice

$$\begin{aligned} 0 &\leq \left| \frac{xy^2 + x^2z + xyz}{\sqrt{x^2 + y^2 + z^2}} \right| \\ &\leq \frac{|xy^2 + x^2z + xyz|}{\sqrt{x^2}} \\ &= \frac{|xy^2 + x^2z + xyz|}{|x|} \\ &\leq \frac{|x|y^2 + x^2|z| + |x||y||z|}{|x|} \\ &= y^2 + |x||z| + |y||z| \end{aligned} \tag{*}$$

Using limit laws it is easy to see the limit approaches 0. * Note in the case that $x = 0$ we have the above is vacuously true since we multiply by 0 in the denominator.

Example: To prove $\lim_{(x,y) \to (0,0)} \frac{xy^2}{x^2 + y^4}$ does not exist, consider two sequences. First, as $(\frac{1}{n}, 0) \to (0, 0)$ we have the limit goes to 0. Second, as $(\frac{1}{n^2}, \frac{1}{n}) \to (0, 0)$ we have the limit goes to $\frac{1}{2}$. Since the limit is unique, there can be no such limit.

4.2 Continuity

Notation: Unless specified otherwise, all functions are of the form $f : A \to W$ where $A \subseteq V$ and V and W are NVSs.

Definition. Continuous at a Point: We say f is continuous (cts) at $a \in A$ if for all $\epsilon > 0$ there is a $\delta > 0$ such that if $x \in A$ with $\|x - a\| < \delta$ then $\|f(x) - f(a)\| < \epsilon$.

Remark: If $a \in \overline{A \setminus \{a\}}$ (i.e. a is not isolated) then f is continuous at a if and only if $\lim_{x \to a} f(x) = f(a)$. If $a \notin \overline{A \setminus \{a\}}$ then f is necessarily continuous at a. This follows since if a is isolated then there is an $r > 0$ such that $B_r(a) \cap A = \{a\}$. Hence for any $\epsilon > 0$ picking $\delta = r$ will have all $x \in A$ with $\|x - a\| < \delta$ be such that $x = a$.

Definition. Continuous Function: If f is continuous at every $a \in A$, then we say f is continuous.

Proposition: The following are equivalent

1. f is continuous
2. f preserves convergence
3. For all open $U \subseteq W$, $f^{-1}(U)$ is relatively open in A.

Proof. Notice we have already proved (2) $\iff$ (3) if assignment 2. (1 $\implies$ 2) Suppose f is continuous. Let $(a_n) \subseteq A$ such that $a_n \to a \in A$. Let $\epsilon > 0$ be given. There is a $\delta > 0$ such that if $x \in A$ and $\|x - a\| < \delta$ then $\|f(x) - f(a)\| < \epsilon$.

Taken $N \in \mathbb{N}$ such that $\|a_n - a\| < \delta$ for all $n \geq N$. Then for $n \geq N$, $\|f(a_n) - f(a)\| < \epsilon$, therefore $f(a_n) \to f(a)$.

(2 $\implies$ 1) Assume f preserves convergence. For the sake of contradiction, suppose f is discontinuous at a. Therefore, there is an $\epsilon > 0$ and a sequence $(a_n) \subseteq A$ such that $\|a_n - a\| < \frac{1}{n}$, but $\|f(a_n) - f(a)\| \geq \epsilon$. Therefore $a_n \to a$ but $f(a_n) \not\to f(a)$, this is a contradiction. □

Example: Consider $P_i : \mathbb{R}^n \to \mathbb{R}$ for $1 \leq i \leq n$. Then $P_i(x_1, \ldots, x_n) = x_i$ is continuous (called the ith projection). Proof: Let $(a_k) \subseteq \mathbb{R}^n$ such that $a_k \to a \in \mathbb{R}^n$. Suppose $a_k = (a_k^{(1)}, a_k^{(2)}, \ldots, a_k^{(n)})$ and $a = (b_1, b_2, \ldots, b_n)$. We know $a_k^{(i)} \to b_i$ but this implies $P_i(a_k) \to P_i(a)$ therefore P_i is continuous.

Proposition: (1) If $f, g : A \to W$ are continuous then $f + g$ and αf (for $\alpha \in \mathbb{R}$) are continuous. (2) If $f : A \to W_1$ is continuous and $g : B \to W_2$ where $B \subseteq W_1$ is continuous, then $g \circ f$ is continuous.

Proof. (1) Let $(a_n) \subseteq A$ is such that $a_n \to a$. Since f, g are continuous $f(a_n) \to f(a)$ and $g(a_n) \to g(a)$. By the limit laws, $f(a_n) + g(a_n) \to f(a) + g(a)$ and $\alpha f(a_n) \to \alpha f(a)$, therefore $f + g$ and αf are continuous.

(2) Let $(a_n) \subseteq A$ such that $a_n \to a$. Since f is continuous, $f(a_n) \to f(a)$. Since g is continuous, $g(f(a_n)) \to g(f(a))$ so $g \circ f$ is continuous. □

Definition. Isometric Isomorphism: An isometric isomorphism between two NVS's V and W is an isomorphism (invertible linear transformation) $T : V \to W$ such that $||T(x)|| =$

$||x||$ for all $x \in V$. If there exists an isometric isomorphism between V and W, we say V and W are isometrically isomorphic and write $V \cong W$.

Theorem. Completion: Let V be a NVS. Let $\sim$ be the equivalence relation such that if $a = (a_n) \subseteq V$ and $b = (b_n) \subseteq V$, then $a \sim b$ if and only if $\lim\limits_{n\to\infty} \|a_n - b_n\| = 0$. Let $\hat{V}$ be the set of equivalence classes of Cauchy sequences of V. Then $\hat{V}$ is a Banach NVS equipped with the norm (distance) $\|[a]-[b]\| := \lim\limits_{n\to\infty} \|a_n - b_n\|$. Further, V is isometrically isomorphic to a subspace of $\hat{V}$.

Proof. See A4 Q4. □

4.3 Uniform Continuity

Definition. Uniformly Continuous: Let V, W be NVSs and $A \subseteq V$. We say $f : A \to W$ is uniformly continuous fif for all $\epsilon > 0$ there is a $\delta > 0$ such that if $x, a \in A$ with $\|x - a\| < \delta$ then $\|f(x) - f(a)\| < \epsilon$.

Remark: The big idea is that one δ works for all $a \in A$. Notice uniform continuity implies continuity.

Definition. Lipschitz: Let V, W be NVSs and $A \subseteq V$. $f : A \to W$ is Lipschitz if there is an $M > 0$ such that $\|f(a) - f(b)\| \leq M\|a - b\|$ for all $a, b \in A$.

Proposition: Let $f : A \to W$. If f is Lipschitz then f is uniformly continuous.

Proof. Let $\epsilon > 0$ be given. Let $\delta = \frac{\epsilon}{M}$. If $a, b \in A$ with $\|a - b\| < \delta$, then $\|f(a) - f(b)\| \leq M\|a - b\| < M\delta = \epsilon$ and so f is uniformly continuous. □

Example: Let $f : [a, \infty) \to \mathbb{R}$ with $f(x) = x^2$. f is not uniformly continuous. To see this, let $a_n = n + \frac{1}{n}$ and $b_n = n$ for $n \in \mathbb{N}$. Notice $\|a_n - b_n\| = \frac{1}{n} \to 0$, however $\|f(a_n) - f(b_n)\| = (n + \frac{1}{n})^2 - n^2 = 2 + \frac{1}{n^2} \geq 2$. Hence picking $\epsilon = 1$, there is no point δ such that $\|a_n - b_n\| < \delta \implies \|f(a_n) - f(b_n)\|$ since their difference is always at least 2.

Example: Let $f : (0, 1] \to \mathbb{R}$ with $f(x) = \ln x$. f is not uniformly continuous. To see this, let $a_n = \frac{1}{n}$ and $b_n = \frac{1}{n^2}$ for $n \in \mathbb{N}$. Notice $\|a_n - b_n\| = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2} \to 0$. However, $\|f(a_n) - f(b_n)\| = \ln(\frac{1}{n}) - \ln(\frac{1}{n^2}) = \ln(n^2) - \ln(n) = \ln(n) \geq \ln(2)$ for $n \geq 2$. Thus even as $a_n \to 0$ and $b_n \to 0$ the difference of the function values is atleast $\ln 2$.

Theorem: If $C \subseteq V$ is compact and $f : C \to W$ is continuous then f is uniformly continuous.

Proof. Suppose for the sake of contradiction f is not uniformly continuous. Then there is $(a_n), (b_n) \subseteq C$ such that for some $\epsilon > 0$, $\|a_n - b_n\| < \frac{1}{n}$ but $\|f(a_n) - f(b_n)\| \geq \epsilon$.

By compactness: $a_{n_k} \to a \in C$ and $b_{n_k} = b_{n_k} - a_{n_k} + a_{n_k} \to 0 + a = a$. By continuity, $f(a_{n_k}) \to f$ and $f(b_{n_k}) \to f(b)$, therefore $\|f(a_{n_k}) - f(b_{n_k})\| \to 0$, however we have that $\|f(a_{n_k}) - f(b_{n_k})\| \geq \epsilon$, a contradiction. □

4.4 Extreme Value Theorem

Proposition: Let $C \subseteq V$ be compact, and $f : C \to W$ be continuous. Then $f(C)$ is compact.

Proof. Let $(f(a_n)) \subseteq f(C)$ with $a_n \in C$. So $(a_n) \subseteq C$ and by compactness $a_{n_k} \to a \in C$. Since f is continuous, $f(a_{n_k}) \to f(a) \in f(C)$, therefore $f(C)$ is compact. □

Lemma: If $A \subseteq \mathbb{R}$ is bounded and non-empty, then $\inf A, \sup A \in \overline{A}$.

Proof. For all $n \in \mathbb{N}$, $\sup A - \frac{1}{n} < a_n \leq \sup A$ where $(a_n) \subseteq A$. Therefore, by squeeze theorem $a_n \to \sup A$ making $\sup A \in \overline{A}$. The proof for $\inf A$ follows similarly. □

Theorem. Extreme Value Theorem: (EVT) Let V be a NVS. Let $\emptyset \neq C \subseteq V$ be compact and $f : C \to \mathbb{R}$ be continuous. There is $a, b \in C$ such that $f(a) = \min f(C)$ and $f(b) = \max f(C)$.

Proof. Since C is compact, $f(C)$ is compact and therefore closed and bounded. Thus by our lemma, $\sup f(C), \inf f(C) \in \overline{f(C)} = f(C)$ (by closedness). Therefore, there are $a, b \in C$ such that $f(a) = \inf f(C) = \min f(C)$ and $f(b) = \sup f(C) = \max f(C)$. □

Proposition: Let V, W be NVSs. Let $K \subseteq V$ be compact. Define $C(K, W) := \{f : K \to W : f \text{ is continuous}\}$. Then $C(K, W)$ equipped with the uniform norm $\|f\|_\infty = \max\{\|f(x)\| : x \in K\}$ is a NVS.

Proof. Notice $f : K \to W$ is continuous and $\|\cdot\| : W \to \mathbb{R}$ is continuous. Therefore, $\|\cdot\| \circ f : K \to \mathbb{R}$ is continuous. By the extreme value theorem $\max\{\|f(x)\| : x \in K\}$ exists. The rest of the properties of a norm follow from max properties. □

Week 5 Sequences and Spaces of Functions

5.1 Sequences of Functions

Definition. Connected: Let V be a NVS. A subset $A \subseteq V$ is said to be connected if there does not exist non-empty, disjoint, relatively open $U, W \subseteq A$ such that $A = U \cup W$.

Proposition. Path-Connected: Let $A \subseteq V$. A is connected if and only if for all $a, b \in A$ there exists a continuous function $f : [0, 1] \to A$ such that $f(0) = a$ and $f(1) = b$.

Proof. See A5 Q2 for backwards direction. Forwards direction left as exercise. □

Definition. Pointwise Convergence: Let V, W be NVSs and $A \subseteq V$. Let $f_n : A \to W$ be a sequence of functions and let $f : A \to W$. We say f_n converges pointwise to f if $f_n(x) \to f(x)$ for all $x \in A$.

Definition. Uniform Convergence: Let V, W be NVSs and $A \subseteq V$. Let $f_n : A \to W$ be a sequence of functions and let $f : A \to W$. We say f_n converges to f uniformly if for all $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that $\|f_n(x) - f(x)\| < \epsilon$ for all $n \geq N$ and $x \in A$.

Remark: $f_n \to f$ pointwise if $\forall x \in A, \forall \epsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N, \|f_n(x) - f(x)\| < \epsilon$. $f_n \to f$ uniformly if $\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall x \in A, \forall n \geq N, \|f_n(x) - f(x)\| < \epsilon$. That is uniform convergence requires one N for all $x \in A$, whereas pointwise requires for all $x \in A$ that there is an N.

Remark: We define $\|f_n - f\|_\infty := \sup\{\|f_n(x) - f(x)\| : x \in A\}$ where $f_n, f : A \to W$ and $A \subseteq V$. Note that $f_n \to f$ uniformly if and only if $\|f_n - f\|_\infty < \infty$ eventually and $\|f_n - f\|_\infty \to 0$ for a tail where $\|f_n - f\|_\infty < \infty$.

5.2 Examples

Example: Consider $f_n : (0,1) \to \mathbb{R}$ given by $f_n(x) = \frac{nx}{1+nx}$. Clearly for $x \in (0,1)$, $f_n(x) = \frac{nx}{1+nx} \to 1$ hence $f_n \to 1$ pointwise. However, $forn \geq 1$, $|f_n(\frac{1}{n}) - 1| = \frac{1}{2}$, hence $\|f_n - f\|_\infty \not\to 0$. So the convergence is not uniform.

Example: Consider $f_n : C_0 \to \mathbb{R}$ given by $f_n((a_k)) = a_n$ (C_0 is space of 0 tail sequences). Notice for $(a_k) \in C_0$ $f_n((a_k)) = a_n \to 0$ as $n \to \infty$ since $a_n \to 0$. Hence $f_n \to 0$ pointwise. Now notice for $n \in \mathbb{N}$, $|f_n(\underbrace{1, 1, \ldots, 1}_{n \text{ times}}, 0, 0, \ldots) - 0| = |1 - 0| = 1$ hence $\|f_n - 0\|_\infty \geq 1$ implying $\|f_n - 0\|_\infty \not\to 0$ and thus $f_n \to 0$ uniformly.

Example: Consider $f_n : [0,1] \times [0,1] \to \mathbb{R}$ given by $f_n(a,b) = \frac{a^n}{n} + \frac{1}{b+n}$. Notice both $\frac{a^n}{n} \to 0$ and $\frac{1}{b+n} \to 0$ hence $f_n \to 0$ pointwise. Now notice $|f_n(a,b) - 0| = \frac{a^n}{n} + \frac{1}{b+n} \leq \frac{1}{n} + \frac{1}{n} = \frac{2}{n}$, thus $|f_n(a,b) - 0| \to 0$. Since this was for arbitrary a, b, we have $\|f_n - 0\|_\infty \leq \frac{2}{n} \to 0$ thus $f_n \to 0$ uniformly.

5.3 Theorem A

Example: Consider $f_n : [0,1] \to \mathbb{R}$ given by $f_n(x) = x^n$. Notice each f_n is continuous, however

$$f_n \to f = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{if } x \neq 1 \end{cases}$$

pointwise, where f is not continuous. Notice however this is not true for continuous sequences of uniformly convergent functions.

Theorem: Let V, W be NVSs and $A \subseteq V$. Let (f_n) be a sequences of functions $f_n : A \to W$. If f_n is continuous for all $n \in \mathbb{N}$ and $f_n \to f$ uniformly, then f is continuous.

Proof. Let $\epsilon > 0$. Let $(a_n) \subseteq A$ be such that $a_n \to a$. We may find $N \in \mathbb{N}$ such that $\|f_N - f\| < \frac{\epsilon}{3}$. Since f_N is continuous, there is an $M \in \mathbb{N}$ such that $\|f_N(a_n) - f_N(a)\| < \frac{\epsilon}{3}$ for all $n \geq M$. For $n \geq M$,

$$\begin{aligned}\|f(a_n) - f(a)\| &= \|f(a_n) - f_N(a_n) + f_N(a_n) - f_N(a) + f_N(a) - f(a)\| \\ &\leq \|f(a_n) - f_N(a_n)\| + \|f_N(a_n) - f_N(a)\| + \|f_N(a) - f(a)\| \\ &\leq \|f - f_N\|_\infty + \|f_N(a_n) - f_N(a)\| + \|f_N - f\|_\infty \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon\end{aligned}$$

□

5.4 Theorem B

Theorem: Let V be a NVS, $A \subseteq V$ be compact, and W be a Banach space. Then $(C(A, W), \|\cdot\|_\infty)$ is a Banach space.

Proof. Let $(f_n) \subseteq C(A, W)$ be Cauchy. Let $\epsilon > 0$ be given. There is an $N \in \mathbb{N}$ such that $\|f_n - f_m\|_\infty < \epsilon$, for all $n, m \geq N$. Then for any $x \in A$ and $n, m \geq N$, $\|f_n(x) - f_m(x)\| \leq \|f_n - f_m\|_\infty < \epsilon$. hence $(f_n(x)) \subseteq W$ is Cauchy. Since W is a Banach space, we know that $f_n(x) \to f(x) \in W$, for some $f(x) \in W$. Since this is true for all $x \in A$, we have constructed a function $f : A \to W$ such that $x \mapsto f(x)$. Notice further that $f_n \to f$ pointwise.

Now for all $x \in A$ and $n \geq N$, we have $\lim_{m\to\infty} \|f_n(x) - f_m(x)\| \leq \epsilon$ since limits preserve order. Therefore, (recall $f_m(x) \to f(x)$ and $\|\cdot\|$ is continuous, i.e. preserves convergence) $\|f_n(x) - f(x)\| \leq \epsilon$. Since x was arbitrary, $\|f_n - f\|_\infty \leq \epsilon$. Therefore $f_n \to f$ uniformly. By the previous theorem, $f \in C(A, W)$, and so $f_n \to f \in C(A, W)$ so $C(A, W)$ is a Banach space. □

Unit 2 Differentiation

Week 6 Multi-variable Derivatives

6.1 Partial Derivatives

Definition. Scalar Function: A scalar function is a function of the form $f : A \to \mathbb{R}$ for $A \subseteq \mathbb{R}^n$.

Remark: If $f : A \to \mathbb{R}^m$ for $A \subseteq \mathbb{R}^n$ is a function, then there are scalar functions $f_1, \ldots, f_m$ such that $f = (f_1, f_2, \ldots, f_m)$.

Example: Consider $f : \mathbb{R}^3 \to \mathbb{R}^2$ defined as $f(x, y, z) = (cze^y, x^2 + z^2)$. Then $f_1(x, y, z) = xze^y$ and $f_2(x, y, z) = x^2 + z^2$ are the associated scalar functions, i.e. $f = (f_1, f_2)$.

Definition. Partial Derivative of Scalar Function: Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$. Let $\{e_1, \dots, e_n\}$ be the standard basis for $\mathbb{R}^n$. For $1 \leq i \leq n$, we define the ith partial derivative of f at $a = (a_1, a_2, \dots, a_n) \in A$ by

$$f_{x_i}(a) \overset{\text{Notation}}{=} \frac{\partial f}{\partial x_i}(a) := \lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h},$$

provided the limit exists.

Remark: For $a \in A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}$, writing $f(x_1, \dots, x_n)$.

1. $f_{x_i}(a)$ is the derivative at a with respect to the variable x_i, treating x_j for $j \neq i$ as constant. (Since we add a scalar multiple of e_i, leaving all other variable alone.)

2. $f_{x_i}(a)$ is the slope of the tangent line to the surface $y = f(x_1, \dots, x_n)$ which is parallel to e_i.

Example: Let $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y)$. Let $a = (a_1, a_2) \in \mathbb{R}^2$. We have

$$f_x(a) = \lim_{h \to 0} \frac{f(a + h, e_1) - f(a)}{h} = \lim_{h \to 0} \frac{f(a_1 + h, a_2) - f(a_1, a_2)}{h}$$

this is why we say we are holding the other variable(s) constant.

Notation: Since we can think of $f_{x_i} = \frac{\partial f}{\partial x_i}$ as a function, plugging in variables as we wish, we write

$$f_{x_i}(x_1, x_2, \dots, x_n) = \frac{\partial}{\partial x_i} f(x_1, x_2, \dots, x_n)$$

to denote the derivative function.

Example: Consider $f : \mathbb{R}^3 \to \mathbb{R}$ given by $f(x, y, z,) = xy^2z + e^{xy}$. Then we have

$$\begin{aligned} f_x(x, y, z) &= \frac{\partial}{\partial x} f(x, y, z) = y^2 z + ye^{xy} \\ f_y(x, y, z) &= \frac{\partial}{\partial y} f(x, y, z) = 2xyz + xe^{xy} \\ f_z(x, y, z) &= \frac{\partial}{\partial z} f(x, y, z) = xy^2 \end{aligned}$$

This further shows why we say we consider the other variables as constant, because you can evaluate it as a usual derivative with respect only to your current variable consider all else as constant.

Definition. Partial Derivative: Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$ where $f = (f_1, \dots, f_n)$. For $a \in A$ and $1 \leq i \leq n$, the ith partial derivative of f is

$$f_{x_i}(a) \overset{\text{Notation}}{=} \frac{\partial f}{\partial x_i}(a) := \left(\frac{\partial f_1}{\partial x_i}(a), \frac{\partial f_2}{\partial x_i}(a), \dots, \frac{\partial f_m}{\partial x_i}(a) \right)$$

provided it exists.

Example: Consider $f : \mathbb{R}^2 \to \mathbb{R}^3$ given by $f(x, y) = (2x^2y, 4x, e^{xy})$. We have then

$$\begin{aligned} f_x(x, y) &= (4xy, 4, ye^{xy}) \\ f_y(x, y) &= (2x^2, 0, xe^{xy}) \end{aligned}$$

6.2 Differentiability

Remark: Recall for $A \subseteq \mathbb{R}$, $f : A \to \mathbb{R}$, we say f is differentiable at $a \in A$ if and only if (1) $a \in \text{Int}(A)$ (i.e. it's not a boundary point) and (2)

$$\lim_{h\to 0} \frac{f(a+h)-f(a)}{h} = f'(a) \text{ exists} \iff \lim_{h\to 0} \frac{f(a+h)-f(a)-mh}{h} = 0 \text{ for some } m \in \mathbb{R}$$

Part of the reason we need $a \in \text{Int}(A)$ is that for small enough h, the limit is possibly existent. If it weren't, then for all $h \neq 0$, $a+h$ is not in the domain of f.

Remark: $T : \mathbb{R} \to \mathbb{R}$ is a linear transformation if and only if $T(x) = mx$ for some $m \in \mathbb{R}$. (Proof) Let $x \in \mathbb{R}$. So $T(x) = x \cdot T(1) = mx$. Let $x, y \in \mathbb{R}$ and $c \in \mathbb{R}$. Then $T(cx + y) = m(cx + y) = cmx + my = cT(x) + T(y)$.

Therefore,

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

if and only if f over $[a, a+h]$ can be approximated arbitrarily well by the line $T(x) = mx$.

Notation: Recall $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) = \{T : \mathbb{R}^n \to \mathbb{R}^m : T \text{ is linear}\}$.

Definition. Differentiability: Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. We say f is differentiable at $a \in A$ if (1) $a \in \text{Int}(A)$ and (2) there is a line $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ such that

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

Note: This means equivalently f is differentiable if and only if there is a matrix A such that

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

Remark: By (1) of differentiability, $f(a+h)$ is defined for small enough h. Thus for an open ball centered at a, the way the function f changes can be arbitrarily well approximated by the change in a linear transformation.

Remark: Recall from MATH 146 that for $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. Where B is the standard matrix of T (matrix representation of T relative to standard basis), $T(x) = Bx$ for all $x \in \mathbb{R}^n$. Recall that from A1Q2 we have the operator norm $\|A\|_{op} := \sup\{\|Ax\| : x \in \mathbb{R}^n, \|x\| = 1\}$ is a norm and such that $\|Ax\|_{op} \leq \|A\|_{op} \cdot \|x\|$.

Theorem: Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. If f is differentiable at $a \in A$, then f is continuous at a.

Proof. Since f is differentiable at a, there is a $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ such that

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

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

where $B = [T]_E$ where E is the standard basis. This implies we may find $\delta > 0$ such that if $0 < \|h - 0\| = \|h\| < \delta$ then

$$\begin{aligned}\left\|\frac{f(a+h)-f(a)-Bh}{\|h\|}\right\| &< 1\\ \|f(a+h)-f(a)-Bh\| &< \|h\|\\ \|f(a+h)-f(a)\| - \|Bh\| &< \|h\|\\ \|f(a+h)-f(a)\| &< \|Bh\| + \|h\|\\ \|f(a+h)-f(a) &< \|B\|_{op}\cdot\|h\| + \|h\|\end{aligned}$$

So as $h \to 0$ we have $\|B\|_{op}\cdot\|h\|+\|h\| \to 0$. Therefore by the squeeze theorem $\lim_{h\to 0} f(a+h) = f(a)$. Letting $x = a+h$, we have $\lim_{x\to a} f(x) = f(a)$, hence f is continuous at a. □

Definition. Differentiable on Set: Let $U \subseteq \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^m$. We say f is differentiable (on U) if f is differentiable at every point in U.

Remark: The openness in the above definition is required because we require each differentiable point to be an interior point.

6.3 Total Derivative

Example: Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} (x^2+y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) & \text{if } (x,y) \neq 0\\ 0 & \text{if } (x,y) = 0\end{cases}$$

We will show f is differentiable at $(0,0)$. (Notice $(0,0) \in \text{Int}(A)$.) Consider $B = (0,0)$. Then

$$\begin{aligned}\lim_{h\to 0}\frac{f(a+h)-f(a)-Bh}{\|h\|} &= \lim_{h\to 0}\frac{f(h)}{\|h\|}\\ &= \lim_{(x,y)\to(0,0)}(x^2+y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right)\cdot\frac{1}{\sqrt{x^2+y^2}}\\ &= \lim_{(x,y)\to(0,0)}\sqrt{x^2+y^2}\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right)\\ &= 0 \qquad \text{By squeeze theorem, } \sqrt{\|x^2+y^2\|} \to 0\end{aligned}$$

How do we know to choose B?

Remark. Investigation of B: Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. Suppose f is differentiable at $a \in A$. Then there is a $B \in M_{m\times n}(\mathbb{R})$ such that $\lim_{h\to 0}\frac{f(a+h)-f(a)-Bh}{\|h\|} = 0$. Now let

$\{e_1, \ldots, e_n\}$ be the standard basis vector for $\mathbb{R}^n$. since $\mathbb{R} \ni t \to 0_{\mathbb{R}}$ we have $\mathbb{R}^n \ni te_i \to 0_{\mathbb{R}^n}$, we also have (since h is arbitrary)

$$\lim_{t\to 0} \frac{f(a+te_i) - f(a) - Bte_i}{|t|} = 0$$

Therefore

$$\lim_{t\to 0^+} \frac{f(a+te_i) - f(a) - tBe_i}{t} = 0$$
$$\lim_{t\to 0^+} \frac{f(a+te_i) - f(a)}{t} = Be_i$$

and

$$\lim_{t\to 0^-} \frac{f(a+te_i) - f(a) - tBe_i}{-t} = 0$$
$$\lim_{t\to 0^-} \frac{f(a+te_i) - f(a)}{-t} = -tBe_i$$
$$\lim_{t\to 0^-} \frac{f(a+te_i) - f(a)}{t} = tBe_i$$

Thus we have

$$Be_i = \lim_{t\to 0} \frac{f(a+te_i) - f(a)}{t}$$
$$Be_i = \frac{\partial f}{\partial x_i}(a)$$
$$\text{Col}_i(B) = \left(\frac{\partial f_1}{\partial x_i}(a), \frac{\partial f_2}{\partial x_i}(a), \ldots, \frac{\partial f_n}{\partial x_i}(a)\right)$$

In particular, for $B = (b_{ij})$, we have $b_{ij} = \frac{\partial f_i}{\partial x_j}(a)$.

Definition. Total Derivative: Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. For $a \in A$ we call the matrix

$$Df(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \in M_{m\times n}(\mathbb{R})$$

the total derivative of f at c, provided it exists.

Theorem: Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. If f is differentiable at $a \in A$, then (1) for all $1 \leq j \leq n$, $\frac{\partial f}{\partial x_j}(a)$ exists (it is $\text{Col}_j(Df(a))$ exactly) and (2)

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

See investigation above for proof.

Definition. Gradient: Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}$ be a scalar function. For $a \in A$, we call $Df(a)$ the gradient of f at a and label it by $\nabla f(a)$. That is

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \frac{\partial f}{\partial x_2}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\right)$$

Remark: In the example at the start of the lecture, we chose B simply by computing $\nabla f(0,0) = (0,0)$.

Notation: Let $U \subseteq \mathbb{R}^n$ be open and $f : U \to \mathbb{R}^m$. Notice we have then the following

$$f_{x_i}(a) = \frac{\partial f}{\partial x_i}(a) = \left(\frac{\partial f_1}{\partial x_i}(a), \frac{\partial f_2}{\partial x_i}(a), \ldots, \frac{\partial f_m}{\partial x_i}(a)\right) \in M_{m\times 1}$$

and

$$\nabla f_i(a) = \frac{\partial f_i}{\partial x}(a) = \left(\frac{\partial f_i}{\partial x_1}(a), \frac{\partial f_i}{\partial x_2}(a), \ldots, \frac{\partial f_i}{\partial x_n}(a)\right) \in M_{1\times n}$$

thus for $1 \leq i \leq m$ and $1 \leq j \leq n$

$$\text{Row}_i(Df(a)) = \nabla f_i(a) = \frac{\partial f_i}{\partial x}(a) \qquad \text{and} \qquad \text{Col}_j(Df(a)) = f_{x_j}(a) = \frac{\partial f}{\partial x_j}(a)$$

6.4 Continuous Partials

Remark: We know that if $f : A \to \mathbb{R}^m$ for $A \subseteq \mathbb{R}^n$ is differentiable at a, then each $\frac{\partial f_i}{\partial x_j}(a)$ exists, but is the converse true?

Example: Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by

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

So we have $f_x(0,0) = \lim\limits_{h\to 0} \frac{f(h,0)-f(0,0)}{h} = 0$ and $f_y(0,0) = 0$ by symmetry. Hence $\nabla f(0,0) = 0$. However, f is not continuous at 0, hence it cannot be differentiable. To see this consider $(\frac{1}{n}, \frac{1}{n}) \to 0$ as $n \to \infty$. We have

$$f\left(\frac{1}{n}, \frac{1}{n}\right) = \frac{\left(\frac{1}{n^2}\right)}{\left(\frac{2}{n^2}\right)} = \frac{1}{2} \to \frac{1}{2}$$

hence $(\frac{1}{n}, \frac{1}{n}) \to (0,0)$ but $f(\frac{1}{n}, \frac{1}{n}) \not\to f(0,0)$.

Theorem: Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$. If $a \in U$ and for all $1 \leq j \leq n$ $\frac{\partial f}{\partial x_j}$ exists on U and is continuous at a, then f is differentiable at a.

Proof. Suppose $a = (a_1, \ldots, a_n)$. Since U is open, there is $r > 0$ such that $B_r(a) \subseteq U$. For any $h = (h_1, \ldots, h_n) \neq 0$ such that $a + h \in B_r(a)$, we have

$$\begin{aligned}
f(a+h) - f(a) &= f(a_1 + h_1, \ldots, a_n + h_n) - f(a_1, \ldots, a_n) \\
&= f(a_1 + h_1, \ldots, a_n + h_n) - f(a_1, a_2 + h_2, \ldots, a_n + h_n) \\
&\quad + f(a_1, a_2 + h_2, \ldots, a_n + h_n) - f(a_1, a_2, a_3 + h_3, \ldots, a_n + h_n) \\
&\quad + f(a_1, a_2, a_3 + h_3, \ldots, a_n + h_n) - f(a_1, a_2, a_3, a_4 + h_4, \ldots, a_n + h_n) \\
&\quad \vdots \\
&\quad + f(a_1, \ldots, a_{n-1}, a_n + h_n) - f(a_1, a_2, \ldots, a_n) \qquad (1)
\end{aligned}$$

However, by the single variable Mean Value Theorem, for every $1 \leq j \leq n$ there exists a c_j between a_j and $a_j + h_j$ such that

$$\begin{aligned}
&\frac{f(a_1, \ldots, a_{j-1}, a_j + h_j, \ldots, a_n + h_n) - f(a_1, \ldots, a_j, a_{j+1} + h_{j+1}, \ldots, a_n + h_n)}{a_j + h_j - a_j} \\
&= \frac{\partial f}{\partial x_j}(a_1, \ldots, a_{j-1}, c_j, a_{j+1} + h_{j+1}, \ldots, a_n + h_n) \qquad (2)
\end{aligned}$$

This means, by (1) and (2) we get

$$f(a+h) - f(a) = \sum_{j=1}^{n} h_j \frac{\partial f}{\partial x_j}(a_1, \ldots, a_{j-1}, c_j, a_{j+1} + h_{j+1}, \ldots, a_n + h_n)$$

Now for $1 \leq j \leq n$ let

$$\delta_j := \frac{\partial f}{\partial x_j}(a_1, \ldots, a_{j-1}, c_j, a_{j+1} + h_{j+1]}, \ldots, a_n + h_n) - \frac{\partial f}{\partial x_j}(a_1, \ldots, a_n)$$

and let $\delta = (\delta_1, \ldots, \delta_n)$. Then we have

$$f(a+h) - f(a) - \nabla f(a) \cdot h = h \cdot \delta$$

(Recall $\nabla f(a)$ is the vector of partials at a, i.e. the negative terms of δ.) Since all the partial derivatives are continuous at a, as $h \to 0$ each $\delta_j \to 0$ and so $\delta \to 0_{\mathbb{R}^n}$. Therefore,

$$0 \leq \lim_{h \to 0} \frac{|f(a+h) - f(a) - \nabla f(a) \cdot h|}{\|h\|} = \lim_{h \to 0} \frac{|\delta \cdot h|}{\|h\|} \leq \lim_{h \to 0} \frac{\|\delta\| \cdot \|h\|}{\|h\|} = \lim_{h \to 0} \delta = 0$$

where the second inequality holds from the Cauchy-Schwarz inequality. Therefore,

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

and so

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

By definition, this means f is differentiable at a. □

Example: Let $U = \mathbb{R}^2 \setminus \{(0,0)\}$. Consider $f : U \to \mathbb{R}$ given by

$$f(x,y) = \frac{\sin(xy)}{x^2+y^2}$$

We have then

$$f_x(x,y) = \frac{(x^2+y^2)\cos(xy)y - 2x\sin(xy)}{(x^2+y^2)^2}$$

exists on U and is continuous on U. By symmetry, $f_y(x,y)$ exists and is continuous on U as well. Thus by the above theorem we have f is differentiable on U.

Remark: The converse of the above theorem is not true. It is possible that a function is differentiable at a but has a partial derivative discontinuous at a.

Example: Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} (x^2+y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) & \text{if } (x,y) \neq 0 \\ 0 & \text{if } (x,y) = 0 \end{cases}$$

From a previous example we recall f is differentiable at $(0,0)$. Now observe

$$f_x(x,y) = 2x\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) - \cos\left(\frac{1}{\sqrt{x^2+y^2}}\right)\frac{x}{\sqrt{x^2+y^2}}$$

for all $(x,y) \neq (0,0)$. Now notice that $(\frac{1}{n}, 0) \to (0,0)$ but

$$f_x\left(\frac{1}{n}, 0\right) = \frac{2}{n}\sin(n) - \cos(n)$$

diverges. Therefore, f_x is not continuous at $(0,0)$, hence why the converse of the above theorem does not hold.

Week 7 More on Derivatives

7.1 Differentiation Rules

Theorem: Let $U \subseteq \mathbb{R}^n$ be open and let $a \in U$. Let $f : U \to \mathbb{R}^m$. There is a unique matrix $A \in M_{m\times n}(\mathbb{R})$ such that

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

and this matrix is exactly $Df(a)$. Notice this means f is differentiable at a if and only if $Df(a)$ exists and if an A prove f is differentiable at a then it is the total derivative of f at a.

Proof. Let $A \in M_{m\times n}(\mathbb{R})$ be such that $\lim_{h\to 0} \frac{f(a+h)-f(a)-Ah}{\|h\|} = 0$. So we know that f is differentiable at $a \in U$ and we know that $Df(a)$ also satisfies this limit. That is

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

Notice we have then for any $1 \le i \le n$

$$\begin{aligned}
\lim_{h\to 0} \frac{f(a+h)-f(a)-Ah}{\|h\|} - \frac{f(a+h)-f(a)-Df(a)h}{\|h\|} &= 0 \\
\lim_{h\to 0} \frac{(Df(a)h - A)h}{\|h\|} &= 0 \\
\lim_{t\to 0^+} \frac{(Df(a) - A)te_i}{\|te_i\|} &= 0 \\
\lim_{t\to 0^+} \frac{(Df(a) - A)te_i}{|t|} &= 0 \\
\lim_{t\to 0^+} (Df(a) - A)e_i &= 0 \\
\operatorname{Col}_i(Df(a)) &= \operatorname{Col}_i(A)
\end{aligned}$$

Therefore each column is equal, thus $Df(a) = A$, proving it is unique. □

Theorem. Sum and Scalar Multiplication Rules: Let $A \subseteq \mathbb{R}^n$ be open and $a \in A$. Let $f, g : A \to \mathbb{R}^m$ be differentiable at a. For all $\alpha \in \mathbb{R}$ $f + \alpha g$ is differentiable at a and

$$D(f+\alpha g)(a) = Df(a) + \alpha Dg(a)$$

Proof. Let $P = f + \alpha g$. Notice we have by our limit laws

$$\begin{aligned}
&\lim_{h\to 0} \frac{P(a+h)-P(a)-(Df(a)+\alpha Dg(a))h}{\|h\|} \\
&= \lim_{h\to 0} \frac{f(a+h)-f(a)-Df(a)h}{\|h\|} + \alpha\frac{g(a+h)-g(a)-Dg(a)}{\|h\|} \\
&= 0 + \alpha 0 = 0
\end{aligned}$$

□

Remark. Dot Product: Recall the dot product of two vectors $x, y \in \mathbb{R}^n$ is the scalar given by $x \cdot y = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}$.

Theorem. Dot Product Rule: Let $A \subseteq \mathbb{R}^n$ be open and $a \in A$. Let $f, g \to \mathbb{R}^m$ be differentiable at a. Define the dot product of f, g to be $f \cdot g : A \to \mathbb{R}$ defined by $(f \cdot g)(x) = f(x) \cdot g(x)$. Then $f \cdot g$ is differentiable at a and

$$D(f \cdot g)(a) = g(a)Df(a) + f(a)Dg(a)$$

We consider $f(a)$ and $g(a)$ as row vectors in this case.

Proof. Notice we have

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

So we have

$$\begin{aligned}
&\lim_{h\to 0}\frac{(f\cdot g)(a+h)-(f\cdot g)(a)-g(a)Df(a)h-f(a)Dg(a)h}{\|h\|}\\
&=g(a)\cdot\lim_{h\to 0}\frac{f(a+h)-f(a)-Df(a)h}{\|h\|}+f(a)\cdot\lim_{h\to 0}\frac{g(a+h)-g(a)-Dg(a)h}{\|h\|}\\
&\quad+\lim_{h\to 0}\frac{f(a)\cdot g(a)-g(a)\cdot f(a+h)-f(a)\cdot g(a+h)+f(a+h)\cdot g(a+h)}{\|h\|}\\
&=0+\lim_{h\to 0}\frac{g(a)\cdot(f(a)-f(a+h))-g(a+h)\cdot(f(a)-f(a+h))}{\|h\|}\\
&=\lim_{h\to 0}\frac{(g(a)-g(a+h))\cdot(f(a)-f(a+h))}{\|h\|}
\end{aligned}$$

However, by the Cauchy-Schwarz inequality,

$$\frac{|g(a)-g(a+h))\cdot(f(a)-f(a+h))|}{\|h\|}\le\frac{\|g(a)-g(a+h)\|\cdot\|f(a)-f(a+h)\|}{\|h\|}$$

Therefore

$$\begin{aligned}
0&\le\lim_{h\to 0}\frac{|(f\cdot g)(a+h)-(f\cdot g)(a)-Xh|}{\|h\|}\\
&\le\lim_{h\to 0}\frac{\|g(a)-g(a+h)\|\cdot\|f(a)-f(a+h)\|}{\|h\|}\\
&=\lim_{h\to 0}\frac{\|g(a)-g(a+h)\|}{\|h\|}\cdot\frac{\|f(a)-f(a+h)\|}{\|h\|}\cdot\|h\|\\
&=\lim_{h\to 0}\frac{\|g(a)-g(a+h)-Dg(a)h+Dg(a)h\|}{\|h\|}\cdot\frac{\|f(a)-f(a+h)-Df(a)h+Df(a)h\|}{\|h\|}\cdot\|h\|\\
&\le\lim_{h\to 0}\frac{\|g(a)-g(a+h)-Dg(a)h\|+\|Dg(a)h\|}{\|h\|}\cdot\frac{\|f(a)-f(a+h)-Df(a)h\|+\|Df(a)h\|}{\|h\|}\cdot\|h\|\\
&=\lim_{h\to 0}\left(\frac{\|g(a)-g(a+h)-Dg(a)\|}{\|h\|}+\frac{\|Dg(a)h\|}{\|h\|}\right)\cdot\left(\frac{\|f(a)-f(a+h)-Df(a)\|}{\|h\|}+\frac{\|Df(a)h\|}{\|h\|}\right)\cdot\|h\|\\
&=\lim_{h\to 0}\frac{\|Dg(a)h\|}{\|h\|}\cdot\frac{\|Df(a)h\|}{\|h\|}\cdot\|h\|\\
&\le\lim_{h\to 0}\frac{\|Dg(a)\|_{op}\cdot\|h\|}{\|h\|}\cdot\frac{\|Df(a)\|_{op}\cdot\|h\|}{\|h\|}\cdot\|h\|\\
&=\lim_{h\to 0}\|Dg(a)\|_{op}\cdot\|Df(a)\|_{op}\cdot\|h\|\\
&=(\|Dg(a)\|_{op}\cdot\|Df(a)\|_{op})\cdot\lim_{h\to 0}\|h\|\\
&=0
\end{aligned}$$

As desired. □

7.2 Chain Rule

Theorem. Chain Rule: Let $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ be open. Let $f : A \to \mathbb{R}^m$ and $g : B \to \mathbb{R}^k$ with $f(A) \subseteq B$. If f is differentiable at $a \in A$ and g is differentiable at $f(a) \in B$, then $(g \circ f)$ is differentiable at a with

$$D(g \circ f)(a) = Dg(f(a)) \cdot Df(a)$$

Proof. We wish to show

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

where $X = Dg(f(a))Df(a)$. Let $b = f(a)$. Let

$$\begin{aligned} \epsilon(h) &= f(a+h) - f(a) - Df(a)h \\ \delta(k) &= g(b+k) - g(b) - Dg(b)k \end{aligned}$$

so that

$$\lim_{h \to 0} \frac{\epsilon(h)}{\|h\|} = 0 \qquad \text{and} \qquad \lim_{k \to 0} \frac{\delta(k)}{\|k\|} = 0$$

Let $k = f(a+h) - f(a)$. Note that $k \to 0$ as $h \to 0$ by the continuity of f at a. Then,

$$\begin{aligned} &\lim_{h \to 0} \frac{(g \circ f)(a+h) - (g \circ f)(a) - Xh}{\|h\|} \\ &= \lim_{h \to 0} \frac{g(f(a+h)) - g(f(a)) - Dg(f(a))Df(a)h}{\|h\|} \\ &= \lim_{h \to 0} \frac{g(f(a+h) - f(a) + f(a)) - g(b) - Dg(b)Df(a)h}{\|h\|} \\ &= \lim_{h \to 0} \frac{g(k+b) - g(b) - Dg(b)Df(a)h}{\|h\|} \\ &= \lim_{h \to 0} \frac{g(k+b) - g(b) - Dg(b)(-\epsilon(h) + f(a+h) - f(a))}{\|h\|} \\ &= \lim_{h \to 0} \frac{Dg(b)\epsilon(h) + g(k+b) - g(b) - Dg(b)(f(a+h) - f(a))}{\|h\|} \\ &= \lim_{h \to 0} Dg(b)\frac{\epsilon(h)}{\|h\|} + \frac{\delta(k)}{\|h\|} \end{aligned}$$

Now since

$$0 \leq \frac{\|Dg(b)\epsilon(h)\|}{\|h\|} \leq \|Dg(b)\|_{op} \frac{\|\epsilon(h)\|}{\|h\|} \to 0$$

as $h \to 0$, we see that $\lim_{h \to 0} Dg(b)\frac{\epsilon(h)}{\|h\|} = 0$. Now,

$$\lim_{h \to 0} \frac{\delta(k)}{\|h\|} = \lim_{h \to 0} \frac{\delta(k)}{\|k\|} \cdot \frac{\|k\|}{\|h\|}$$

However,

$$\|k\| = \|Df(a)h + \epsilon(h)\| \leq \|Df(a)\|_{op}\|h\| + \|\epsilon(h)\|$$

Therefore,

$$\frac{\|k\|}{\|h\|} \leq \frac{\|Df(a)\|_{op}\|h\| + \|\epsilon(h)\|}{\|h\|} \to \|Df(a)\|_{op}$$

as $h \to 0$. Hence, $\frac{\|k\|}{\|h\|}$ is bounded. Since $\frac{\delta(k)}{\|k\|} \to 0$ as $h \to 0$, we have

$$\lim_{h \to 0} \frac{\delta(k)}{\|h\|} = \lim_{h \to 0} \frac{\delta(k)}{\|k\|} \cdot \frac{\|k\|}{\|h\|} = 0$$

as desired. □

7.3 Mean Value Theorem

Remark. Naive MVT: Recall the single variable mean value theorem is

> *Let $f : [a, b] \to \mathbb{R}$ be continuous. If f is differentiable on (a, b), then there is a $c \in (a, b)$ such that*
>
> $$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Hence, we might conjecture we can extend the mean value theorem as

> *Let $U \subseteq \mathbb{R}^n$ be open. Let $f : U \to \mathbb{R}^m$ be differentiable. If $a, b \in U$ then there is a $c \in L(a, b)$ such that*
>
> $$f(b) - f(a) = Df(c)(b - a)$$
>
> *where $L(a, b) := \{(1 - t)a + tb : t \in [0, 1]\}$ is the line through a and b.*

However, our conjecture is disproved by the counterexample $f : \mathbb{R} \to \mathbb{R}^2$ given by $f(x) = (\cos x, \sin x)$. So $f(0) = f(2\pi) = (1, 0)$, but

$$Df(x) = \begin{bmatrix} -\sin x \\ \cos x \end{bmatrix} \neq 0$$

for all $x \in \mathbb{R}$. Hence we wish to work with only one direction at a time. I.e. for all $x \in \mathbb{R}^m$ find a $c \in L(a, b)$ such that

$$x \cdot (f(b) - f(a) - Df(c)(b - a)) = 0 \iff x \cdot (f(b) - f(a)) = x \cdot (Df(a)(b - a))$$

Lemma: Let $a, b \in \mathbb{R}^n$. The function $\varphi : \mathbb{R} \to \mathbb{R}^n$ given by $\varphi(t) = (1 - t)a + tb$ is differentiable with $D\varphi(t) = b - a$.

Proof.

$$\begin{aligned}\lim_{h\to 0}\frac{\varphi(t+h)-\varphi(t)-(b-a)h}{\|h\|} &= \lim_{h\to 0}\frac{(1-t-h)a+(t+h)b-(1-t)a-tb-(b-a)h}{\|h\|}\\ &= \lim_{h\to 0}\frac{-ha+hb(b-a)h}{\|h\|}\\ &= 0\end{aligned}$$

□

Theorem. Mean Value Theorem: Let $U \subseteq \mathbb{R}^n$ be open. If $f : U \to \mathbb{R}^m$ is differentiable and $a, b \in U$ such that $L(a, b) \subseteq U$, then for all $x \in \mathbb{R}^m$ there exists $c \in L(a, b)$ such that

$$x \cdot (f(b) - f(a)) = x \cdot (Df(c)(b - a))$$

Proof. Let $a, b \in U$ such that $L(a, b) \subseteq U$. Let $x \in \mathbb{R}^m$ be fixed. (1) Let $\varphi(t) = (1-t)a+t(b)$. Since $\varphi([0, 1]) = L(a, b) \subseteq U$ is open, there is a $\delta > 0$ such that $\varphi((0 - \delta, 1 + \delta)) \subseteq U$. (2) For $t \in (-\delta, 1+\delta)$, $D(f \circ \varphi)(t) = Df(\varphi(t))(b-a)$ by the chain rule and the previous lemma.

Let $F : (-\delta, 1 + \delta) \to \mathbb{R}$ be given by $F(t) = x \cdot (f \circ \varphi)(t)$. By the dot product rule, $F'(t) = x \cdot Df(\varphi(t))(b-a)$ (recall x is a constant vector). Hence, by the single variable mean value theorem, we have that there is a t_0 such that

$$\begin{aligned}F(1) - F(0) &= F'(t_0)(1 - 0)\\ x \cdot (f \circ \varphi)(1) - x \cdot (f \circ \varphi)(0) &= x \cdot Df(\varphi(t_0))(b - a)\\ x \cdot (f(b) - f(a)) &= x \cdot Df(\varphi(t_0))(b - a)\end{aligned}$$

Thus let $c = \varphi(t_0) \in L(a, b)$ and we have the result. □

7.4 Tangent Hyperplanes

Remark. Motivation: Notice that in the $n = 1$ case, if $f : U \to \mathbb{R}$ (for open $U \subseteq \mathbb{R}^1$) is differentiable at $a \in U$, then $f'(a)$" = "$\nabla f(a)$ is the slope of the tangent line to the curve $y = f(x)$ at $x = a$. We want to try and derive information from $\nabla f(a)$ about the tangent hyperplane in higher dimensions.

For instance, the information we can determine for $f : U \to \mathbb{R}$ where $U \subseteq \mathbb{R}^2$ is open. We want to derive information about the tangent plane to the surface $z = f(x, y)$ at $(x, y) = a \in U$.

Definition. Hyperplane: A hyperplane in $\mathbb{R}^n$ is a subspace of the form

$$P = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : a_1x_1 + a_2x_2 + \cdots + a_nx_n = d\}$$

Where $a_1, \ldots, a_n \in \mathbb{R}$ are fixed and not all zero and $d \in \mathbb{R}$. Notice a hyperplane in $\mathbb{R}^2$ is a line and in $\mathbb{R}^3$ it is a plane.

Definition. Normal (vector): Let $P = \{x_1, \dots, x_n) : a_1x_1 + \cdots + a_nx_n = d\}$ be a hyperplane in $\mathbb{R}^n$. We call $\vec{n} = (a_1, \dots, a_n)$ the normal or normal vector of P.

Remark. Hyperplanes Geometrically: Let $P = \{x_1, \dots, x_n) : a_1x_1 + \cdots + a_nx_n = d\}$ be a hyperplane. Let $b = (b_1, \dots, b_n) \in P$. So we have $d = a_1b_1 + \cdots + a_nb_n$. Now let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. So

$$\begin{aligned} x \in P &\iff d = a_1x_1 + \cdots + a_nx_n \\ &\iff 0 = a_1x_1 + \cdots + a_nx_n - b \\ &\iff 0 = a_1(x_1 - b_1) + \cdots + a_n(x_n - b_n) \\ &\iff 0 = n \cdot (x - b) && \text{dot product} \end{aligned}$$

Therefore, $P = \{x \in \mathbb{R}^n : n \cdot (x - b) = 0\}$. I.e. $x \in P$ if and only if n is orthogonal/perpendicular to $x - b$.

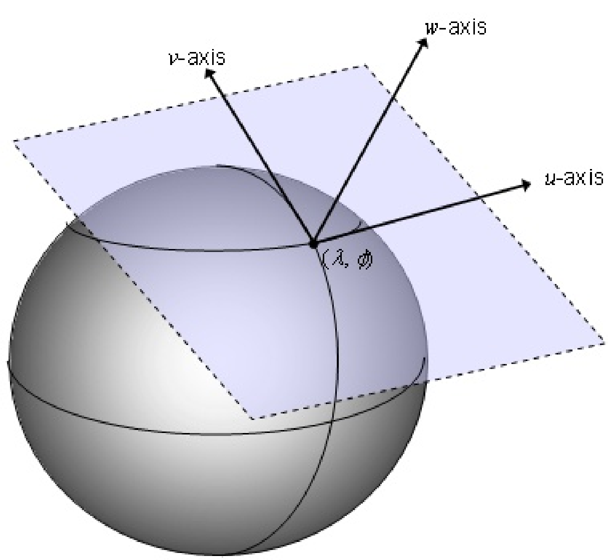
Definition. Tangent Hyperplane: Let $A \subseteq \mathbb{R}^n$ and $a \in A$. A hyperplane $a \in P \subseteq \mathbb{R}^n$ with normal n is said to be tangent to A at a if

$$n \cdot \frac{a_k - a}{\|a_k - a\|} \to 0$$

for all sequences $(a_k) \subseteq A \setminus \{a\}$ such that $a_k \to a$.

Remark. Intuition of Tangent Hyperplanes: Recall that $a, b \in \mathbb{R}^n$ are orthogonal if $a \cdot b = 0$. So $n \cdot \frac{a_j - a}{\|a_j - a\|} \to 0$ says unit vectors in the direction of $a_k - a$ become closer and closer to being orthogonal to n as $k \to \infty$.

Example: The following is a visual representation of a tangent hyperplane tangent to a sphere



Theorem: Let $U \subseteq \mathbb{R}^n$ be open and $a \in U$. Let $f : U \to \mathbb{R}$. If f is differentiable at a, then the surface

$$S = \{(x, z) \in \mathbb{R}^{n+1} : z = f(x), x \in U\}$$

has a tangent hyperplane at $(a, f(a))$ with normal $n = (\nabla f(a), -1)$.

Proof. □

Example: Find the tangent plane, P, to the surface $z = 2x^2 + y^2$ at $(1, 1, 3)$. Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = 2x^2 + y^2$. Notice f_x and f_y exist and are continuous on $\mathbb{R}^2$ thus f is differentiable on $\mathbb{R}^2$. Now consider $\nabla f(x, y) = (4x, 2y)$, in particular $\nabla f(1, 1) = (4, 2)$. So we have the normal of the tangent plane is $n = (4, 2, -1)$. Notice we know that the points of P are such that $4x + 2y - z = d$ and we already know $(1, 1, 3) \in P$. Therefore, $d = 4 \cdot 1 + 2 \cdot 1 - 3 = 3$. Thus

$$P = \{(x, y, z) : 4x + 2y - z = 3, x, y, z \in \mathbb{R}\}$$

Week 8 Applications of Differentiation

8.1 Higher Order Derivatives

Definition. Higher Order Partial: Higher order partial derivatives are defined recursively by

$$\frac{\partial^k f}{\partial x_1 \partial x_2 \cdots \partial x_k} := \frac{\partial}{\partial x_1}\left(\frac{\partial^{k-1} f}{\partial x_2 \cdots \partial x_k}\right),$$

if it exists. We call k the order of the partial derivative. We also use the notation

$$f_{x_k x_{k-1} \cdots x_1} = \frac{\partial^k f}{\partial x_1 \partial x_2 \cdots \partial x_k}.$$

Also note that I am not assuming the x_i's are distinct here.

Definition. $C^k(U, \mathbb{R}^m)$: Let $f : U \to \mathbb{R}^m$ be a function on an open set $U \subseteq \mathbb{R}^n$. We say $f \in C^k(U, \mathbb{R}^m)$ if all partial derivatives of f of order less than or equal to k exist on U and are continuous on U. If $m = 1$ we write $C^k(U, \mathbb{R}) = C^k(U)$.

Remark: For $f : U \to \mathbb{R}$ differentiable at $a \in U$, we may think of $\nabla f(a)$ as a function from $\mathbb{R}^n$ to $\mathbb{R}$ by

$$\nabla f(a)(h_1, \ldots, h_n) = \frac{\partial f}{\partial x_1}(a)h_1 + \cdots + \frac{\partial f}{\partial x_n}(a)h_n = \nabla f(a) \cdot (h_1, \ldots, h_n)$$

Definition. Higher Order Total Derivative: Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$. Let $k \in \mathbb{N}$. Assume all partial derivatives of order less than or equal to k exist at $a \in U$. We define the kth order total derivative of f at a by $D^k f(a) : \mathbb{R}^n \to \mathbb{R}$ by

$$D^k f(a)(h_1, \ldots, h_n) = \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} \frac{\partial^k}{\partial x_{i_1} \cdots \partial x_{i_k}}(a) h_{i_1} h_{i_2} \cdots h_{i_k}$$

Example: Let $f : \mathbb{R}^2 \to \mathbb{R}$. Then

$$D^2 f(a)(h_1, h_2) = f_{xy}(a)h_1^2 + f_{xy}(a)h_1 h_2 + f_{yx}(a)h_2 h_1 + f_{yy}(a)h_2^2$$

8.2 Taylor's Theorem

Theorem. Taylor's Theorem: Let $p \in \mathbb{N}$. Let $U \subseteq \mathbb{R}^n$ and $f \in C^p(U)$. For all $x, a \in U$ with $L(x,a) \subseteq U$, there exists $c \in L(x,a)$ such that

$$f(x) = f(a) + \sum_{k=1}^{p-1} \frac{1}{k!} D^k f(a)(x-a) + \frac{1}{p!} D^p f(c)(x-a)$$

Proof. Let $x, a \in U$ and let $h = x - a = (h_1, \ldots, h_n)$. Since $L(x,a) \subseteq U$ and U is open, there is a $\delta > 0$ such that $a + th \in U$ for all $t \in I := (-\delta, 1+\delta)$. Now, the function $g : I \to \mathbb{R}$ given by $g(t) = f(a+th)$ is differentiable and by the chain rule

$$g'(t) = Df(a+th)h = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a+th)h_i$$

Moreover, it may be shown by induction that for $1 \leq j \leq p$,

$$g^{(j)}(t) = \sum_{i_1=1}^{n} \cdots \sum_{i_j=1}^{n} \frac{\partial^j f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_j}}(a+th)h_{i_1} \cdots h_{i_j}$$

Note that this exactly a higher-order total derivative. Hence we have for all $1 \leq j \leq p-1$

$$g^{(j)}(0) = D^j f(a)(h)$$

and

$$g^{(p)}(t) = D^p f(a+th)(h)$$

Therefore, $g : I \to \mathbb{R}$ is p-times differentiable and in particular we have $g : \mathbb{R} \to \mathbb{R}$ thus by the single variable Taylor's Theorem

$$g(1) - g(0) = \sum_{k=1}^{p-1} \frac{1}{k!} g^{(k)}(0) + \frac{1}{p!} g^{(p)}(t)$$

for some $0 \leq t \leq 1$. Thus,

$$f(x) - f(a) = f(a+h) - f(a) = g(1) - g(0) = \sum_{k=1}^{p-1} \frac{1}{k!} D^k f(a)(h) + \frac{1}{p!} D^p f(a+th)(h)$$

and so taking $c = a + th$, the result is achieved. □

8.3 Optimization

Definition. Local Max (resp. Local Min): Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$. Let $a \in U$. We say $f(a)$ is local max (resp. local min) of f if there is an $r > 0$ such that $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$) for all $x \in B_r(a)$. We say $f(a)$ is local extremum of f if it is a local min or max.

Remark: Suppose $f : U \to \mathbb{R}$ is differentiable at $a \in U$ and that $f(a)$ is a local extremum of f. Suppose $a = (a_1, \ldots, a_n)$. Then $g_i(t) = f(a_1, \ldots, a_{i-1}, t, a_{i+1}, \ldots, a_n)$ has a local extremum at $t = a_i$. Hence, $g_i'(a_i) = \dfrac{\partial f}{\partial x_i}(a) = 0$ for all $1 \leq i \leq n$. Therefore, $\nabla f(a) = 0$.

Definition. Saddle Point: Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$. Let $a \in U$. If $\nabla f(a) = 0$ but $f(a)$ is not a local extremum of f, we say a is a saddle point of f.

Example: Consider $f(x, y) = x^2 - y^2$. So $\nabla f(x, y) = (2x, 2y)$ and so $\nabla f(0, 0) = 0$, however f has no local extrema, hence $(0, 0)$ is a saddle point of f.

Lemma: Let $U \subseteq \mathbb{R}^n$ be open and let $f \in C^2(U)$. If $a \in U$ such that $D^2 f(a)(h) > 0$ for all $0 \neq h \in \mathbb{R}^n$, then there is an $m > 0$ such that

$$D^2 f(a)(x) \geq m\|x\|^2$$

for all $x \in \mathbb{R}^n$.

Proof. Consider the compact set $K = \{x \in \mathbb{R}^n : \|x\| = 1\}$ (Heine-Borel). Since $f \in C^2(U)$, we have that $D^2 f(a)$ is continuous and positive on K. By the EVT, there is an $m > 0$ such that $m = \min\{D^2 f(a)(x) : x \in K\}$. For $0 \neq x \in \mathbb{R}^n$, we then see that $\frac{x}{\|x\|} \in K$ and so

$$D^2 f(a)\left(\frac{x}{\|x\|}\right) = \frac{1}{\|x\|^2} D^2 f(a)(x) \geq m$$

Note that we extract $\frac{1}{\|x\|}$ as a square because we take the second derivative. I.e. we are summing across the product of two components of x, each divided by $\|x\|$, hence we square. □

Lemma: Let $U \subseteq \mathbb{R}^n$ be open and let $f \in C^2(U)$. Suppose $a \in U$ is such that $\nabla f(a) = 0$. Let $r > 0$ be such that $B_r(a) \subseteq U$. There is a function $\epsilon : B_r(0) \to \mathbb{R}$ such that

$$\lim_{h \to 0} \epsilon(h) = 0$$

and

$$f(a + h) - f(a) = \frac{1}{2} D^2 f(a)(h) + \|h\|^2 \epsilon(h)$$

for $\|h\|$ sufficiently small.

Proof. Consider

$$\epsilon(h) := \frac{f(a + h) - f(a) - \frac{1}{2} D^2 f(a)(h)}{\|h\|^2}$$

for $0 \neq h \in B_r(0)$ and define $\epsilon(0) := 0$. We must show $\epsilon(h) \to 0$ as $h \to 0$. Let $h \in B_r(0)$. Since $\nabla f(a) = 0$, by Taylor's Theorem we have that

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

for some $c \in L(a, a+h)$. So

$$\begin{aligned} 0 \leq |\epsilon(h)| \cdot \|h\| &= \left|\frac{1}{2}D^2 f(c)(h) - \frac{1}{2}D^2 f(a)(h)\right| \\ &\leq \sum_i \sum_j \left|\frac{\partial^2 f}{\partial x_i \partial x_j}(c) - \frac{\partial^2 f}{\partial x_i \partial x_j}(a)\right| |h_i h_j| \\ &\leq \sum_i \sum_j \left|\frac{\partial^2 f}{\partial x_i \partial x_j}(c) - \frac{\partial^2 f}{\partial x_i \partial x_j}(a)\right| \|h\| \end{aligned}$$

and

$$\frac{1}{2}\sum_i \sum_j \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(c) - \frac{\partial^2 f}{\partial x_i \partial x_j}(a)\right) \to 0$$

as $h \to 0$ because $c \to a$ as $h \to 0$ and $f \in C^2(U)$. □

Theorem. Second Derivative Test: Let $U \subseteq \mathbb{R}^n$ be open, let $f \in C^2(U)$. Let $a \in U$. If $\nabla f(a) = 0$, then:

1. If $\forall h \neq 0$, $D^2 f(a)(h) > 0$, then $f(a)$ is a local min.
2. If $\forall h \neq 0$, $D^2 f(a)(h) < 0$, then $f(a)$ is a local max.
3. If $\exists h, k \in \mathbb{R}^n$ such that $D^2 f(a)(h) > 0$ and $D^2 f(a)(k) < 0$, then a is a saddle point.

Proof. Let $r > 0$ such that $B_r(a) \subseteq U$. There is a function $\epsilon : B_r(0) \to \mathbb{R}$ such that

$$\lim_{h \to 0} \epsilon(h) = 0$$

and

$$f(a+h) - f(a) = \frac{1}{2}D^2 f(a)(h) + \|h\|^2 \epsilon(h)$$

for $\|h\|$ sufficiently small.

1. Suppose $D^2 f(a)(h) > 0$ for all $0 \neq h \in \mathbb{R}^n$. Let $m > 0$ be such that

   $$D^2 f(a)(x) \geq m\|x\|^2,$$

   for all $x \in \mathbb{R}^n$. Then,

   $$f(a+h) - f(a) = \frac{1}{2}D^2 f(a)(h) + \|h\|^2 \epsilon(h) \geq \left(\frac{m}{2} + \epsilon(h)\right)\|h\|^2 > 0$$

   for all $\|h\|$ sufficiently small, since $m > 0$ and $\epsilon(h) \to 0$ as $h \to 0$. Therefore $f(a+h) > f(a)$ for all $\|h\|$ sufficiently small, and so $f(a)$ is a local minimum of f.
2. Follow from (1) by replacing f with $-f$.

3. Let $h \in \mathbb{R}^n$. For small $t \in \mathbb{R}$,

$$\begin{aligned} f(a+th) - f(a) &= \frac{1}{2} D^2 f(a)(th) + \|th\|^2 \epsilon(th) \\ &= t^2 \left(\frac{1}{2} D^2 f(a)(h) + \|h\|^2 \epsilon(th) \right). \end{aligned}$$

Letting $t \to 0$, we have that $\epsilon(th) \to 0$ and so $f(a+th) - f(a)$ takes the same sign as $D^2 f(a)(h)$, which can be both positive and negative. Therefore a is a saddle point.

□

Lemma: Let $a, b, c \in \mathbb{R}$ and let $D := b^2 - ac$. Consider $\varphi : \mathbb{R}^2 \to \mathbb{R}$ given by $\varphi(h,k) = ah^2 + 2bhk + ck^2$. (1) If $D < 0$ then a and $\varphi(h,k)$ have the same sign for all $h, k \neq 0$. (2) If $D > 0$ then $\varphi(h,k)$ takes on positive and negative values on its domain.

Theorem: Let $U \subseteq \mathbb{R}^2$ be open. Let $f \in C^2(U)$ and let $\nabla f(a,b) = 0$. Let $D := f_{xy}(a,b)^2 - f_{xx}(a,b) f_{yy}(a,b)$ be the discriminant. (1) If $D < 0$ and $f_{xx}(a,b) > 0$, then $f(a,b)$ is a local min. (2) If $D < 0$ and $f_{xx}(a,b) < 0$, then $f(a,b)$ is a local max. (3) If $D > 0$ then (a,b) is a saddle point.

Proof. Follows from second derivative test along with the previous lemma for $a = f_{xx}(a,b)$, $b = f_{xy}(a,b)$, $c = f_{yy}(a,b)$ and $\varphi(h,k) = D^2 f(a,b)(h,k)$. □

8.4 Examples

Example: Classify all local extrema and/or saddle points of $f(x,y) = x^4 - y^4 - 4xy + 2$.

Notice $\nabla f(x,y) = (4x^3 - 4y, 4y^3 - 4x)$ which is zero iff $x^3 = y$ and $y^3 = x$, implying $x = y = 0$, $x = y = 1$, or $x = y = -1$ (these are our "critical points"). Now notice taking derivatives of our gradient we have $f_{xx}(x,y) = 12x^2$, $f_{yy}(x,y) = 12y^2$, $f_{xy} = -4$.

At $(x,y) = (0,0)$, we have $D = 16 - 0 \cdot 0 > 0$ (discriminant), meaning $(0,0)$ is a saddle point. At $(x,y) = (1,1)$, we have $D = 16 - 12 \cdot 12 < 0$, moreover $f_{xx}(1,1) = 12 > 0$ hence $(1,1)$ is a local minimum. At $(x,y) = (-1,-1)$, we have $D = 16 - 12 \cdot 12 < 0$, moreover $f_{xx}(1,1) = 12 > 0$ hence $(-1,-1)$ is a local minimum.

Example: Let $K = \overline{B_1(0,0)}$. Find the absolute max and min of $f : K \to \mathbb{R}$ given by $f(x,y) = 2x^3 + y^4$.

Notice $\nabla f(x,y) = (6x^3, 4y^3)$ which is zero iff $6x^3 = 0$ and $4y^3 = 0$. Hence $(0,0)$ is the only critical point, where we achieve $f(0,0) = 0$. We now consider the boundary $\partial(K) = \{(x,y) : x^2 + y^2 = 1\}$.

Notice on the boundary, we have $f(x,y) = 2x^3 + (1-x^2)^2 = \underbrace{x^4 + 2x^3 - 2x^2 + 1}_{g(x)}$, where we denote $g(x)$ to be the expanded single variable representation of $f(x,y)$. Maximizing $g(x) : [-1,1] \to \mathbb{R}$, we have $g'(x) = 4x^3 + 6x^2 - 4x = 2x(2x^2 + 3x - 2) = 2x(2x-1)(x+2)$.

This is zero where $x = 0, \frac{1}{2}, -2$ (note $x = -2$ is illegal). Notice $g(0) = 1$, $g(1) = \frac{13}{16}$, $g(1) = 2$, $g(-1) = -2$ and $f(0,0) = 0$, so the absolute minimum is at $(x, y) = (-1, 0)$ with $f(x, y) = -2$ and the absolute maximum is at $(x, y) = (1, 0)$ with $f(x, y) = 2$.

Week 9 Local Theorems

9.1 Inverse Function Theorem

Remark: Recall the definition of the single variable inverse function theorem. For $I = (a, b)$, let $f : I \to \mathbb{R}$ is continuous and injective and let $y \in f(I)$. Suppose y is such that f is differentiable at $x = f^{-1}(y) \in I$ and $f'(x) \neq 0$. Then f^{-1} is differentiable at y and $(f^{-1})'(y) = \frac{1}{f'(x)}$. To generalize to multiple variables, we need something more like a matrix inverse given the derivative is a matrix.

Definition. Jacobian: Let $U \subseteq \mathbb{R}^n$ and let $f : U \to \mathbb{R}^n$. We defined the Jacobian of f at $a \in U$ by $Jf(a) := \det(Df(a))$

Lemma: Let $U \subseteq \mathbb{R}^n$ be open. Suppose $a \in U$ is such that we may find $r > 0$ such that $\overline{B_r(a)} \subseteq U$. Let $f : U \to \mathbb{R}^n$ be continuous and injective when restricted to $\overline{B_r(a)}$ and assume its first order partials exist on $B_r(a)$. If $Jf \neq 0$ on $B_r(a)$ then there exists $\epsilon > 0$ such that $B_\epsilon(f(a)) \subseteq f(B_r(a))$.

Proof. Consider $g : \overline{B_r(a)} \to \mathbb{R}$ given by $g(x) = \|f(x) - f(a)\|$. Since f is continuous and injective on $\overline{B_r(a)}$, we have that g is continuous and $g(x) > 0$ for all $x \neq a$. By the EVT,

$$m = \inf\{g(x) : \|x - a\| = r\| > 0$$

Take $\epsilon = \frac{m}{2}$. We claim that $B_\epsilon(f(a)) \subseteq f(B_r(a))$. Let $y \in B_\epsilon(f(a))$. Again by the EVT, there exists $b \in B_r(a)$ such that

$$\|f(b) - y\| = \inf\{\|f(x) - y\| : x \in \overline{B_r(a)}\}$$

For the sake of contradiction suppose that $\|b - a\| = r$. Then,

$$\epsilon > \|f(a)-y\| \geq \|f(b)-y\| \geq \|f(b)-f(a)\|-\|f(a)-y\| = g(b)-\|f(a)-y\| \geq m-\epsilon = 2\epsilon-\epsilon = \epsilon$$

which is a contradiction. Therefore we have that $b \in B_r(a)$.

If we can show that $y = f(b)$, we are done. This is where the information about the partial derivatives and the Jacobian come into play. Consider the continuous function $h : \overline{B_r(a)} \to \mathbb{R}$ given by $h(x) = \|f(x)-y\|$. By construction, $h(b)$ is the minimum value of b. Moreover, $h^2(b)$ is also the minimum value of h^2. Since $b \in B_r(a)$, which is open, we have that $\nabla h^2(b) = 0$ (derivative of h^2, note that in last week's proof we really just needed partials to exist at a, not necessarily be differentiable at a). However,

$$h^2(x) = \sum_{i=1}^{n} (f_i(x) - y_i)^2$$

and so for ever $1 \leq j \leq n$,

$$0 = \frac{\partial h^2}{\partial x_j}(b) = \sum_{i=1}^{n} 2(f_i(b) - y_i)\frac{\partial f_i}{\partial x_j}(b)$$

Thus, $Df(b)x = 0$, where $x = (2(f_1(b) - y_1), \ldots, 2(f_n(b) - y_n))^T \in \mathbb{R}^n$. Since $Df(b)$ is invertible ($Jf(b) \neq 0$ means non-zero determinant) we have that $x = 0$. Hence $f(b) = y$, as desired. □

Lemma: Let $U \subseteq \mathbb{R}^n$ be open and nonempty. If $f : U \to \mathbb{R}^n$ is continuous, injective, has all first-order existing on U, and is such that $Jf \neq 0$ on U, then f^{-1} is continuous on $f(U)$.

Proof. To prove that $f^{-1} : f(U) \to \mathbb{R}^n$ is continuous, it suffices to prove that $f(W)$ is open whenever W is open in $\mathbb{R}^n$ and $W \subseteq U$. This is since in this case W is relatively open in U and U is open, so W is open. So whenever W is open subset of U then $(f^{-1})^{-1}(W) = f(W)$ is open. By a result in 4.2 f^{-1} is continuous.

Now let $W \subseteq U$ be open and let $b \in f(W)$ such that $b = f(a)$ for some $a \in W$. Since W is open, there is an $r > 0$ such that $\overline{B_r(a)} \subseteq W$. By the previous lemma, there is an $\epsilon > 0$ such that

$$B_\epsilon(b) \subseteq f(B_r(a))$$

Thus, $B_\epsilon(b) \subseteq f(W)$, and so since b was arbitrary $f(W)$ is open. □

Lemma: Let $U \subseteq \mathbb{R}^n$ be open and let $f \in C^1(U, \mathbb{R}^n)$. If $a \in U$ is such that $Jf(a) \neq 0$, then there is an $r > 0$ such that $B_r(a) \subseteq U$, f is injective on $B_r(a)$, $Jf \neq 0$ on $B_r(a)$ and

$$\det\left(\left[\frac{\partial f_i}{\partial x_j}(c_i)\right]_{n\times n}\right) \neq 0$$

for all $c_1, \ldots, c_n \in B_r(a)$.

Proof. Let $W = U \times U \times \cdots \times U$ n times. Consider $h : W \to \mathbb{R}$ given by

$$h(x_1, \ldots, x_n) = \det\left(\left[\frac{\partial f_i}{\partial x_j}(x_i)\right]_{n\times n}\right)$$

Since $f \in C^1(U, \mathbb{R}^n)$ and a determinant is a polynomial in its entries, we have that h is continuous. Note that $h(a, a, \ldots, a) = Jf(a) \neq 0$. Thus we may find an open interval $h(a, a, \ldots, a) \in I \subseteq \mathbb{R}$ such that $0 \notin I$. Then by the continuity of h, $h^{-1}(I)$ is open (note that W is open, hence relatively open in W is open). And so, there is an $R > 0$ such that $B_R(a, a, \ldots, a) \subseteq h^{-1}(I)$. But then we may find $r > 0$ such that

$$B_r(a) \times \cdots \times B_r(a) \subseteq B_r(a, \ldots, a) \subseteq h^{-1}(I)$$

We see then that $Jf \neq 0$ on $B_r(a)$ and

$$\det\left(\left[\frac{\partial f_i}{\partial x_j}(c_i)\right]_{n\times n}\right) \neq 0$$

for all $c_1, \ldots, c_n \in B_r(a)$.

It remains to show that f is injective on $B_r(a)$. By way of contradiction, suppose there are $x \neq y$ in $B_r(a)$ with $f(x) = f(y)$. Since f is differentiable on $B_r(a)$, every f_i is differentiable on $B_r(a)$. Fix $1 \leq i \leq n$. By the MVT, there is a $c_i \in L(x, y)$ such that $0 = f_i(x) - f_i(y) = Df_i(c_i)(x - y)$. Letting $A = \left[\frac{\partial f_i}{\partial x_j}(c_i)\right]$ we see that $A(x - y) = 0$. Since $x - y \neq 0$, A is not invertible and so

$$\det\left(\frac{\partial f_i}{\partial x_j}(c_i)\right) = 0$$

a contradiction. □

Remark. Cramer's Rule: Recall Cramer's Rule: Let A be an $n \times n$ invertible matrix and consider a system of equations $Ax = b$. This system has a unique solution $(x_1, \ldots, x_n)^T \in \mathbb{R}^n$ given by

$$x_i = \frac{\det(A(i))}{\det A}$$

where $A(i)$ is the matrix obtained from A by replacing its ith column by b.

Theorem. Inverse Function Theorem: Let $U \subseteq \mathbb{R}^n$ be open and let $f \in C^1(U, \mathbb{R}^n)$. If $a \in U$ is such that $Jf(a) \neq 0$, then there is an open $a \in W \subseteq U$ such that

1. f is injective on W.
2. $f^{-1} \in C^1(f(W), \mathbb{R}^n)$.
3. And for all $y \in f(W)$ $D(f^{-1})(y) = [Df(x)]^{-1}$ where $x = f^{-1}(y)$.

Proof. 1. By the third lemma, there is an $r > 0$ with $W := B_r(a) \subseteq U$ such that f is injective on W, $Jf \neq 0$ on W, and

$$\det\left(\frac{\partial f_i}{\partial x_j}(c_i)\right) \neq 0$$

for all $c_1, \ldots, c_n \in W$. Moreover, by the second lemma, f^{-1} is continuous on $f(W)$.

2. We wish to show $f^{-1} \in C^1(f(W), \mathbb{R}^n)$. Fix $y_0 \in f(W)$ and $1 \leq i, j \leq n$. Choose $0 \neq t \in \mathbb{R}$ sufficiently small so that $y_0 + te_j \in f(W)$. We may find then $x_0, x_1 = x_1(t) \in W$ such that $f(x_0) = y_0$ and $f(x_1) = y_0 + te_j$. By the MVT, for every $1 \leq i \leq n$ there exists $c_i = c_i(t) \in L(x_0, x_1)$ such that

$$\nabla f_i(c_i)(x_1 - x_0) = f_i(x_1) - f_i(x_0) = \begin{cases} t & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Therefore,

$$\nabla f_i(c_i)\left(\frac{x_1 - x_0}{t}\right) = \frac{1}{t}(f_i(x_1) - f_i(x_0)) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Now let A_j be the $n \times n$ matrix whose ith row is $\nabla f_i(c_i)$. By assumption, $\det(A_j) \neq 0$. Moreover, $A_j(\frac{x_1 - x_0}{t}) = e_j$. For $1 \leq k \leq n$, we see that

$$\frac{(f^{-1})_k(y_0 + te_j) - (f^{-1})_k(y_0)}{t} = \frac{x_{1,k} - x - 0, k}{t}$$

where by Cramer's Rule, $Q_k(t) := \frac{x_{1,k} - x_{0,k}}{t}$ is a quotient of determinants of matrices whose entries are either 0, 1, or a first-order partial of f evaluated at a c_ℓ. As $t \to 0$ we clearly have that $y_0 + te_j \to y_0$. But then, by the continuity off f^{-1}, we have that $x_1 \to x_0$ and so $c_i \to x_0$. Since f is C^1, we therefore that that $Q_k(t) \to Q_k$, where Q_k is a quotient of determinants of matrices whose entries are either 0, 1, or a first-order partial of f evaluated at $x_0 = f^{-1}(y_0)$. Since $f \in C^1$ and f^{-1} is continuous at y_0,it follows that Q_k is continuous at each $y_0 \in f(W)$. Moreover

$$\lim_{t \to 0} \frac{(f^{-1})_k(y_0 + te_j) - (f^{-1})_k(y_0)}{t} = \lim_{t \to 0} \frac{x_{1,k} - x_{0,k}}{t} = Q_k$$

Hence all of the partial derivatives of f^{-1} exist and are continuous at y_0 (i.e. $f^{-1} \in C^1(f(W), \mathbb{R}^n)$).

3. Finally, we quickly run the chain rule and note that for $y \in f(W)$,

$$I = D\,\mathrm{Id}(y) = D(f \circ f^{-1})(y) = Df(f^{-1}(y))D(f^{-1})(y)$$

The result follows.

□

Example: Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(x, y) = (x + y, \sin x + \cos y)$. Note that $f_x(x, y) = (1, \cos x)$ and $f_y(x, y) = (1, -\sin y)$ hence $f \in C^1(\mathbb{R}^2, \mathbb{R}^2)$. Prove that f^{-1} exists and is differentiable on some open set containing $(0, 1)$ and compute $D(f^{-1})(0, 1)$.

First we find the preimage:

$$\begin{aligned}
&f(x, y) = (0, 1) \\
\iff &(x + y, \sin x + \cos y) = (0, 1) \\
\iff &y = -x, \sin x + \cos(-x) = 1 \\
\iff &y = -x, \sin x + \cos x = 1 \\
\iff &(x, y) = (2k\pi, -2k\pi) \text{ OR } (x, y) = \left(\frac{\pi}{2} + 2k\pi, -\frac{\pi}{2} - 2k\pi\right) \qquad \text{where } k \in \mathbb{Z}
\end{aligned}$$

Case 1: $a = (2k\pi, -2k\pi)$ for $k \in \mathbb{Z}$. Then we have

$$Jf(a) = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1 \neq 0$$

So by Inverse Function Theorem, there is an open $a \in W \subseteq \mathbb{R}^2$ such that f is injective on W and $f^{-1} \in C^1(f(W), \mathbb{R}^2)$. Note $(0, 1) \in f(W)$. Moreover,

$$D(f^{-1})(0, 1) = [Df(a)]^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$$

Case 2: $a = (\frac{\pi}{2} + 2k\pi, -\frac{\pi}{2} - 2k\pi)$ for $k \in \mathbb{Z}$. Then we have

$$Jf(a) = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 \neq 0$$

Again, there is an open $a \in W$ such that $f^{-1} \in C^1(f(W), \mathbb{R}^2)$ with

$$D(f^{-1})(0,1) = [Df(a)]^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

Remark: The way we restrict f to make it injective depends on our choice for $f^{-1}(y)$. This will affect the total derivative and the also the domain of the inverse. For instance, a true inverse on $\mathbb{R}^2$ of f in the above example is impossible due to the trigonometric functions.

9.2 Implicit Function Theorem

Note: We want to determine when and where we can solve $f(x, y, z) = 0$ with z as a function of x, y. I.e. when there is a $g(x, y) : \mathbb{R}^2 \to \mathbb{R}$ with

$$\{(x, y, z) \in \mathbb{R}^3 : f(x, y, z) = 0\} = \{(x, y, g(x, y)) : f(x, y, g(x, y)) = 0\}$$

Example: Solve $f(x, y, z) = (x^2 + y^2 + z^2 - 1) = 0$. Where $U = \{(x, y, z) \in \mathbb{R}^3 : z > 0\}$ then with

$$z = g(x, y) = \sqrt{1 - x^2 - y^2}$$

we have on U $f(x, y, g(x, y)) = 0$.

Theorem. Implicit Function Theorem: Let $U \subseteq \mathbb{R}^{n+p}$ be an open set and let $f = (f_1, \ldots, f_n) \in C^1(U, \mathbb{R}^n)$. Let $x_0 \in \mathbb{R}^n$ and $t_0 \in \mathbb{R}^p$ be such that $f(x_0, t_0) = 0$. If

$$\det \left[\frac{\partial f_i}{\partial x_j}(x_0, t_0)\right]_{n \times n} \neq 0$$

Then, there is an open set V with $t_0 \in V \subseteq \mathbb{R}^p$ and a unique $g \in C^1(V, \mathbb{R}^n)$ such that $g(t_0) = x_0$ and $f(g(t), t) = 0$ for all $t \in V$.

Proof. For every $(x, t) \in U$, let

$$F(x, t) := (f(x, t), t) = (f_1(x, t), \ldots, f_n(x, t), t_1, \ldots, t_p)$$

Notice that $F(x_0, t_0) = (0, t_0)$. Then F is a function from U to $\mathbb{R}^{n+p}$ with

$$DF = \begin{bmatrix} \left(\frac{\partial f_i}{\partial x_j}\right)_{n \times n} & B \\ O_{p \times n} & I_{p \times p} \end{bmatrix}$$

Taking the determinant of this matrix evaluated at (x_0, t_0), we have by our hypothesis

$$JF(x_0, t_0) = \det \left(\frac{\partial f_i}{\partial x_j}(x_0, t_0)\right)_{n \times n} \cdot \det I_{n \times p} \neq 0$$

Therefore, by the inverse function theorem, there is an open set $(x_0, t_0) \in W \subseteq U$ such that F is injective on W and $F^{-1} \in C^1(F(W), \mathbb{R}^{n+p})$.

To ease notation, let $G = F^{-1} = (G_1, \ldots, G_n, G_{n+1}, \ldots, G_{n+p})$. Consider $\varphi : F(W) \to \mathbb{R}^n$ given by $\varphi = (G_1, \ldots, G_n)$

By construction, we have that

$$\varphi(F(x,t)) = x$$

for all $(x,t) \in W$ and

$$F(\varphi(x,t), t) = (x,t)$$

for all $(x,t) \in W$.

Consider $V = \{t \in \mathbb{R}^p : (0,t) \in F(W)\}$ and the function $g : V \to \mathbb{R}^n$ given by $g(t) = \varphi(0,t)$. Since G is C^1, it follows that φ is also C^1. Hence $g \in C^1(V, \mathbb{R}^n)$. Also note that V is open since $F(W)$ is open. Finally, we compute that

$$g(t_0) = \varphi(0, t_0) = \varphi(F(x_0, t_0)) = x_0$$

and note that for all $(x,t) \in F(W)$,

$$f(\varphi(x,t), t) = x$$

In particular,

$$0 = f(\varphi(0,t), t) = f(g(t), t) = 0$$

for all $t \in V$. Uniqueness follows from the injectivity of F. $\square$

Remark: A summary of the above theorem: we keep the last p variables, using $t \in V \subseteq \mathbb{R}^p$ as these variables. Conversely $g(t) \in \mathbb{R}^n$, are the first n variables which we replace by an implicit function of t.

Example: Consider the equation $xyz + \sin(x, y + z) = 0$ and in particular the function $f(x,y,z) = xyz + \sin(x + y + z)$ so that $f \in C^1(\mathbb{R}^3)$. We want to replace the z variable with an implicit function. Note that

$$f(\underbrace{0, 0}_{t_0}, \underbrace{0}_{x_0}) = 0$$

hence we have $n = 1$ in this case. Now

$$f_z(x,y,z) = xy + \cos(x + y + z)$$

thus

$$\det[f_z(0,0,0)] = \det[1] = 1 \neq 0$$

hence by the implicit function theorem, there is an open $V \subseteq \mathbb{R}^2$ with $(0,0) \in V$ and $g(x,y) \in C^1(V)$ such that $g(0,0) = 0$ and $f(x, y, g(x,y)) = 0$ for all $x, y \in V$. I.e. $z = g(x,y)$ on V.

Example: Prove there are functions $u, v : \mathbb{R}^4 \to \mathbb{R}$ and there is an open set U with $(2, 1, -1, -2) \in U \subseteq \mathbb{R}^4$ such that

1. $u, v \in C^1(U)$.
2. $u(2,1,-1,-2) = 4$ and $v(2,1,-1,-2) = 3$.
3. For all $x, y, z, w \in U$, $u^2 + v^2 + w^2 = 29$ and $\frac{u^2}{x^2} + \frac{v^2}{y^2} + \frac{w^2}{z^2} = 17$.

where u and v are abuses of notation meaning $u(x, y, z, w)$ and $v(x, y, z, w)$ respectively in clause 3.

Consider the function $f : \mathbb{R}^6 \to \mathbb{R}^2$ given by

$$f(u, v, x, y, z, w) = \left(u^2 + v^2 + w^2 - 29, \frac{u^2}{x^2} + \frac{v^2}{y^2} + \frac{w^2}{z^2} - 17\right)$$

The aim is to use the implicit function theorem to keep x, y, z, w and rewrite u, v as functions of x, y, z, w. We have that $x_0 = (4, 3)$ and $t_0 = (2, 1, -1, -2)$. Notice that $f(4, 3, 2, 1, -1, -2) = 0$ and

$$\det \begin{bmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v} \end{bmatrix} = \begin{vmatrix} 2u & 2v \\ \frac{2u}{x^2} & \frac{2v}{y^2} \end{vmatrix} = 4uv\left(\frac{1}{y^2} - \frac{1}{x^2}\right)$$

Notice in particular this is non-zero at $(4, 3, 2, 1, -1, -2)$. By the implicit function theorem, there is an open set U with $(2, 1, -1, -2) \in U \subseteq \mathbb{R}^4$ and there is a $g \in C^1(U, \mathbb{R}^2)$ such $g(\underbrace{2, 1, -1, -2}_{t_0}) = (\underbrace{4, 3}_{x_0})$ and for all $(x, y, z, w) \in U$,

$$f(g(x, y, z, w), x, y, z, w) = 0$$

We're done. Why? Write g as a scalar function $g = (u, v)$ where $u, v \in C^1(U)$ (components inherit continuous differentiability). Notice we have $g(2, 1, -1, -2) = (4, 3)$, thus $u(2, 1, -1, -2) = 4$ and $v(2, 1, -1, -2) = 3$. Also notice we have $f(g(x, y, z, w), x, y, z, w) = 0$ for $(x, y, z, w) \in U$. But this means we have (using our abuse of notation again)

$$u^2 + v^2 + w^2 - 29 = 0 = \frac{u^2}{x^2} + \frac{v^2}{y^2} + \frac{w^2}{z^2} - 17$$

or $u^2 + v^2 + w^2 = 29$ and $\frac{u^2}{x^2} + \frac{v^2}{y^2} + \frac{w^2}{z^2} = 17$ for $(x, y, z, w) \in U$, as desired.

9.3 Lagrange Multipliers

Definition. Constrained Local Max (resp. Min): Let $U \subseteq \mathbb{R}^n$ be open and $f : U \to \mathbb{R}$. Let $a \in U$. We say $f(a)$ is a local max (resp. local min) of f subject to the constraints $g_i : U \to \mathbb{R}$ for $1 \leq i \leq m$ if $g_i(a) = 0$ for all $1 \leq i \leq m$ and there is an $r > 0$ such that whenever $x \in B_r(a)$ and $g_i(x) = 0$ for all i, then $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$).

Theorem. Lagrange Multipliers: Let $U \subseteq \mathbb{R}^n$ be open and let $m < n$. Let $f, g_1, g_2, \ldots, g_m \in C^1(U)$. Suppose $a \in U$ is such that

$$\det \left[\frac{\partial g_i}{\partial x_j}(a)\right]_{m \times m} \neq 0$$

If $f(a)$ is a local extremum of f subject to the constraints g_i (for $1 \leq i \leq m$), then there is $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that

$$\nabla f(a) + \sum_{i=1}^{m} \lambda_i \nabla g_i(a) = 0$$

Proof. This not a proof but rather the idea of how to prove such a theorem for $m = 2$, $n = 3$. Let

$$A = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(a) & \frac{\partial g_1}{\partial x_2}(a) \\ \frac{\partial g_2}{\partial x_1}(a) & \frac{\partial g_2}{\partial x_2}(a) \end{bmatrix}$$

with $\det(A) \neq 0$. We wish to show that there is $\lambda_1, \lambda_2 \in \mathbb{R}$ such that

$$\lambda_1 \frac{\partial g_1}{\partial x_j}(a) + \lambda_2 \frac{\partial g_2}{\partial x_j}(a) = -\frac{\partial f}{\partial x_j}(a) \tag{1}$$

for $j = 1, 2, 3$. Notice by invertible matrix theorem there is a unique solution (in λ_1, λ_2) to the equation

$$\begin{bmatrix} \lambda_1, \lambda_2 \end{bmatrix} A = \begin{bmatrix} -\frac{\partial f}{\partial x_1}(a) & -\frac{\partial f}{\partial x_2}(a) \end{bmatrix}$$

This covers cases $j = 1, 2$ for (1). It remains to prove for $j = 3$

$$\lambda_1 \frac{\partial g_1}{\partial x_3}(a) + \lambda_2 \frac{\partial g_2}{\partial x_3}(a) = -\frac{\partial f}{\partial x_3}(a) \tag{2}$$

Here the proof becomes less rigorous. The idea is to use the implicit function theorem to find replace x_3 with a function such that $x_3 = h(x_1, x_2)$. We then prove (2) using what we know about x_1, x_2 and the chain rule. □

Example: Maximize and minimize $f(x, y, z) = x + 2y$ subject to the constraints (1) $x + y + z = 1$ and (2) $y^2 + z^2 = 4$. Geometrically this is the intersection of a plane and a cylinder.

Let $f(x, y, z) = x + 2y$, $g_1(x, y, z) = x + y + z - 1$ and $g_2(x, y, z) = y^2 + z^2 - 4$. Note that

$$\det \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{bmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 2y \end{vmatrix} = 2y \neq 0 \qquad \text{for } y \neq 0$$

Now if $g_1(x, 0, z) = g_2(x, 0, z) = 0$ then $z = \pm 2$ and $x = 1 \mp 2$. In this case we have $f(1 \mp 2, 0, \pm 2) = 1 \mp 2$. Otherwise, a max/min must be of the form

$$\nabla f = \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2$$
$$[1, 2, 0] = \lambda_1 [1, 1, 1] + \lambda_2 [0, 2y, 2z]$$

Thus we get the systems

$$1 = \lambda_1 \qquad 2 = \lambda_1 + 2\lambda_2 y \qquad 0 = \lambda_1 + 2\lambda_2 z$$

Solving gives $y = \frac{1}{2\lambda_2}$ and $z = -\frac{1}{2\lambda_2}$. But, $y^2 + z^2 = 4$ means $\frac{2}{4\lambda_2^2} = 4$ which yields

$$\lambda_2 = \pm\frac{1}{2\sqrt{2}} \qquad y = \pm\sqrt{2} \qquad z = \mp\sqrt{2} \qquad x = 1$$

plugging these in, we get a maximum and minimum respectively of

$$f(1, \sqrt{2}, -\sqrt{2}) = 1 + 2\sqrt{2} \qquad \text{and} \qquad f(1, -\sqrt{2}, \sqrt{2}) = 1 - 2\sqrt{2}$$

Unit 3 Integration

Week 10 Multi-variable Integrals

10.1 Riemann Integration

Remark: Recall for $f : [a, b] \to \mathbb{R}$ which is bounded:

1. A partition of $[a, b]$ is a set $P = \{x_0, x_1, \dots, x_n\}$ with $a = x_0 < x_1 < \cdots < x_n = b$.
2. For a partition $P = \{x_0, \dots, x_n\}$:

$$\begin{aligned} U(f, P) &= \sum_{i=1}^{n} M_i(x_i - x_{i-1}) \\ L(f, P) &= \sum_{i=1}^{n} m_i(x_i - x_{i-1}) \\ M_i &= \sup\{f(x) : x \in [x_{i-1}, x_i]\} \\ m_i &= \inf\{f(x) : x \in [x_{i-1}, x_i]\} \end{aligned}$$

3. We say f is integrable if and only if the upper and lower and lower Riemann integrals are equal:

$$\inf_P\{U(f, P)\} := \overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx =: \sup_P\{L(f, P)\}$$

When f is integrable, we simply write its integral as

$$\int_a^b f(x)dx$$

Definition. Rectangle: A rectangle in $\mathbb{R}^n$ is a set of the form

$$R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

where for all $1 \le i \le n$ we have $a_i \le b_i$.

Definition. Partition: Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$ be a rectangle. A partition of R is a grid of rectangles on R obtained by partition each $[a_i, b_i]$.

Definition. Volume of Rectangle: Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$ be a rectangle. The volume of R is

$$v(R) := (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n)$$

Definition. Upper Sum & Lower Sum: Let $R \subseteq \mathbb{R}^n$ be a rectangle and let $f : R \to \mathbb{R}$ be bounded. Let $P = \{R_1, \ldots, R_m\}$ be a partition of R.

- The upper Riemann sum of f relative to P is

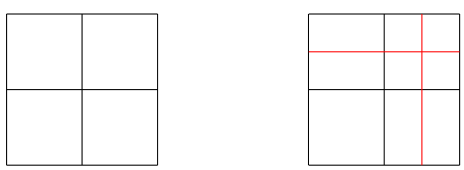
$$U(f, P) = \sum_{i=1}^{n} M_i v(R_i)$$
$$M_i = \sup\{f(x) : x \in R_i\}$$

- The lower Riemann sum of f relative to P is

$$L(f, P) = \sum_{i=1}^{n} m_i v(R_i)$$
$$m_i = \inf\{f(x) : x \in R_i\}$$

Definition. Refinement: Let P, Q be partitions of $R \subseteq \mathbb{R}^n$. We say P is a refinement of Q, written $P \leq Q$, if P is obtained from Q by partitioning the sides of R even furthers.

Example: Consider the following partitions of $\mathbb{R}^2$ with $P \leq Q$:



Q P

Proposition: Let P, Q be partitions of a rectangle R and let $f : R \to \mathbb{R}$ be bounded. If $P \leq Q$ then $U(f, P) \leq U(f, Q)$ and $L(f, P) \geq L(f, Q)$.

Proof. The proof would follow by generalizing the single variable case. See Snew's notes, lemma 1.10. General idea is that by adding a rectangle, we are taking the infimum on two parts as opposed to a single. This gives makes the infimum larger in general. See Blake's lecture module 10.1 for a geometric proof. □

Definition. Riemann Integral: Let $R \subseteq \mathbb{R}^n$ be a rectangle and let $f : R \to \mathbb{R}$ be bounded.

- The lower integral of f is
$$\underline{\int_R} f := \sup\{L(f,P) : P\}$$
- The upper integral of f is
$$\overline{\int_R} f := \inf\{U(f,P) : P\}$$
- We say f is (Riemann) integrable over R if the upper and lower integrals of f are equal and in this case
$$\int_R f := \underline{\int_R} f = \overline{\int_R} f$$

Example: Let $f : [0,1] \times [0,1] \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$$

then for all partitions P we have $U(f,P) = 1$ and $L(f,P) = 0$ by the density of the rationals and irrationals. So
$$\underline{\int_R} f = 0 \neq 1 = \overline{\int_R} f$$

10.2 Characterization Theorem

Lemma: Let $R \subseteq \mathbb{R}^n$ be a rectangle and $f : R \to \mathbb{R}$ be bounded. If P, Q are partitions of R then $L(f,P) \leq U(f,Q)$.

Proof. Find a common refinement $S \leq P$, $S \leq Q$ (e.g. S overlaps P and Q) then

$$L(f,P) \leq L(f,S) \leq U(f,S) \leq U(f,Q)$$

□

Remark: Let $R \subseteq \mathbb{R}^n$ be a rectangle and $f : R \to \mathbb{R}$ be bounded. For all partitions P, Q, $L(f,P) \leq U(f,Q)$. Thus
$$L(f,P) \leq \underline{\int_R} f \leq \overline{\int_R} f \leq U(f,Q)$$

Theorem: Let $R \subseteq \mathbb{R}^n$ be a rectangle and $f : R \to \mathbb{R}$ be bounded. Then f is integrable if and only if for all $\epsilon > 0$, there is a partition P such that $U(f,P) - L(f,P) < \epsilon$.

Proof. ($\implies$) Suppose f is integrable. So we have

$$\underline{\int_R} f = \overline{\int_R} f$$

Let $\epsilon > 0$ be given. We may find partitions P, Q such that

$$\underline{\int_R} f - \frac{\epsilon}{2} < L(f, P)$$

and

$$\overline{\int_R} f + \frac{\epsilon}{2} > U(f, Q)$$

So, we have

$$U(f, Q) < \overline{\int_R} f + \frac{\epsilon}{2} = \underline{\int_R} f - \frac{\epsilon}{2} + \epsilon < L(f, P) + \epsilon$$

Let S be a common refinement of P, Q ($S \leq P, Q$). Therefore,

$$U(f, S) < U(f, Q) < L(f, P) + \epsilon < L(f, S) + \epsilon$$

and so

$$U(f, S) - L(f, S) < \epsilon$$

($\implies$) Assume for all $\epsilon > 0$ there is a partition P such that $U(f, P) - L(f, P) < \epsilon$. Let $\epsilon > 0$ be given. We may a partition P of R such that

$$U(f, P) - L(f, P) < \epsilon$$

Now

$$0 \leq \overline{\int_R} f - \underline{\int_R} f \leq U(f, P) - L(f, P) < \epsilon$$

therefore,

$$\overline{\int_R} f = \underline{\int_R} f$$

and so f is integrable □

10.3 Content and Measure

Definition. Lebesgue Measure Zero: A set $A \subseteq \mathbb{R}^n$ has (Lebesgue) measure zero if for all $\epsilon > 0$ there exists rectangles R_i (for $i \in \mathbb{N}$) such that

$$A \subseteq \bigcup_{i=1}^{\infty} R_i$$

and

$$\sum_{i=1}^{\infty} v(R_i) < \epsilon$$

Definition. Jordan Content Zero: A set $A \subseteq \mathbb{R}^n$ has (Jordan) content zero if for all $\epsilon > 0$ there exists rectangles $R_1, \dots, R_m$ such that

$$A \subseteq \bigcup_{i=1}^{m} R_i$$

and

$$\sum_{i=1}^{m} v(R_i) < \epsilon$$

Remark: The idea of having Lebesgue measure zero is that A is in a sense small and already close to having volume of zero. We can cover A as finely as we want as to not add too much volume in the covering.

Proposition: If $A \subseteq \mathbb{R}^n$ has content zero, then it has measure zero.

Proof. Let $\epsilon > 0$ be given and suppose A has content zero. We know then there are rectangles $R_1, \ldots, R_m$ such that $A \subseteq R_1 \cup \cdots \cup R_m$ and $\sum_{i=1}^{m} v(R_i) < \epsilon$. For $i > m$, let $R_i \subseteq \mathbb{R}^n$ be any rectangle with volume zero. (Note we never mentioned rectangle had to be non-degenerate.) Therefore, $A \subseteq \bigcup_{i=1}^{\infty} R_i$ and

$$\sum_{i=1}^{\infty} v(R_i) = \sum_{i=1}^{m} Vv(R_i) < \epsilon$$

□

Example: Consider $A = \mathbb{Q} \subseteq \mathbb{R}^1$. We claim $\mathbb{Q}$ has measure zero. Note that $\mathbb{Q}$ is countable. We can therefore write $\mathbb{Q} = \{q_1, q_2, q_3, \ldots\}$. Let $\epsilon > 0$. Let $R_i = [q_i - \frac{\epsilon}{2^{i+2}}, q_i + \frac{\epsilon}{2^{i+2}}]$. Clearly $\mathbb{Q} \subseteq \bigcup_{i=1}^{\infty} R_i$. Further

$$\sum_{i=1}^{\infty} v(R_i) = \sum_{i=1}^{\infty} \frac{\epsilon}{2^{i+1}} = \epsilon \sum_{i=1} \frac{1}{2^{i+1}} = \frac{\epsilon}{2}$$

By the properties of geometric series. Therefore, $\mathbb{Q}$ has measure zero. Since $\mathbb{Q}$ is unbounded, it cannot be covered by finitely many rectangles, therefore it does not have content zero. Thus, the converse of the above proposition is not true. Note the above proof also shows that any countable set has measure zero.

Proposition: If $A_1, A_2, A_3, \ldots \subseteq \mathbb{R}^n$ have measure zero, then $A = \bigcup_{i=1}^{\infty} A_i$ has measure zero.

Proof. Let $\epsilon > 0$. We may find for each $i \in \mathbb{N}$, we may find rectangles such that $A_i \subseteq \bigcup_{j=1}^{\infty} R_{i,j}$ and such that $\sum_{j=1}^{\infty} v(R_{i,j}) < \frac{\epsilon}{2^i}$. So we have

$$A = \bigcup_{i=1}^{\infty} A_i \subseteq \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{\infty} R_{i,j} = \bigcup_{(i,j) \in \mathbb{N} \times \mathbb{N}} R_{i,j}$$

and

$$\sum_{(i,j) \in \mathbb{N} \times \mathbb{N}} v(R_{i,j}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} v(R_{i,j}) < \sum_{i=1}^{\infty} \frac{\epsilon}{2^i} = \epsilon$$

□

Proposition: If $A \subseteq \mathbb{R}^n$ is compact and has measure 0 then A has content 0.

Proof. Let $\epsilon > 0$ be given. By A9 Q3 there are open rectangles R_i such that $A \subseteq \bigcup_{i=1}^{\infty} R_i$ and $\sum_{i=1}^{\infty} v(R_i) < \epsilon$. Since A is compact and $\bigcup_{i=1}^{\infty} R_i$ forms an open cover of A, there is a finite subcover, say $R_{i_1}, \dots, R_{i_m}$.

Hence $A \subseteq \bigcup_{j=1}^{m} R_{i_j}$ and clearly $\sum_{j=1}^{m} v(R_{i_j}) \leq \sum_{i=1}^{\infty} v(R_i) < \epsilon$. Therefore A has content zero. □

10.4 Integrability and Measure

Definition. Oscillation: Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be bounded. For $a \in A$ and $\delta > 0$ let

$$\begin{aligned} M(a, f, \delta) &= \sup\{f(x) : x \in A, \|x - a\| < \delta\} \\ m(a, f, \delta) &= \inf\{f(x) : x \in A, \|x - a\| < \delta\} \\ o(f, a) &= \lim_{\delta \to 0} M(a, f, \delta) - m(a, f, \delta) \end{aligned}$$

We call $o(f, a)$ the oscillation of f at a.

Remark: The oscillation of f at a always exists (use MCT to show), f is continuous at a if and only if $o(f, a) = 0$.

Proposition: Let $A \subseteq \mathbb{R}^n$ be closed and $f : A \to \mathbb{R}$ be bounded. Then for all $\epsilon > 0$, the set $\{x \in A : o(f, x) \geq \epsilon\}$ is closed.

Proof. Let $B = \{x \in A : o(f, x) \geq \epsilon\}$. We will show that $A \setminus B$ is relatively open, thereby proving B is relatively closed in A, or simply closed. Let $x \in A \setminus B$. We know then that $o(f, x) < \epsilon$, and so there is a $\delta > 0$ with

$$M(x, f, \delta) - m(x, f, \delta) < \epsilon$$

Consider $y \in B_{\delta/2}(x) \cap A$. Then for any $z \in A$ with $|y - z| < \frac{\delta}{2}$ we have

$$|z - x| \leq |z - y| + |y - x| < \delta$$

and so $m(x, f, \delta) \leq f(z) \leq M(x, f, \delta)$. Therefore, since z was any point in with $\frac{\delta}{2}$ of y, we have

$$M\left(y, f, \frac{\delta}{2}\right) - m\left(y, f, \frac{\delta}{2}\right) \leq M(x, f, \delta) - m(f, x, \delta) < \epsilon$$

So $o(f, y) < \epsilon$. Now since $y \in B_{\delta/2}(x) \cap A$ was arbitrary, we have $B_{\delta/2}(x) \cap A \subseteq A \setminus B$. Thus since for an arbitrary $x \in A \setminus B$ we can find an $r > 0$ such that $B_r(x) \cap A \subseteq A \setminus B$ and so $A \setminus B$ is relatively open. Thus as as mentioned above, B is closed as desired. □

Proposition: Let $R \subseteq \mathbb{R}^n$ be a rectangle and $f : R \to \mathbb{R}$ be bounded. Let $\epsilon > 0$. If $o(f, x) < \epsilon$ for all $x \in R$, then there is a partition P of R with $U(f, P) - L(f, P) < \epsilon v(R)$.

Proof. Let $\epsilon > 0$. For all $x \in R$ we may find a $\delta_x > 0$ such that

$$M(x, f, \delta_x) - m(x, f, \delta_x) < \epsilon$$

For all $x \in R$, let R_x be an open rectangle such that

$$x \in R_x \subseteq B_{\frac{\delta_x}{2}}(x)$$

Let $U_x = R_x \cap R$. We see that

$$R = \bigcup_{x \in R} U_x$$

is relatively open cover of R. Since R is compact, there exists $x_1, \ldots, x_m \in R$ such that

$$R = U_{x_1} \cup \cdots \cup U_{x_m}$$

Let P be a partition of R so fine that each subrectangle in P is contained in some $\overline{U_{x_i}}$ (possible since the $R = U_{x_1} \cup \cdots \cup U_{x_m}$). Note that

$$\overline{U_{x_i}} = \overline{R_{x_i}} \cap R \subseteq \overline{B_{\frac{\delta_{x_i}}{2}}} \cap R \subseteq B_{\delta_{x_i}} \cap R$$

Therefore, for ever $R_i \in P$, $M_i - m_i < \epsilon$. This is since each $x \in R_i$ has $\|x - x_j\| < \delta_{x_j}$ for some x_j. Thus

$$m(x_j, f, \delta_{x_j}) \leq f(x) \leq M(x_j, f, \delta_{x_j})$$

(Alternatively think of it as the rectangle is a subset of $B_{\delta_{x_j}}(x_j)$, thus the supremum is smaller and infimum larger.) Therefore,

$$U(f, P) - L(f, P) = \sum_{R_i \in P} (M_i - m_i) v(R_i) < \sum_{R_i \in P} \epsilon v(R_i) = \epsilon v(R)$$

as desired □

Theorem: Let $R \subseteq \mathbb{R}^n$ be a rectangle and $f : R \to \mathbb{R}$ be bounded. Let

$$A = \{x \in R : f \text{ is not continuous at } x\}$$

Then f is integrable if and only if A has measure zero.

Proof. ($\implies$) Suppose f is integrable. Let $\epsilon > 0$ be given. For every $n \in \mathbb{N}$, let

$$B_n = \left\{x \in R : o(f, x) \geq \frac{1}{n}\right\}$$

Since $A = B_1 \cup B_2 \cup \cdots$ (recall a point is continuous iff $o(f, x) = 0$), it suffices to show that each B_n has measure zero. Fix $n \in \mathbb{N}$.

Since f is integrable, we may find a partition P of R such that $U(f, P) - L(f, P) < \frac{\epsilon}{n}$. Let X denote the collection of rectangles in P which intersect with B_n. In particular, the elements of X cover B_n and are rectangles. Now if $R_i \in X$, then $M_i - m_i \geq \frac{1}{n}$ by the definition of the oscillation function. This is since it contains both $M(f,x,\delta)$ and $m(f,x,\delta)$ for some x and δ, therefore the difference it at least $\frac{1}{n}$. Then

$$\begin{aligned}\sum_{R_i\in X}\frac{1}{n}\cdot v(R_i) &\leq \sum_{R_i\in X}(M_i-m_i)v(R_i)\\ &\leq \sum_{R_i\in P}(M_i-m_i)v(R_i)\\ &= U(f,P)-L(f,P)<\frac{\epsilon}{n}\end{aligned}$$

In particular, $\sum\limits_{R_i\in X} v(R_i)<\epsilon$ and X covers B_n, so B_n has measure (in fact content) zero.

($\Longleftarrow$) Suppose A has measure zero. Let $\epsilon>0$ be given. Let $B=\{x\in R : o(f,x)\geq\epsilon\}$ so that B is compact (follows since R is bounded and we proved it is closed). Since $B\subseteq A$ (again point is continuous iff $o(f,x)=0$), we have that B also has measure zero. Since B is compact, B also has content zero. In particular, we may find finitely many rectangles $U_1,\dots,U_m$ whose interiors cover B (A9 Q3) such that $\sum\limits_{i=1}^m v(U_i)<\epsilon$.

Let X denote the set of subrectangles of R which are contained in at least one U_i. Let Y denote the set of subrectangles of R which are contained in $R\setminus B$. Now since the U_i's cover B, we may find a partition $P=\{R_1,\dots,R_k\}$ fine enough so that the elements are from either X or Y.

Since f is bounded, there exists M such that $|f(x)|\leq M$ for all $x\in R$. In particular for every $R_i\in P$, $M_i-m_i\leq 2M$. By the definition of X,

$$\sum_{R_i\in X}(M_i-m_i)v(R_i)\leq 2M\sum_{R_i\in X}v(R_i)\leq 2M\sum_{i=1}^m v(U_i)<2M\epsilon$$

Now, if $R_i\in Y$ and $x\in R_i$, we have that $o(f,x)<\epsilon$ (since $R_i\in R\setminus B$). By a proposition above, we may find a partition $P_i=\{S_{i_1},\dots,S_{i_{\alpha(i)}}\}$ of R_i such that

$$\sum_{j=1}^{\alpha(i)}(M_j-m_j)v(S_{i_j})<\epsilon v(R_i)$$

(Same trick as usual to show $M_j-m_j<\epsilon$.) By replacing each $R_i\in Y$ with $S_{i_1},\dots,S_{i_{\alpha(i)}}$ (and leaving the $R_i\in X$ alone), this creates a refinement $Q\leq P$. Finally,

$$\begin{aligned}U(f,Q)-L(f,Q) &= \sum_{R_i\in X}(M_i-m_i)v(R_i)+\sum_{R_i\in Y}\sum_{j=1}^{\alpha(i)}(M_j-m_j)v(S_{i_j})\\ &< 2M\epsilon+\sum_{R_i\in Y}\epsilon v(R_i)\\ &\leq 2M\epsilon+\epsilon v(R)\\ &\leq \epsilon(2M+v(R))\end{aligned}$$

Since M and R are fixed, the result holds by the alternative characterization of integrability (we can make the final term arbitrarily small). □

Remark: If A from the previous theorem has countably many points of discontinuity, then A has measure zero and thus is integrable.

Week 11 Theorems of Integration

Notation: The following are notational equivalences with the latter being preferred:

1. Grid on R – Partition of R.
2. $|R|$ – $v(R)$ (e.g. volume of rectangle).
3. $L\int_R f$ – $\underline{\int_R} f$.
4. $U\int_R f$ – $\overline{\int_R} f$.

11.1 General Integrability

Definition. Characteristic Function: Let $A \subseteq \mathbb{R}^n$ be bounded. Let R be a rectangle such that $A \subseteq R$. The characteristic function of A on R is $\chi_A : R \to \mathbb{R}$ (symbol is the Greek letter chi) given by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Definition. General Integrability: Let $A \subseteq \mathbb{R}^n$ be bounded. Let $f : A \to \mathbb{R}$ be bounded and let R be a rectangle with $A \subseteq R$. We extend f to $f : R \to \mathbb{R}$ by setting $f(x) = 0$ for all $x \in R \setminus A$. We say $f : A \to \mathbb{R}$ is integrable if and only if $f \cdot \chi_A : R \to \mathbb{R}$ is integrable, in which case we defined

$$\int_A f = \int_R f \cdot \chi_A$$

Note: The above definition is independent of the choice of R.

Remark: Let $A \subseteq R$ and $f : A \to \mathbb{R}$ be bounded. If $f : R \to \mathbb{R}$ and $\chi_A : R \to \mathbb{R}$ are integrable, then $f \cdot \chi_A$ is integrable. Hence f would be integrable over A. We ask ourselves then, when is $\chi_A : R \to \mathbb{R}$ integrable?

Theorem: Let $A \subseteq \mathbb{R}^n$ be bounded and let $A \subseteq R$ be a rectangle. The function $\chi_A : R \to \mathbb{R}$ is integrable if and only if $\partial(A)$ has measure zero.

Proof. Let $a \in R$. We consider the following cases

1. $a \in \text{Int}(A)$. Then there is an open ball $B_\delta(a) \subseteq A$. Since $\chi_A = 1$ on $B_\delta(a)$, χ_A is clearly continuous at a.
2. $a \notin \overline{A}$. Then $a \in \text{Int}(\mathbb{R}^n \setminus A)$ and so there is an open ball $B_\delta(a) \subseteq \mathbb{R}^n \setminus A$. Since $\chi_A = 0$ on $B_\delta(a) \cap R$, χ_A is clearly continuous at a.

3. $a \in \overline{A} \setminus \text{Int}(A) = \partial(A)$. Then $a \in \overline{A}$ and $a \in R^n \setminus \text{Int}(A) = \overline{\mathbb{R}^n \setminus A}$. In particular, for all $\delta > 0$ there exists $x \in A$ and $y \in R \setminus A$ such that $\|x - a\|, \|y - a\| < \delta$. Thus, $O(\chi_A, a) \geq 1$ and so χ_A is not continuous at a.

So χ_A is discontinuous at a if and only if $a \in \partial(A)$. Therefore, the set of discontinuities of A is $\partial(A)$ and so χ_A is integrable if and only if $\partial(A)$ has measure zero. □

Definition. Jordan Region: Let $A \subseteq \mathbb{R}^n$ be bounded. We call A a Jordan region if and only if $\partial(A)$ has measure zero (if and only if χ_A is integrable on $A \subseteq R$).

Definition. Volume of Jordan Region: If A is a Jordan region with $A \subseteq R$ for some rectangle, we defined the volume of A by

$$Vol(A) = \int_R \chi_A = \int_A 1$$

Proposition: Let $A, B \subseteq \mathbb{R}^n$ be Jordan regions. Then (1) $A \cup B$ is a Jordan Region and (2) if $A \cap B = \emptyset$ and $f : A \cup B \to \mathbb{R}$ is integrable, then

$$\int_{A \cup R} = \int_A f + \int_B f$$

Proof. 1. Notice that

$$\begin{aligned} \partial(A \cup B) &= (\overline{A \cup B}) \setminus \text{Int}(A \cup B) \\ &\subseteq (\overline{A} \cup \overline{B}) \setminus \text{Int}\, A \cup \text{Int}\, B \\ &\subseteq (\overline{A} \setminus \text{Int}\, A) \cup (\overline{B} \setminus \text{Int}\, B) \\ &= \partial(A) \cup \partial(B) \end{aligned}$$

Hence since union of measure zero sets have measure zero, $\partial(A \cup B)$ has measure zero. Since $\partial(A \cup B)$ heas measure zero, $A \cup B$ is a Jordan Region.

2. Let $A \cup B \subseteq R$ for some rectangle R. Then

$$\begin{aligned} \int_{A \cup B} f &= \int_R f \cdot \chi_{A \cup B} \\ &= \int_R f(\chi_A + \chi_B) \qquad (*) \\ &= \int_R f\chi_A + \int_R f\chi_B \\ &= \int_A f + \int_B f \end{aligned}$$

where * holds since if $x \in A \cup B$ then given $A \cap B = \emptyset$, x is distinctly in A or in B, so one of χ_A or χ_B is 1, while the other is zero. Similarly, if $(\chi_A + \chi_B)(x) = 1$, then either x in A or $x \in B$, either way $x \in A \cup B$.

And so, the result holds as desired. □

11.2 Fubini's Theorem

Notation: Let $B \subseteq \mathbb{R}^2$ be a Jordan region and $f : B \to \mathbb{R}$ be integrable. We denote

$$\int_B f(v)dv \equiv \iint_B f(x,y)dA$$

Here the v is a pair (x, y) in B and the A is to denote the area of B. Similarly, where $B \subseteq \mathbb{R}^3$, we may write

$$\int_B f(v)dv \equiv \iiint_B f(x,y,z)dV$$

where the V denotes the volume of B, and again v is a tuple (x, y, z).

Lemma: Let $R = [a, b] \times [c, d] \subseteq \mathbb{R}^2$ and $f : R \to \mathbb{R}$ be bounded. If $f(x, \cdot) : [c, d] \to \mathbb{R}$ given by $f(x, \cdot)(y) = f(x, y)$ is integrable for all $x \in [a, b]$, then

$$\underline{\iint_R} f(x,y)dA \leq \underline{\int_a^b} \left(\underline{\int_c^d} f(x,y)dy \right) dx \leq \overline{\int_a^b} \left(\overline{\int_c^d} f(x,y)dy \right) dx \leq \overline{\iint_R} f(x,y)dA$$

Proof. The middle inequality is obvious by a result from week 10. We prove the third inequality as the first inequality follows similarly.

Let $\epsilon > 0$ be given. Let P be a partition on R such that

$$U(f,P) - \epsilon \leq \overline{\iint_R} f(x,y)dA$$

Without loss of generality suppose $P = \{R_{ij} : 1 \leq i \leq k, 1 \leq j \leq \ell\}$ and suppose $R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$ where $x_0 = a, x_k = b, y_0 = c, y_\ell = d$. Let $M_{ij} = \sup\{f(v) : v \in R_{ij}\}$. Then we have

$$\begin{aligned}
\overline{\int_a^b} \left(\int_c^d f(x,y)dy \right) dx &= \sum_{i=1}^k \overline{\int_{x_{i-1}}^{x_i}} \left(\sum_{j=1}^\ell \int_{y_{i-1}}^{y_i} f(x,y)dy \right) dx \\
&\leq \sum_{i=1}^k \sum_{j=1}^\ell \overline{\int_{x_{i-1}}^{x_i}} \left(\int_{y_{i-1}}^{y_i} f(x,y)dy \right) dx \qquad (*) \\
&\leq \sum_{i=1}^k \sum_{j=1}^\ell \overline{\int_{x_{i-1}}^{x_i}} \left(\int_{y_{i-1}}^{y_i} M_{ij}dy \right) dx \\
&= \sum_{i=1}^k \sum_{j=1}^\ell M_{ij}(x_i - x_{i-1})(y_j - y_{j-1}) \\
&= \sum_{R_{ij} \in P} M_{ij} v(R_{ij}) \\
&= U(f,P) \\
&\leq \overline{\iint_R} f(x,y)dA + \epsilon
\end{aligned}$$

Note * holds because for all functions $f, g : [a, b] \to \mathbb{R}$, we have

$$\overline{\int_a^b} (f+g)(x)dx \leq \overline{\int_a^b} f(x)dx + \overline{\int_a^b} g(x)dx$$

The proof would follow mostly from writing Riemann sums and using the triangle inequality. □

Theorem. Fubini's Theorem: Let $R = [a, b] \times [c, d] \subseteq \mathbb{R}^2$ and let $f : R \to \mathbb{R}$ be integrable. If $f(x, \cdot)$ and $f(\cdot, y)$ (given as $f(x, \cdot : [c, d] to \mathbb{R}$ with $f(x, \cdot)(y) = f(x, y)$ and similarly for $f(\cdot, y)$) are integrable over $[c, d]$ and $[a, b]$ respectively, for all $x \in [a, b]$ and $y \in [c, d]$, then

$$\iint_R f(x,y)dA = \int_a^b \int_c^d f(x,y)dydx = \int_c^d \int_a^b f(x,y)dxdy$$

Proof. Since f is integrable,

$$\underline{\iint_R} f(x,y)dA = \overline{\iint_R} f(x,y)dA$$

By our Lemma, this means we have

$$\underline{\iint_R} f(x,y)dA = \underline{\int_a^b} \left(\underline{\int_c^d} f(x,y)dy \right) dx = \overline{\int_a^b} \left(\overline{\int_c^d} f(x,y)dy \right) dx = \overline{\iint_R} f(x,y)dA$$

and so

$$\iint_R f(x,y)dA = \int_a^b \int_c^d f(x,y)dydx$$

Reversing the roles of x and y in the lemma, we get the other equality and so the theorem is complete. □

Remark: When we integrate a function $f : A \to \mathbb{R}$ with $A \subseteq \mathbb{R}^n$ as

$$\int_{a_n}^{b_n} \int_{a_{n-1}}^{b_{n-1}} \cdots \int_{a_1}^{b_1} f(x_1, x_2, \ldots, x_n)dx_1 dx_2 \cdots dx_n$$

we call this an iterated integral.

Remark: Let the setup be like that of the above theorem. If f is continuous, then all the premises are met.

Example: Let $R = [1, 2] \times [0, \pi]$ and calculate

$$\iint_R y \sin(xy)dA$$

Note that $f(x, \cdot)$ and $f(\cdot, y)$ are both continuous on any closed Jordan region and therefore integrable. We have by Fubini's theorem that

$$\begin{aligned}\iint_R y\sin(x,y)dA &= \int_0^\pi \int_1^2 y\sin(xy)dxdy \\ &= \int_0^\pi \Big[-\cos(xy)\Big]_{x=1}^2 dy \\ &= \int_0^\pi \cos(y) - \cos(2y)dy \\ &= \Big[\sin(y) - \frac{1}{2}\sin(2y)\Big]_{y=0}^\pi \\ &= 0\end{aligned}$$

Note if we had done integration on y then on x, we would have had to do a difficult integration by parts. Thus always try and pick the easier integral to evaluate.

11.3 Iterated Integrals

Theorem. Generalized Fubini's Theorem: Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$ and $f : R \to \mathbb{R}$ be integrable. Denote $R_n = [a_1, b_1] \times [a_{n-1}, b_{n-1}] \subseteq \mathbb{R}^{n-1}$. If $f(x, \cdot)$ is integrable for all $x \in R_n$, then $\int_{a_n}^{b_n} f(x,t)dt$ is integrable on R_n and

$$\int_R f(v)dv = \int_{R_n}\int_{a_n}^{b_n} f(x,t)dtdx$$

Remark: If $f : R \to \mathbb{R}$ is continuous (for $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$), then

$$\int_R f(v)dv = \int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n} f(x_1, x_2, \ldots, x_n)dx_n dx_{n-2}\cdots dx_1$$

Proof. The proof follows very similarly from the $n = 2$ case above. □

Definition. Type 1, 2 Regions in $\mathbb{R}^2$: We say $A \subseteq \mathbb{R}^2$ is type 1 if

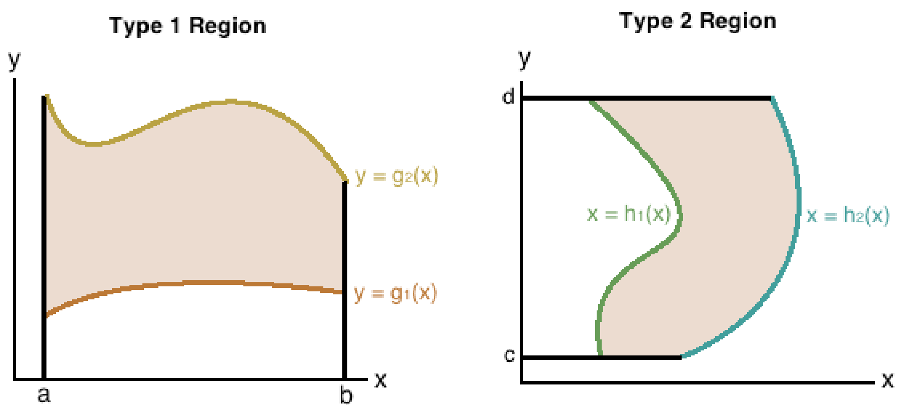
$$A = \{(x, y) : x \in [a, b], \varphi(x) \le y \le \psi(x)\}$$

for some continuous functions $\varphi, \psi : [a, b] \to \mathbb{R}$. Conversely, we say $A \subseteq \mathbb{R}^2$ is type if

$$A = \{(x, y) : y \in [a, b], \varphi(y) \le x \le \psi(y)\}$$

for some continuous functions $\varphi, \psi : [a, b] \to \mathbb{R}$.

Example: The following are type 1 and type 2 regions respectively



Notice in both cases there is a sort of "line" cutting off the region vertically for a type 1 region and horizontally for a type 2 region.

Definition. Type 1, 2, 3 Regions in $\mathbb{R}^3$: We say $A \subseteq \mathbb{R}^3$ is of

Type 1 If $A = \{(x, y, z) : (x, y) \in H, \varphi(x, y) \leq z \leq \psi(x, y)\}$

Type 2 If $A = \{(x, y, z) : (x, z) \in H, \varphi(x, y) \leq y \leq \psi(x, y)\}$

Type 3 If $A = \{(x, y, z) : (y, z) \in H, \varphi(x, y) \leq x \leq \psi(x, y)\}$

where $H \subseteq \mathbb{R}^2$ is a closed Jordan region and $\varphi, \psi : H \to \mathbb{R}$ are continuous.

Proposition: Regions of type 1, 2, and 3 are all Jordan regions.

Proof. Intuitively this is because they will have boundaries which are simply lines. These lines can be covered by infinitely small rectangles, giving them measure zero. □

Theorem: Let $A \subseteq \mathbb{R}^2$ and $f : A \to \mathbb{R}$ be *continuous*. (1) If A is of type 1 such that

$$A = \{(x, y) : x \in [a, b], \varphi(x) \leq y \leq \psi(x)\}$$

for some continuous functions $\varphi, \psi : [a, b] \to \mathbb{R}$, then

$$\int_A f(v)dv = \int_a^b \int_{\varphi(x)}^{\psi(x)} f(x, y)dydx$$

(2) If A is of type 2 such that

$$A = \{(x, y) : y \in [a, b], \varphi(y) \leq y \leq \psi(y)\}$$

for some continuous function $\varphi, \psi : [a, b] \to \mathbb{R}$, then

$$\int_A f(v)dv = \int_a^b \int_{\varphi(y)}^{\psi(y)} f(x, y)dxdy$$

Proof. We will prove (1), the proof of (2) follows similarly.

Let $R = [a, b] \times [c, d]$ (same a, b as in hypothesis; note we can find c, d by the EVT since φ, ψ are continuous) be a rectangle containing A. We extend f from A to R by setting $f(v) = 0$ for $v \in R \setminus A$. By Fubini's theorem,

$$\int_A f(v)dv = \int_R f(v)dv = \int_a^b \int_c^d f(x, y)dydx$$

However, if it is <u>not</u> the case that

$$\varphi(x) \leq y \leq \psi(x)$$

then $f(x, y) = 0$ and so integrating over this part is integrating over zero. We may thus write

$$\int_A f(v)dv = \int_a^b \int_{\varphi(x)}^{\psi(x)} f(x, y)dydx$$

□

Theorem: Let $A \subseteq \mathbb{R}^3$ and $f : A \to \mathbb{R}$ be continuous. (1) If A is of type 1 such that

$$A = \{(x, y, z) : (x, y) \in H, \varphi(x, y) \leq z \leq \psi(x, y)\}$$

for some continuous function $\varphi, \psi : H \to \mathbb{R}$ where H is a closed Jordan region, then

$$\int_A f(v)dv = \int_H \int_{\varphi(x,y)}^{\psi(x,y)} f(u, z)dzdu$$

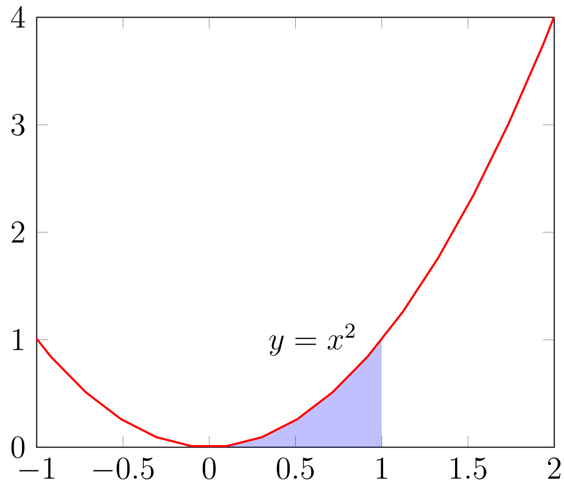
A similar result holds for A being of type 1 and type 2, replacing the appropriate variables.

Proof. Follows similarly to the $\mathbb{R}^2$ case. □

Remark: Note in the above theorem, if H were to be not only a closed Jordan region, but also to be of type 1 or 2 (in $\mathbb{R}^2$), then we may split the integral again as in the theorem for the $\mathbb{R}^2$ case.

11.4 Examples

Example: Let $D \subseteq \mathbb{R}^2$ be the region bounded between $0 \leq x \leq 1$ and $0 \leq y \leq x^2$. Compute the integral of $x \cos y$ on this region. The region is shaded in blue in the figure below



Notice clearly this is a type 1 region, since 0 and x^2 are continuous, thus

$$\begin{aligned}\iint_D x\cos y dA &= \int_0^1 \int_0^{x^2} x\cos y dy dx \\ &= \int_0^1 \Big[x\sin y\Big]_{y=0}^{y=x^2} dx \\ &= \int_0^1 x\sin x^2 dx \\ &= \Big[-\frac{1}{2}\cos x^2\Big]_{x=0}^1 \\ &= -\frac{1}{2}\cos 1 + \frac{1}{2}\end{aligned}$$

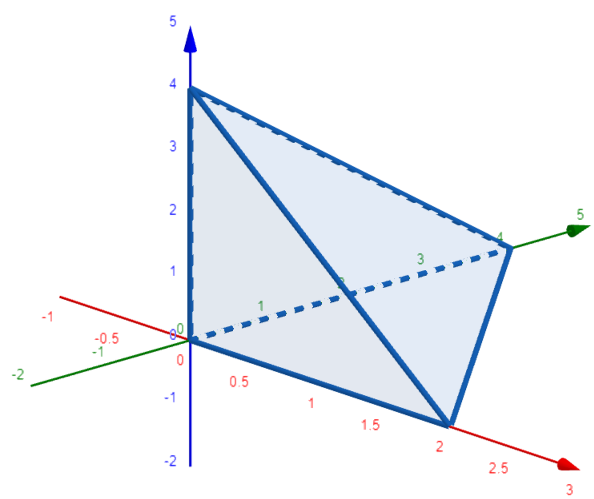
Example: Compute

$$\int_0^1 \int_{3y}^3 e^{x^2} dx dy$$

Notice e^{x^2} cannot be easily calculated, so we're in a tricky situation. We remark that this is equivalent to integrating over the region $0 \le y \le 1$ and $3y \le x \le 3$. Since the region is delimited by $x = 3y$ for $y \in [0, 1]$, we can also write $y = \frac{1}{3}x$ as $x \in [0, 3]$ giving the equivalent region D given by $0 \le x \le 3$, $0 \le y \le \frac{1}{3}x$. Therefore,

$$\begin{aligned} \int_0^1 \int_{3y}^3 e^{x^2} &= \iint_D e^{x^2} dA \\ &= \int_0^3 \int_0^{\frac{1}{3}x} e^{x^2} dy dx \\ &= \int_0^3 \left[e^{x^2} y \right]_{y=0}^{y=\frac{1}{3}x} dx \\ &= \int_0^3 \frac{1}{3} x e^{x^2} \\ &= \left[\frac{1}{6} e^{x^2} \right]_{x=0}^3 \\ &= \frac{1}{6}(e^9 - 1) \end{aligned}$$

Example: Find the volume of the tetrahedron T enclosed by $x = 0, y = 0, z = 0$ and $2x + y + z = 4$ shown below



Notice the bottom face is a triangle with edges $x = 0$, $y = 0$, and hypotenuse $y = 2x - 4$ (keeping z constant). We may then write the tetrahedron as

$$T = \{(x, y, z) : 0 \le x \le 2,\ 0 \le y \le 4 - 2x,\ 0 \le z \le 4 - 2x - y\}$$

We will write

$$H = \{(x, y) : 0 \le x \le z,\ 0 \le y \le 4 - 2x\}$$

denoting the bottom face as a type 1 region, so that we may write

$$T = \{(x, y, z) : (x, y) \in H, 0 \le z \le 4 - 2x - y\}$$

also being a type 1 region. So we have

$$\begin{aligned}\iiint_T 1dv &= \iint_H \underbrace{\int_0^{4-2x-y} 1dz}_{\text{continuous}} dA \\ &= \int_0^2 \int_0^{4-2x} \int_0^{4-2x-y} 1dzdydx \\ &= \int_0^2 \int_0^{4-2x} 4-2x-y \\ &= \int_0^2 \left[(4-2x)y - \frac{1}{2}y^2\right]_{y=0}^{y=4-2x} dx \\ &= \int_0^2 \frac{1}{2}(4-2x)^2 dx \\ &= \frac{16}{3}\end{aligned}$$

Week 12 Change of Variables

12.1 Polar Co-ordinates

Remark: Recall for $f : A \to \mathbb{R}^m$ with $a \in A \subseteq \mathbb{R}^n$, the Jacobian of f at a is

$$Jf(a) = \det(Df(a))$$

Theorem. Change of Variables: Let $U \subseteq \mathbb{R}^n$ be open, and $A \subseteq U$ be a closed Jordan region. Let $f : A \to \mathbb{R}$ be continuous and let $\varphi \in C^1(U, \mathbb{R}^n)$.

Suppose there is a set $B \subseteq A$ with (1) $vol(B) = 0$, (2) φ is injective on $A \setminus B$, and (3) $J\varphi(a) \neq 0$ for all $a \in A \setminus B$. Suppose $f : \varphi(A) \to \mathbb{R}^n$ is continuous. Then $\varphi(A)$ is a Jordan region, f is integrable on $\varphi(A)$ and

$$\int_{\varphi(A)} f(x)dx = \int_A f(\varphi(x))|J\varphi(x)|dx$$

Proof. The proof is long and technical, what follows is only some cases which give reason to the claim. In the case where $n = 1$, $J\varphi(x) = \varphi'(x)$ and so this looks more like the u-substitution formula of first year calculus. We take the absolute since the injectivity of φ tells us φ' is either always positive or always negative (otherwise would eventually have φ achieve same value twice). □

Remark. Polar Co-ordinates: Recall polar co-ordinates are an alternative representation of points in $\mathbb{R}^2$ represented as (r, θ) for $r \geq 0$ and $0 \leq \theta < 2\pi$.

Note: Note that the function $\varphi \in C^1(\mathbb{R}^2, \mathbb{R}^2)$ given by $\varphi(r, \theta) = (r\cos\theta, r\sin\theta)$ converts from polar to cartesian co-ordinates. Further, φ is injective on

$$\mathbb{R}^2 \setminus \{(0, \theta) : 0 \leq \theta < 2\pi\}$$

and in particular

$$\{(0,\theta) : 0 \leq \theta < 2\pi\}$$

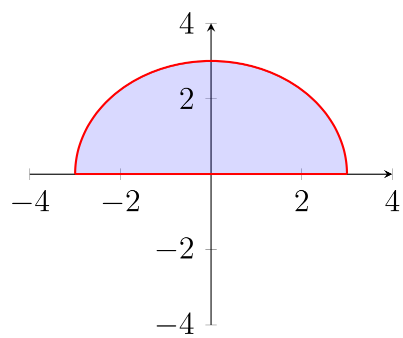
is a line and thus has volume zero. Finally notice

$$\begin{aligned} |J\varphi(r,\theta)| &= \left|\det\begin{pmatrix}\cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta\end{pmatrix}\right| \\ &= |r\cos^2\theta + r\sin^2\theta| \\ &= |r| = r \end{aligned}$$

This makes φ incredibly useful for integrals in polar co-ordinates because

$$\iint_{\varphi(D)} f(x,y)dA = \iint_D f(r\cos\theta, r\sin\theta)r dA$$

Example: Let D be the region bounded above by $x^2+y^2=9$ and below by the x-axis shown below



Compute $\displaystyle\iint_D \cos(x^2+y^2)dA$.

Notice D can equally be given by

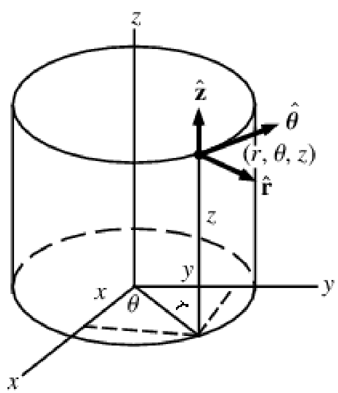
$$D = \{(r,\theta) : 0 \leq r \leq 3, 0 \leq \theta \leq \pi\}$$

So we have

$$\begin{aligned} \iint_D \cos(x^2+y^2)dA &= \iint_D \cos(r^2\cos^2\theta + r^2\sin^2\theta)r\ dA \\ &= \iint_D \cos(r^2)r\ dA \\ &= \int_0^\pi \int_0^3 \cos(r^2)r\ drd\theta \\ &= \int_0^\pi \left[\frac{1}{2}\sin(r^2)\right]_{r=0}^3 d\theta \\ &= \int_0^\pi \frac{1}{2}\sin 9 d\theta \\ &= \frac{\pi}{2}\sin 9 \end{aligned}$$

12.2 Cylindrical Co-ordinates

Remark. Cylindrical Co-ordinates: We can equally use polar co-ordinates in $\mathbb{R}^3$ by using polar co-ordinates to denote the location in the xy plane and keeping the z value constant. These are called cylindrical co-ordinates based on how one finds these co-ordinates. Below is an image representing a cylindrical co-ordinate system



Note: We may convert these cartesian co-ordinate by

$$\varphi(r, \theta, z) = (r\cos\theta, r\sin\theta, z)$$

We remark that

$$|J\varphi(r,\theta,z)| = \left|\det\begin{pmatrix}\cos\theta & -r\sin\theta & 0\\ \sin\theta & r\cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}\right| = r$$

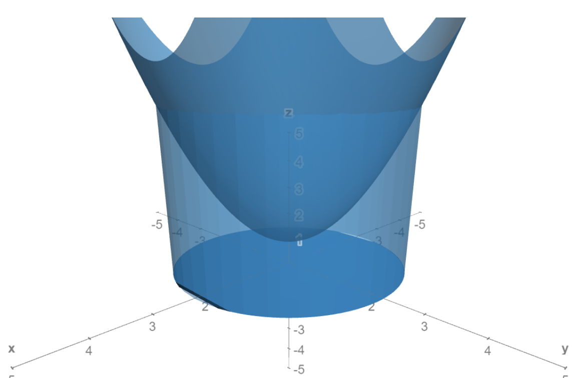
and that φ is injective everywhere except

$$\{(0,\theta,z) : 0 \le \theta < 2\pi, z \in \mathbb{R}\}$$

which is a region of volume zero. Hence

$$\iiint_{\varphi(A)} f(x,y,z)dV = \iiint_A f(r\cos\theta, r\sin\theta, z)r\ dV$$

Example: Let A be the region enclosed by (1) the paraboloid $z = 1 + x^2 + y^2$, (2) the cylinder $x^2 + y^2 = 5$, and (3) the xy plane as shown below



Compute $\iiint_A e^z dV$.

Notice we have

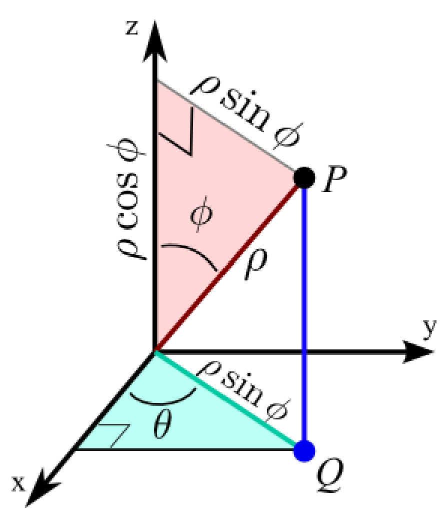
$$\begin{aligned} A &= \{(r,\theta,z) : 0 \le r \le \sqrt{5}, 0 \le \theta < 2\pi, 0 \le z \le x^2 + y^2\} \\ &= \{(r,\theta,z) : 0 \le r \le \sqrt{5}, 0 \le \theta < 2\pi, 0 \le z \le r^2\} \end{aligned}$$

Therefore,

$$\begin{aligned} \iiint_A e^z dV &= \int_0^{\sqrt{5}} \int_0^{2\pi} \int_0^{r^2} e^z r \, dz d\theta dr \\ &= \int_0^{\sqrt{5}} \int_0^{2\pi} \left[e^z r \right]_{z=0}^{z=1+r^2} d\theta dr \\ &= \int_0^{\sqrt{5}} \int_0^{2\pi} r e^{1+r^2} - r \, d\theta dr \\ &= 2\pi \int_0^{\sqrt{5}} r e^{1+r^2} - r \, dr \\ &= 2\pi \left[\frac{1}{2} e^{1+r^2} - \frac{1}{2} r^2 \right]_{r=0}^{\sqrt{5}} \\ &= 2\pi \left(\frac{1}{2} e^6 - \frac{5}{2} - \frac{1}{2} e \right) \\ &= \pi(e^6 - e - 5) \end{aligned}$$

12.3 Spherical Co-ordinates

Remark. Spherical Co-ordinates: We may also write co-ordinates in $\mathbb{R}^3$ using spherical co-ordinates, where we specify the angle relative to the xy axis, θ, and the angle relative to the zx axis, ϕ, of the point on a sphere of radius ρ. This gives us a tuple (ρ, θ, ϕ). A repreentation of a spherical co-ordinate system follows



Note: We may convert these to cartesian co-ordinate by the following equations

$$\begin{aligned} x &= r\cos\theta & r &= \rho\sin\phi \\ y &= r\sin\theta & z &= \rho\cos\phi \end{aligned}$$

Notice that $x^2 + y^2 + z^2 = \rho^2$. We may also use the formula

$$\varphi(\rho, \theta, \phi) = (\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)$$

We remark that

$$|J\varphi(\rho,\theta,\phi)| = \left|\det\begin{pmatrix} \sin\phi\cos\theta & -\rho\sin\phi\sin\theta & \rho\cos\phi\cos\theta \\ \sin\phi\sin\theta & \rho\sin\phi\cos\theta & \rho\cos\phi\sin\theta \\ \cos\phi & 0 & -\rho\sin\phi \end{pmatrix}\right| = \rho^2\sin\phi$$

and that φ is injective everywhere except

$$\{(0, \theta, \phi) : 0 \le \theta < 2\pi, 0 \le \phi < \pi\}$$

which is a region of volume zero. Hence

$$\iiint_{\varphi(A)} f(x,y,z)dV = \iiint_A f(\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)\rho^2\sin\phi\, dV$$

Example: Find the volume of a sphere delimited by

$$x^2 + y^2 + z^2 = a^2$$

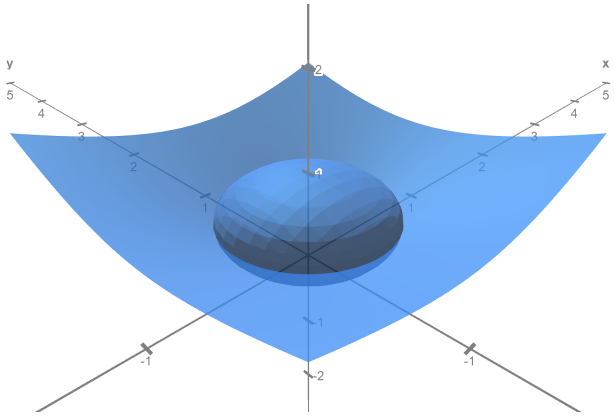
Notice this area can equally be given by

$$S = \{(\rho, \theta, \phi) : 0 \le \rho \le a, 0 \le \theta < 2\pi, 0 \le \phi < \pi\}$$

and so

$$\begin{aligned} Vol(S) &= \int_S 1dv \\ &= \iiint_S 1dxdydz \\ &= \int_0^a \int_0^{2\pi} \int_0^\pi \rho^2\sin\phi\, d\phi d\theta d\rho \\ &= \int_0^a \int_0^{2\pi} \Big[-\rho^2\cos\phi\Big]_{\phi=0}^{\pi} d\theta d\rho \\ &= \int_0^a \int_0^{2\pi} 2\rho^2\, d\theta d\rho \\ &= 2\pi \int_0^a 2\rho^2\, d\rho \\ &= 2\pi\left[\frac{2}{3}\rho^3\right]_{\rho=0}^{a} \\ &= \frac{4\pi}{3}a^3 \end{aligned}$$

Example: Find the volume of the solid which is bounded below by the cone $z = \sqrt{x^2 + y^2}$ and above by the sphere $x^2 + y^2 + z^2 = z$ shown below



Notice that

$$x^2 + y^2 + z^2 = z \iff x^2 + y^2 + \left(z - \frac{1}{2}\right)^2 = \frac{1}{2}$$

We can also write the cone as

$$\rho \cos \phi = \sqrt{\rho^2 \sin^2 \phi \cos^2 \theta + \rho^2 \sin^2 \phi \sin^2 \theta} = \rho \sin \phi$$

and so

$$C = \left\{(\rho, \theta, \phi) : \rho = 0 \vee \phi = \frac{\pi}{4}\right\}$$

The sphere can be written as as

$$\rho^2 = \rho \cos \phi$$

and so

$$S = \{(\rho, \theta, \phi) : \rho = 0 \vee \rho = \cos \phi\}$$

Letting D be the above solid, then its volume is

$$\begin{aligned}
\iiint_D 1 dV &= \int_0^{2\pi} \int_0^{\frac{\pi}{4}} \int_0^{\cos \phi} \rho^2 \sin \phi \; d\rho d\phi d\theta \\
&= \int_0^{2\pi} \int_0^{\frac{\pi}{4}} \left[\frac{1}{3}\rho^3 \sin \phi\right]_{\rho=0}^{\rho=\cos\phi} d\phi d\theta \\
&= \int_0^{2\pi} \int_0^{\frac{\pi}{4}} \frac{1}{3} \cos^3 \phi \sin \phi \; d\phi d\theta \\
&= \int_0^{2\pi} \left[-\frac{1}{12} \cos^4 \phi\right]_{\phi=0}^{\frac{\pi}{4}} \\
&= \int_0^{2\pi} \frac{1}{16} \\
&= \frac{\pi}{8}
\end{aligned}$$

Remark: All the above co-ordinate If z can have its upper and lower curves defined as a function of x and y, then it is often preferable to use cylindrical co-ordinates, otherwise spherical may be be better.

Index