MATH 247 by Camila Sehnem

Eason Li

2024 S

Contents

1 The Euclidean Inner Product and Distance in $\mathbb{R}^n$ 4

1.1 Standard Inner Product . . . 5
1.2 (Euclidean) Norm . . . 6
1.3 Cauchy-Schwartz Inequality . . . 6

2 Angles between Vectors in $\mathbb{R}^N$ 8

2.1 Orthogonal . . . 8

3 Topology on $\mathbb{R}^N$ - Open Sets and Closed Sets 8

3.1 Permanence Properties of Open Sets . . . 10
3.2 De Morgan's Law . . . 11

4 Sets that are neither closed not open 11

4.1 Cluster Point . . . 12
4.2 Characterization of Closed Sets . . . 13
4.3 Closure . . . 13
4.4 $\mathbb{R}^N$ is the Disjoint Union . . . 17

5 Compactness 17

5.1 Heine-Boul Theorem . . . 19
5.2 N-cell is Compact . . . 21

6 Connected Sets 23

6.1 Interval is Connected . . . 23
6.2 Higher-Dimensional Examples . . . 24
6.3 Convex is Connected . . . 24
6.4 Only $\mathbb{R}^N$ and $\varnothing$ are Clopen . . . 25

7 Sequence and Limits in $\mathbb{R}^N$ 26

7.1 Bounded if (Cauchy iff Convergent) . . . 27

8 Sequential Characterization of Compact Set 28

8.1 Compact and Sequential Compact (in Metric Space $\mathbb{R}^N$) . . . 29

9 Limits of Function and Continuity 31

9.1 Limit . . . 31
9.2 Continuity . . . 32
9.3 Properties of Continuous Functions . . . 33
9.3.1 Example and Application . . . 35
9.4 Continuity and Compactness . . . 36
9.4.1 Extreme Value Theorem . . . 37
9.5 Uniform Continuity . . . 37
9.6 Continuity and Connectedness . . . 38
9.6.1 Intermediate Value Theorem . . . 38

10 Differentiability on $\mathbb{R}^N$ **39**
10.1 Uniqueness of Derivative 40
10.2 Chain Rule 43
10.3 Partial Derivative 45
10.3.1 Geometrix Interpretation 45
10.4 Product Rule + Linearity 51
10.5 Higher Order Partial Derivatives 53
10.5.1 Partial Derivatives are Commutative 55

11 Vector Fields **56**
11.1 Other Operations on a Vector Field 57
11.2 Derivative as Linear Approximation 58

12 Taylor's Theorem **58**
12.1 Single Variable Taylor's Theorem 58
12.2 Multivariable Taylor's Theorem 60
12.3 Multivariate Polynomial 63
12.4 The Hessian 64
12.5 Critiacal Points 65

13 Local Properties of Continuously differentiable function **70**
13.1 Inverse Function Theorem 70
13.2 Implicit Function Theorem 74
13.2.1 The Linear Case 76

14 Integration on $\mathbb{R}^N$ **79**
14.1 Riemann Sum 81
14.2 Riemann Integrable 82
14.2.1 Cauchy Criterion for Riemann Integrable 82
14.3 Content Zero 85
14.4 Measure Zero 86
14.5 Lebesgue Theorem 88
14.5.1 Mean Value Theorem for Integration 91
14.6 Fubini's Theorem 92
14.7 Change of Variables 96
14.8 Integration with Cylindrical Coordinates 97
14.9 Spherical Coordinates 98

Lecture 1 - Monday, May 6

Recall that if $S_1, \dots, S_n$ are sets, then the **Cartesian Product** $S_1 \times \cdots \times S_n$, also denoted as $\prod_{i=1}^{n} S_i$, is the set

$$S_1 \times \cdots \times S_n = \{(x_1, \dots, x_n) \mid x_j \in S_j,\ j = 1, \dots, n\}$$

Definition 0.1: N-dimensional Euclidean Space, Vector

The **N-dimensional Euclidean Space** is the N-fold Cartesian product $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$. Element $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ is called a **vector** is simply a point in $\mathbb{R}^n$. The numbers $x_1, \dots, x_n$ are called the coordinates.

Recall that $\mathbb{R}^n$ is a vector space over $\mathbb{R}$ with coordinate-wise operations: that is, if $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, we have

$$\begin{aligned} x + y &= (x_1 + y_1, \dots, x_n + y_n) \\ \lambda x &= (\lambda x_1, \dots, \lambda x_n) \end{aligned}$$

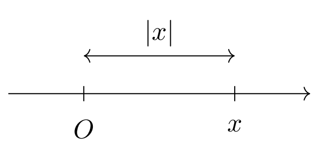
Definition 0.2: Zero Vector / Origin

The **zero vector**, or the **origin**, is the vector $\vec{0} = (0, \dots, 0)$.

1 The Euclidean Inner Product and Distance in $\mathbb{R}^n$

Example 1.1: Absolute Vector

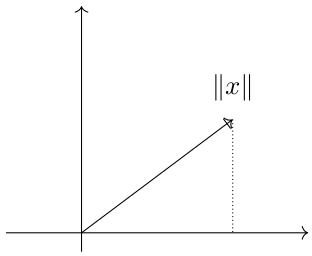
In $\mathbb{R}$, the distance of $x \in \mathbb{R}$ is from O in the **absolute vector**, $|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{otherwise} \end{cases}$.



For $x, y \in \mathbb{R}$, the distance of x and y is $|x - y|$.

Example 1.2: In $\mathbb{R}^2$

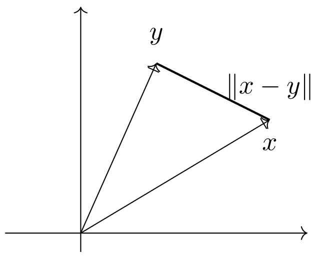
In $\mathbb{R}^2$, there is a natural notion of distance of a vector $x = (x_1, x_2)$ to 0.



$$\|x\| := \sqrt{x_1^2 + x_2^2}$$

For $x, y \in \mathbb{R}^2$, we can define the distance of x and y by

$$\|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$



We can find that $\|x - y\|^2 = \|x\|^2 + \|y\|^2$ if and only if the "dot product", $x \circ y = x_1 y_1 + x_2 y_2$ is zero, because $x \circ y = \|x\| \cdot \|y\| \cos\theta$ (follow from the law of cosine).

We extend this to $\mathbb{R}^n$

1.1 Standard Inner Product

Definition 1.1: Euclidean Inner Product (Dot Product)

The Euclidean inner product (or dot product) on $\mathbb{R}^N$ is the function

$$\begin{aligned} \circ : \mathbb{R}^N \times \mathbb{R}^N &\to \mathbb{R} \\ (x, y) &\to \sum_{i=1}^{N} x_i y_i \end{aligned}$$

Proposition 1.1

The dot product satisfies that for all $x, y \in \mathbb{R}^N$ and $\lambda \in \mathbb{R}$, the following holds:

1. $x \circ x \geq 0$
2. $x \circ x = 0$ if and only if $x = 0$
3. $x \circ y = y \circ x$
4. $x \circ (y + z) = x \circ y + x \circ z$
5. $(\lambda x) \circ y = \lambda(x \circ y)$

Result 1.1

Properties 3, 4 and 5 imply that $\circ$ is **bilinear**.

1.2 (Euclidean) Norm

Definition 1.2: Norm

For $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$, we define the (Euclidean) **norm** of x by

$$\|x\| = \sqrt{x \circ x} = \sqrt{\sum_{i=1}^{N} x_i^2}$$

Proposition 1.2

The function $\|\cdot\| : \mathbb{R}^N \to [0, \infty)$ satisfies

1. $\|x\| \geq 0$
2. $\|x\| = 0$ if and only if $x = 0$
3. $\|\lambda x\| = |\lambda| \, \|x\|$

We would also like to show that this satisfies the **triangle inequality**:

$$\|x + y\| \leq \|x\| + \|y\| \quad \text{for all } x, y \in \mathbb{R}^N$$

For this we need the Cauchy-Schwartz inequality.

1.3 Cauchy-Schwartz Inequality

Theorem 1.1: Cauchy-Schwartz

For all $x, y \in \mathbb{R}^N$ we have

$$|x \circ y| \leq \|x\| \cdot \|y\|$$

Moreover, equality holds if and only if $x = ty$ or $y = tx$ for some $t \in \mathbb{R}$.

Proof. We may assume that both x and y are non-zero. For all $t \in \mathbb{R}$, we know that

$$(x - ty) \circ (x - ty) \geq 0$$

then we have

$$p(t) = x \circ x - 2t(x \circ y) + t^2(y \circ y) \geq 0$$

Notice that this is a quadratic function of t, which implies that $p(t)$ has at most one root, thus

$$\Delta = \left[2(x \circ y)^2\right] - 4(x \circ x)(y \circ y) \leq 0$$

and the remaining follows naturally. $\square$

Corollary 1.1: Triangle Inequality

For all $x, y \in \mathbb{R}^N$ we have

$$\|x + y\| \leq \|x\| + \|y\|$$

Proof. We simply have

$$\begin{aligned} \|x + y\|^2 &= (x + y) \circ (x + y) \\ &= \|x\|^2 + \|y\|^2 + 2(x \circ y) \\ &\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \\ &= (\|x\| + \|y\|)^2 \end{aligned}$$

thus completing the proof. $\square$

Lecture 2 - Wednesday, May 8

Theorem 1.2: Properties of the Euclidean Norm

The Euclidean norm $\|\cdot\| : \mathbb{R}^N \to [0, \infty)$ satisfies the following for all $x, y \in \mathbb{R}^N$ and $\lambda \in \mathbb{R}$:

1. Proposition 1.2
2. Triangle inequality
$$\|x + y\| \leq \|x\| + \|y\|$$
3. Reversed triangle inequality
$$|\, \|x\| - \|y\| \,| \leq \|x - y\|$$

Proof. exercise. $\square$

Definition 1.3: Distance

For $x, y \in \mathbb{R}^N$, define the distance of x and y by

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

Notice that for all $z \in \mathbb{R}^N$,

$$d(x, y) \leq d(x, z) + d(z, y)$$

which is a direct consequence of the Triangle Inequality 1.1.

2 Angles between Vectors in $\mathbb{R}^N$

In $\mathbb{R}^2$, we know that $x \circ y = \|x\| \, \|y\| \cos\theta$, where θ is the angle between x and y.

In $\mathbb{R}^N$, Cauchy-Schwartz inequality 1.1 implies that for $x, y \neq 0$, then

$$\frac{x \circ y}{\|x\| \, \|y\|} \in [-1, 1]$$

we can find a unique $\theta \in [0, \pi]$ such that

$$\cos\theta = \frac{x \circ y}{\|x\| \, \|y\|}$$

Definition 2.1: Angle between x and y

We define the **angle between** x **and** y as θ.

2.1 Orthogonal

Definition 2.2: Orthogonal

We say x and y are **orthogonal** if $\theta = \pi/2$.

3 Topology on $\mathbb{R}^N$ - Open Sets and Closed Sets

In topology, we study the notion of **closeness** (limits, convergence, continuity, etc.) through the collection of open sets / closed sets.

Definition 3.1: Open Ball and Closed Ball

The **open ball** in $\mathbb{R}^N$ of radius $r > 0$ centered at $x \in \mathbb{R}^N$ is the set

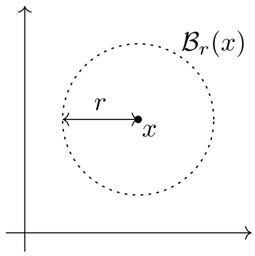
$$\mathcal{B}_r(x) = \{y \in \mathbb{R}^N \ : \ \|x - y\| < r\}$$

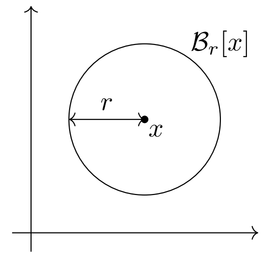
Remark: the other notation is $\mathcal{B}(x, r)$.
The **closed ball** in $\mathbb{R}^N$ of radius $r > 0$ centered at $x \in \mathbb{R}^N$ is the set

$$\mathcal{B}_r[x] = \{y \in \mathbb{R}^N \ : \ \|x - y\| \leq r\}$$

Example 3.1

1. In $\mathbb{R}$, $\mathcal{B}_r(x)$ is the open interval $(x - r, x + r)$. Similarly, $\mathcal{B}_r[x]$ is the closed interval $[x - r, x + r]$.
2. In $\mathbb{R}^2$, we have





Definition 3.2: Open Set and Closed Set

1. We say that $U \subseteq \mathbb{R}^N$ is **open** if for all $x \in U$, there exists $\varepsilon > 0$ (depending on x) such that $\mathcal{B}_\varepsilon(x) \subseteq U$.

2. We say that $F \subseteq \mathbb{R}^N$ is **closed** if its complement,

$$F^c = \{y \in \mathbb{R}^N \ : \ y \notin F\},$$

is open.

Result 3.1: "Clopen"

Notice that $\varnothing$ and $\mathbb{R}^N$ are open; and they are also closed. They are known as **clopen**.

Proposition 3.1: Open Balls are Open, and Vice Versa

1. The open ball $\mathcal{B}_r(x)$ is open.

2. The closed ball $\mathcal{B}_r[x]$ is closed.

Proof. The proof consists of two parts:
(Part 1):
Let $y \in \mathcal{B}_r(x)$, we want to find $\varepsilon > 0$ such that $\mathcal{B}_\varepsilon(y) \subseteq \mathcal{B}_r(x)$. We know that for $z \in \mathbb{R}^N$

$$d(x, z) \ \leq \ d(x, y) + d(y, z)$$

hence we can take $\varepsilon = r - d(x, y)$, then $\varepsilon > 0$ and $\mathcal{B}_\varepsilon(y) \subseteq \mathcal{B}_r(x)$.
(Part 2):
Use the Reversed Triangle Inequality:

$$|\,\|x - z\|\,| = \|x - y + y - z\| \geq |\,\|x - y\| - \|z - y\|\,|$$

We want to show that

$$\mathcal{B}_r[x]^c = \{y \in \mathbb{R}^N \ : \ \|y - x\| > r\}$$

is open. Choose y such that $\|y - x\| > r$. Let $\varepsilon = \|x - y\| - r$, so $\varepsilon > 0$. Also let $z \in \mathcal{B}_\varepsilon(y)$, then we have

$\|z-y\| < \varepsilon$, which implies that $-\|z-y\| > -\varepsilon = r - \|x-y\|$. Therefore,

$$\begin{aligned} \|x-z\| &\geq |\, \|x-y\| - \|y-z\| \,| \\ &= |\, \|x-y\| - \|z-y\| \,| \\ &> \|x-y\| + r - \|x-y\| \\ &= r \end{aligned}$$

Hence $z \in \mathcal{B}_r[x]^c$ is needed wich means that $\mathcal{B}_\varepsilon(y) \subseteq \mathcal{B}_r[x]^c$. □

3.1 Permanence Properties of Open Sets

Theorem 3.1: Permanence Properties of Open Sets

1. The union of an arbitrary collection of open sets is open.
   Precisely, if Λ are indices and $\{E_\alpha \mid \alpha \in \Lambda\}$ are open sets, then

$$E \equiv \bigcup_{\alpha \in \Lambda} E_\alpha$$

   is open.

2. The intersect of a *finite* collection of open sets is open.

Proof. 1. Let $x \in E$, then there exists $\alpha \in \Lambda$ such that $x \in E_\alpha$. Since E_α is open, then there exists some $\varepsilon > 0$ such that

$$\mathcal{B}_\varepsilon(x) \subseteq E_\alpha \subseteq \bigcup_{\alpha \in \Lambda} E_\alpha = E$$

which implies that E is also open

2. Let $E_1, E_2, \ldots, E_m$ be open sets in $\mathbb{R}^N$ and we let $E \equiv \bigcap_{i=1}^m E_i$. Let $x \in E$. For $i = 1, \ldots, m$, we can find $\varepsilon_i > 0$ such that $\mathcal{B}_\varepsilon(x) \subseteq E_i$. So we can set $\varepsilon \equiv \min\{\varepsilon_i \ : \ i = 1, \ldots, m\}$. Then

$$\mathcal{B}_\varepsilon(x) \subseteq \bigcap_{i=1}^{m} E_i = E$$

giving that E is open.

□

Lecture 3 - Friday, May 10

Example 3.2

The intersection of an infinite collection of open sets need not to be open, Consider that for all $m \geq 1$. take $E_m \equiv \mathcal{B}_{1/m}(n)$, then E_m is open, but the intersect is a single point n, which is indeed closed.

3.2 De Morgan's Law

Theorem 3.2: De Morgan's Law

Let $\{E_\alpha \;:\; \alpha \in \Lambda\}$ be a collection of subsets of a set A, then

$$\left(\bigcup_{\alpha\in\Lambda} E_\alpha\right)^c = \bigcap_{\alpha\in\Lambda} E_\alpha^c$$
$$\left(\bigcap_{\alpha\in\Lambda} E_\alpha\right)^c = \bigcup_{\alpha\in\Lambda} E_\alpha^c$$

Corollary 3.1: Properties of Closed Sets

1. The intersection of an arbitrary collection of closed sets is closed
2. The union of a finite collection of closed sets is closed

Proof. This follows the De Morgan's Law 3.2. □

Example 3.3

The sphere

$$\partial\mathcal{B}_r(x) = \{y \in \mathbb{R}^N \;:\; \|y - x\| = r\}$$

is closed because

$$\partial\mathcal{B}_r(x) = \mathcal{B}_r[x] \cap \mathcal{B}_r(x)^c$$

Example 3.4

The union of an infinite collection of closed sets need not be closed: Take $F_m = \{1/m\}$ (i.e. $(1/m, \ldots, 1/m) \in \mathbb{R}^N$), then F_m is closed,
Exercise: Show that $\bigcup_{m=1}^{\infty} F_m$ is not closed.

Proof. To show that $\bigcup_{m=1}^{\infty} F_m$ is not closed, it suffices to show that the complement is not open. Consider the point $\mathcal{O} = \{0\}$, we can easily find that we are not able to construct an open ball that is contained in the complement, thus completing the proof. □

4 Sets that are neither closed not open

Discovery 4.1

In general, an arbitrary subset S of $\mathbb{R}^N$ need not be closed nor open.

Example 4.1

In $\mathbb{R}$, consider $(a, b]$.

Example 4.2

Let

$$S \equiv \{(x, y, z) \in \mathbb{R}^3 \ : \ y^2 + z^2 = 1, x > 0\}$$

then S is neither closed nor open.

Proof. 1. (not open)
Take $p = (1, 0, 1) \in S$, then for $\varepsilon > 0$, we claim that $\mathcal{B}_\varepsilon(p) \cap S^c \neq \varnothing$ (i.e. there are points in the open ball around p but not in S). We can simple set the point to be $q = (1, 0, 1 + \varepsilon/2)$.

2. (not closed)
Take $p = (0, 0, 1) \in S^c$, given that $\varepsilon > 0$, we want to show that S^c is not open. Take $q = (\varepsilon/2, 0, 1)$, then $q \in S$ and $q \in \mathcal{B}_\varepsilon(p)$, so $\mathcal{B}_\varepsilon(p) \cap S \neq \varnothing \ \Rightarrow \ \mathcal{B}_\varepsilon(p) \not\subseteq S^c$.

□

4.1 Cluster Point

Definition 4.1: Cluster Point

1. A point $p \in \mathbb{R}^N$ is called a **cluster point** (or accumulation point) of S if for every $\varepsilon > 0$, we have

$$(\mathcal{B}_\varepsilon(p) \backslash \{p\}) \cap S \neq \varnothing$$

Equivalently, for every open set U with $p \in U$, there exists $x \in S \cap U$ and $x \neq p$.

2. We denote by S' the set of all cluster points of S.

Example 4.3: Every $p \in \mathbb{R}^N$ is a cluster point of $\mathbb{Q}^N$

Every $p \in \mathbb{R}^N$ is a cluster point of $\mathbb{Q}^N = \{(q_1, \ldots, q_N) \in \mathbb{R}^N \ : \ q_i \in \mathbb{Q}, \ i = 1, \ldots, N\}$.

Proof. To see this, let $p = (p_1, \ldots, p_N) \in \mathbb{R}^N$ and $\varepsilon > 0$. By density of $\mathbb{Q}$ in $\mathbb{R}$, for each $i = 1, \ldots, N$, we can find $c_i \in \mathbb{Q}$, $c_i \neq p_i$ such that $|p_i - c_i| < \varepsilon/\sqrt{N}$, set $c = (c_1, \ldots, c_N) \in \mathbb{Q}^N$, then

$$\|p - c\| = \sqrt{\sum_{i=1}^{N} (p_i - c_i)^2} < \varepsilon$$

and $p \neq c$. Hence $c \in (\mathcal{B}_\varepsilon(p) \backslash \{p\}) \cap \mathbb{Q}^N$ is needed. □

Example 4.4

Let S be a finite set,

$$S = \{x_1, \ldots, x_N\} \in \mathbb{R}^N$$

then S has no cluster point.

Proof. To see this, take $p \in \mathbb{R}^N$ and $\varepsilon > 0$ with

$$\varepsilon < \min\{\|p - x\| \ : \ x \in S, x \neq p\}$$

□

4.2 Characterization of Closed Sets

Theorem 4.1: Characterization of Closed Sets

Let $F \subseteq \mathbb{R}^N$, TFAE

1. F is closed
2. $F' \subseteq F$

Proof.

1. $(1 \Rightarrow 2)$
   Suppose F is closed. Let $p \in F^c$, we have to show that $p \notin F'$. Since F is closed, F^c is open, hence there exists $\varepsilon > 0$ such that $\mathcal{B}_\varepsilon(p) \subseteq F^c$. In particular,

   $$\mathcal{B}_\varepsilon(p) \cap F = \varnothing$$

   giving that $p \notin F'$, we have $F' \subseteq F$.

2. $(2 \Rightarrow 1)$
   Suppose $F' \subseteq F$, we will show that F^c is open. Take $p \in F^c$, then $p \notin F'$, so there exists $\varepsilon > 0$ such that

   $$(\mathcal{B}_\varepsilon(p) \backslash \{p\}) \cap F = \varnothing$$

   Thus $\mathcal{B}_\varepsilon(p) \cap F = \varnothing$. Since $p \in F^c$, so $\mathcal{B}_\varepsilon(p) \subseteq F^c$ and thus F^c is open.

□

4.3 Closure

Definition 4.2: Closure

Let $S \subseteq \mathbb{R}^N$, define the **closure** of S by $\overline{S} = S \cup S'$.

Proposition 4.1

Let $S \subseteq \mathbb{R}^N$. Then

$$S' = \overline{S}'$$

In particular, we have $\overline{S}$ is closed.

Corollary 4.1

$\overline{S}$ is the smallest closed set contans S. i.e. if $S \subseteq F$ and F is closed, then $\overline{S} \subseteq F$.

$$\overline{S} = \bigcap_{\substack{F \supseteq S \\ F \text{ open}}} F$$

Definition 4.3: Boundary and Interior

Let $S \subseteq \mathbb{R}^N$,

1. We say that a point $p \in \mathbb{R}^N$ is a **boundary point** of S if for every $\varepsilon > 0$, we have

$$\mathcal{B}_\varepsilon(p) \cap S \neq \varnothing \qquad \& \qquad \mathcal{B}_\varepsilon(p) \cap S^c \neq \varnothing$$

   The **boundary**, ∂S, is the set of all boundary points of S.

2. We say that a point $p \in \mathbb{R}^N$ is an **interior point** of S if there exists $\varepsilon > 0$ such that $\mathcal{B}_\varepsilon(p) \subseteq S$. The **interior** of S, denoted by S°, is the set of all interior points of S.

Result 4.1

We have

$$S^\circ \subseteq S \subseteq \overline{S}$$

Example 4.5

Let $S = (0,1] \cup \{2\}$, we have

$$\begin{aligned} \partial S &= \{0, 1, 2\} \\ S' &= [0,1] \\ S^\circ &= (0,1) \\ \overline{S} &= [0,1] \cup \{2\} \end{aligned}$$

Proposition 4.2

Let $x \in \mathbb{R}^N$ and $r > 0$, then

1. $\partial\mathcal{B}_r(x) = \partial\mathcal{B}_r[x] = \{y \in \mathbb{R}^N \ : \ \|y - x\| = r\}$
2. $\overline{\mathcal{B}_r(x)} = \mathcal{B}_r[x]$

Proof. 1. Let $y \in \mathbb{R}^N$ with $\|y - x\| = r$. It suffices to show that for all $\varepsilon > 0$,

$$\mathcal{B}_\varepsilon(y) \cap \mathcal{B}_r(x) \neq \varnothing \quad \& \quad \mathcal{B}_\varepsilon(y) \cap \mathcal{B}_r[x]^c \neq \varnothing$$

since $\mathcal{B}_r(x)$ and $\mathcal{B}_r[x]^c$ are open. Let $\lambda > 0$, so we have

$$\|\lambda(y - x)\| = \lambda \|y - x\| = \lambda r$$

Set $z_\lambda = x + \lambda(y - x)$. Notice that if $\lambda < 1$, then $z_\lambda \in \mathcal{B}_r(x)$, and if $\lambda > 1$, then $z_\lambda \in \mathcal{B}_r[x]^c$. Take $0 < \lambda < 1$ with $1 - \lambda < \varepsilon/r$, then $z_\lambda \in \mathcal{B}_r(x)$ and

$$\begin{aligned} \|z_\lambda - y\| &= \|x + \lambda(y - x) - y\| \\ &= (1 - \lambda) \|y - x\| \\ &< \frac{\varepsilon}{r} \cdot r = \varepsilon \end{aligned}$$

To get $z_\lambda \in \mathcal{B}_\varepsilon(y) \cap \mathcal{B}_r[x]^c$, take $\lambda > 0$ with $\lambda - 1 < \varepsilon/r$, then $z_\lambda \in \mathcal{B}_r[x]^c$ and is above $z_\lambda \in \mathcal{B}_\varepsilon(y)$.

2. We know that

$$\overline{\mathcal{B}_r(x)} = \mathcal{B}_r(x) \cup \mathcal{B}_r(x)'$$

If $p \in \mathcal{B}_r[x]^c$, then $p \notin \mathcal{B}_r(x)'$, so

$$\overline{\mathcal{B}_r(x)} \subseteq \mathcal{B}_r[x]$$

By part a), if $p \in \mathbb{R}^N$ and $\|p - x\| = r$, then $p \in \partial\mathcal{B}_r(x)$ and hence $p \in \mathcal{B}_r(x)$, thus

$$\mathcal{B}_r[x] \subseteq \overline{\mathcal{B}_r(x)}$$

□

Proposition 4.3

Let $S \subseteq \mathbb{R}^N$, then

1. S° is open, and

$$S^\circ = \bigcup_{\substack{U \subseteq S \\ U \text{ open}}} U$$

2. $S^\circ = S \backslash \partial S$

Proof. 1. Let $x \in S^\circ$, since x is an interior point, so we can find $\varepsilon_x > 0$ such that

$$\mathcal{B}_{\varepsilon_x}(x) \subseteq S$$

If $y \in \mathcal{B}_{\varepsilon_x}(x)$, then there exists $\delta > 0$ such that

$$\mathcal{B}_\delta(y) \subseteq \mathcal{B}_{\varepsilon_x}(x) \subseteq S$$

So y is also an interior point. This gives that

$$\mathcal{B}_{\varepsilon_x}(x) \subseteq S^\circ$$

This shows that S° is open, and

$$S^\circ = \bigcup_{x \in S^\circ} \mathcal{B}_{\varepsilon_x}(x) \subseteq \bigcup_{\substack{U \subseteq S \\ U \text{ open}}} U$$

Now let $U \subseteq S$, U open and let $x \in U$. Since U is open, there exists $\varepsilon > 0$ such that

$$\mathcal{B}_\varepsilon(x) \subseteq U \subseteq S$$

suggesting that $x \in S^\circ$, hence completes the proof.

2. Let $x \in S^\circ$, we want to show that $x \notin \partial S$. We know there exists $\varepsilon > 0$ such that

$$\mathcal{B}_\varepsilon(x) \subseteq S$$

hence we have

$$\mathcal{B}_\varepsilon(x) \cap S^c = \varnothing \quad \implies \quad S^\circ \subseteq S \backslash \partial S$$

On the other hand, let $x \in S \backslash \partial S$, hence we can find $\varepsilon > 0$ such that

$$\mathcal{B}_\varepsilon(x) \cap S^c = \varnothing \quad \implies \quad x \in S^\circ$$

□

Lecture 5 - Wednesday, May 15

Discovery 4.2

S° is the largest open set contained in S.

4.4 $\mathbb{R}^N$ is the Disjoint Union

Theorem 4.2

Let $S \subseteq \mathbb{R}^N$, then $\mathbb{R}^N$ is the disjoint union

$$\mathbb{R}^N = S^\circ \sqcup \partial S \sqcup (S^c)^\circ$$

Remark: The symbol $\sqcup$ implies that this is a disjoint union.

Proof. Clearly $S^\circ \cap (S^c)^\circ = \varnothing$ since $S^\circ \subseteq S$ and $S^{c\circ} \subseteq S^c$, and if $p \in S^\circ \cup (S^c)^\circ$, then $p \notin \partial S$, thus the above union is disjoint. To see that $\mathbb{R}^N = S^\circ \cup \partial S \cup (S^c)^\circ$, let $x \in \mathbb{R}^N$, if $x \in S^\circ \cup (S^c)^\circ$, we are done. Otherwise given $\varepsilon > 0$, we have $\mathcal{B}_\varepsilon(x) \cap S^c \neq \varnothing$ because $x \notin S^\circ$ and $\mathcal{B}_\varepsilon(x) \cap S \neq \varnothing$ because $x \notin (S^c)^\circ$. Since ε is arbitrary, thus we have $x \in \partial S$. ☐

Corollary 4.2

For any $S \subseteq \mathbb{R}^N$, we have

$$\overline{S} = S \cup \partial S$$

Proof. Exercise. ☐

5 Compactness

Compactness is an important concept in topology especially in connection with continuity.

Definition 5.1: Open Cover, Compact

1. Let $S \subseteq \mathbb{R}^N$. An **open cover** of S is a collection, $g = \{g_\alpha\}_{\alpha \in \Lambda}$, of open subsets of $\mathbb{R}^N$ that covers S. i.e.

$$S \subseteq \bigcup_{\alpha \in \Lambda} g_\alpha$$

2. We say that $K \subseteq \mathbb{R}^N$ is **compact** if every open cover $g = \{g_\alpha\}_{\alpha \in \Lambda}$ of K admits a finite subcover. i.e. there exists a finite subcollection $g' = \{g_{\alpha_i} : i = 1, \ldots, n\}$ of sets from g such that

$$K \subseteq \bigcup_{i=1}^{n} g_{\alpha_i}$$

Example 5.1: Finite Sets Are Compact

If $S = \{x_1, \ldots, x_n\}$ is finite, then S is compact.

Proof. Let $g = \{g_\alpha\}_{\alpha \in \Lambda}$ be an open cover of S. Since $S \subseteq \bigcup_{\alpha \in \Lambda} g_\alpha$, for each $i = 1, \ldots, n$, we can find $\alpha_i \in \Lambda$ such that $x_i \in g_{\alpha_i}$. Set $g' = \{g_{\alpha_i} : i = 1, \ldots, n\}$, then g' is a finite collection of sets from g that cover S, thus S is indeed compact. ☐

Example 5.2: Open Balls Are Not Compact

Let $r > 0$, $x \in \mathbb{R}^N$, then $\mathcal{B}_r(x)$ is not compact.

Proof. We need to exhibit an open cover $g = \{g_\alpha\}_{\alpha\in\Lambda}$ that admits no finite subcover. Let $k \geq 0$ be such that $1/k < r$. For each $m \geq k$, we set $g_m = \mathcal{B}_{r-1/m}(x)$. Then each g_m is open and we set $g = \{g_m\}_{m\geq k}$. Then g is a open cover of $\mathcal{B}_r(x)$. We claim that g admits no finite subcover. SFAC that $g' = \{g_{m_i} : i = 1, \ldots, l\}$ is a finite subover for $\mathcal{B}_r(x)$. Let j be such that $m_j = \max\{m_i : i = 1, \ldots, l\}$. Then

$$\mathcal{B}_r(x) \subseteq g_{m_j} = \mathcal{B}_{r-1/m_j}(x)$$

which is a contradiction because if $u \in \mathbb{R}^N$, $\|u\| = 1$, and let $r - 1/m_j < q < r$, then $z = x + qu$, $z \in \mathcal{B}_r(x)$, but $z \notin \mathcal{B}_{r-1/m_j}(x)$. □

Proposition 5.1

Suppose $K \subseteq \mathbb{R}^N$ is compact and $F \subseteq K$ is closed, then F is compact.

Proof. Let $g = \{g_\alpha\}_{\alpha\in\Lambda}$ be an arbitrary open cover of F, then

$$K \subseteq F \cup F^c \subseteq \left(\bigcup_{\alpha\in\Lambda} g_\alpha\right) \cup F^c$$

so $\overline{g} = g \cup \{F^c\}$ is an open cover of K because F^c is open. Since K is compact, $\overline{g}$ admits a finite subcover $\overline{g}' = \{g_{\alpha_i} : i = 1, \ldots, n\}$. Now

$$\begin{aligned} F = F \cap K &\subseteq F \cap \left(\bigcup_{i=1}^n g_{\alpha_i}\right) \\ &= \bigcup_{i=1}^n F \cap g_{\alpha_i} \\ &\subseteq \bigcup_{g\in\overline{g}', g\neq F^c} g \end{aligned}$$

Setting $g' = \overline{g}' \backslash \{F^c\}$, we see that g' is a finite subcover of F containing of sets from g. □

Definition 5.2: Bounded

We will say that a set $S \subseteq \mathbb{R}^N$ is **bounded** if there exists $m \geq 1$ such that

$$S \subseteq \mathcal{B}_m[0]$$

Theorem 5.1

Suppose $K \subseteq \mathbb{R}^N$ is compact, then K is closed and bounded.

Proof. Suppose K is compact

1. *Bounded*:
   For each $m \geq 1$, let $g_m = \mathcal{B}_m(0)$, then $g_m \subseteq g_{m+1}$, and g_m is open. Let $g = \{g_m\}_{m\geq 1}$, then g is now an open cover of K. By compactness of K, g admits a fintie subcover $g' = \{g_{m_i} : i = 1, \dots, l\}$. Let j be such that $m_j = \max\{m_i : i = 1, \dots, l\}$. Then $K \subseteq g_{m_j} \subseteq \mathcal{B}_{m_j}[0]$.

2. *Closed*:
   For each $x \in K^c$, we need to find $\varepsilon > 0$ such that $\mathcal{B}_\varepsilon(x) \subseteq K^c$. For each $y \in K$, we set $\varepsilon_y = \|x - y\|/2$, then $\varepsilon_y > 0$ because $x \notin K$. By the reverse triangle inequality, we have

   $$\mathcal{B}_{\varepsilon_y}(x) \cap \mathcal{B}_{\varepsilon_y}(y) = \varnothing$$

   For each $y \in K$, we set $g_y = \mathcal{B}_{\varepsilon_y}(y)$ and let

   $$g := \{g_y : y \in K\}$$

   Then g is an open cover of K. By compactness, we can find a finite subcover from g, say $g' = \{g_{y_j} : j = 1, \dots, n\}$. Let $\varepsilon = \min\{\varepsilon_{y_j} : j = 1, \dots, n\}$.

□

Lecture 6 - Friday, May 17

Discovery 5.1

If $F \subseteq \mathbb{R}^N$ is closed and $K \subseteq \mathbb{R}^N$ is compact (so it's also closed and bounded), then $F \cap K$ is compact, since $F \cap K$ is closed (3.1) and $F \cap K \subseteq K$ (5.1).

Theorem 5.2

If $E \subseteq K$ is an infinite set and K is compact, then E has a cluster point in K.

Proof. SFAC that E has no cluster point in K. Since $E \subseteq K$, by A01-Q4, we have

$$E' \subseteq K' \subseteq K$$

because K is closed. Thus we must have $E' = \varnothing$. Then E is closed since $E' = \varnothing \subseteq E$. It follows that E is compact (5.1). Now if $p \in E$, it is clear that $p \notin E'$, so we get $\varepsilon_p > 0$ such that

$$\mathcal{B}_{\varepsilon_p}(p) \cap E = \{p\}$$

Then the open cover $\{\mathcal{B}_{\varepsilon_p}(p) : p \in E\}$ admits no finite subcover because E is infinite. □

5.1 Heine-Boul Theorem

We wish to prove the converse, that is, we want to show that if $K \sqsubseteq \mathbb{R}^N$ is closed and bounded, then K is compact.

Theorem 5.3: Nested Interval Principle

Recall the nested interval principle:

If $I_m = [a_m, b_m] \subseteq \mathbb{R}$ is a nested sequence of closed and bounded intervals in $\mathbb{R}$, then

$$\bigcap_{m=1}^{\infty} I_m \neq \varnothing$$

i.e. $I_m \supseteq I_{m+1} \supseteq \cdots$ for all m. Moreover, if $\lim_m (b_m - a_m) = 0$, then

$$\bigcap_{m=1}^{\infty} = \{z\}$$

is a single point.

Definition 5.3: N-cell

For each $j = 1, \ldots, N$, let $a_j, b_j \in \mathbb{R}$ with $a_j < b_j$. We call the Cartesian product

$$I = [a_1, b_1] \times \cdots \times [a_N, b_N]$$

an N-cell.

Theorem 5.4

Let $I_1 \supseteq I_2 \supseteq \cdots$ be an increasing sequence of N-cells, then

$$\bigcap_{m=1}^{\infty} I_m \neq \varnothing$$

Moreover, if $\lim_m \|b_m - a_m\| = 0$, then

$$\bigcap_{m=1}^{\infty} I_m = \{z\}$$

is a single point, where here $a_m, b_m \in \mathbb{R}^N$ and $I_m = [a_{m,1}, b_{m,1}] \times \cdots \times [a_{m,N}, b_{m,N}]$.

Proof. Since $I_m \supseteq I_{m+1}$, we have

$$[a_{m,j}, b_{m,j}] \supseteq [a_{m+1,j}, b_{m+1,j}]$$

By nested interval principle in $\mathbb{R}$, there exists

$$z_j \in \bigcap_{m=1}^{\infty} [a_{m,j}, b_{m,j}] \qquad j = 1, \ldots, N$$

We set $z = (z_1, \ldots, z_N)$, then $z \in \bigcap_{m=1}^{\infty} I_m$. If $\lim_{m\to\infty} \|b_m - a_m\| = 0$, then since $(b_{m,j} - a_{m,j}) \leq$

$\|b_m - a_m\|$, we deduce that $\lim_{n\to\infty}(b_{m,j} - a_{m,j}) = 0$. Hence

$$\bigcap_{m=1}^{\infty} [a_{m,j}, b_{m,j}] = \{z_j\}$$

Then

$$\bigcap_{m=1}^{\infty} I_m = \{z\}$$

□

5.2 N-cell is Compact

Theorem 5.5

Let $I = [a_1, b_1] \times \cdots [a_N, b_N]$ be an N-cell, then I is compact.

Theorem 5.6: Heine-Boul Theorem

Let $K \subseteq \mathbb{R}^N$, then TFAE

1. K is compact,
2. K is closed and bounded.

Proof. We have shown that compact implies closed and bounded (5.1). Therefore it suffices to show the other direction: Suppose K is closed and bounded. Since it is bounded, we find that there exists some $M > 0$ such that $K \subseteq \mathcal{B}_M[0]$. Then if $x \in K$, we have $|x_j| \leq \|x_j\| \leq M$ and so K is contained in the N-cell

$$I_M = \underbrace{[-M, M] \times \cdots \times [-M, M]}_{N \text{ terms}}$$

By the previous theorem 5.5, I_m is compact, and because $K \subseteq I_M$ and K is closed, thus K is compact (5.1). □

Lecture 7 - Tuesday, May 21

LMAO Camila didn't show up to the class today.

Lecture 8 - Wednesday, May 22

Proof. This is the proof of Theorem (5.5).
Let $a = (a_1, \ldots, a_N)$ and $b = (b_1, \ldots, b_N)$. Set

$$\delta = \|b - a\| = \sqrt{\sum_{i=1}^{N}(b_i - a_i)^2}$$

Notice that if $x, y \in I$, then $\|x - y\| \leq \delta$. SFAC that I is not compact, then there exists an open cover $g = \{g_\alpha\}_{\alpha\in\Lambda}$ for I that admits no finite subcover.

1. **Step 1**:
   For each $j = 1, \ldots, N$, let $c_j = \frac{a_j+b_j}{2}$. Then the intervals $[a_j, c_j], [c_j, b_j]$ gives 2^N N-cells,

   $$J_1 = \{I_{1,l} : l = 1, \ldots, 2^N\}$$

   such that $I = \bigcup_{l=1}^{2^N} I_{1,l}$, where each N-cell $I_{1,l}$ is the Cartesian product

   $$[d_1, e_1] \times \cdots \times [d_N, e_N]$$

   with

   $$[d_j, e_j] \in \{[a_j, c_j], [c_j, b_j]\}$$

   It follows that there is some $l \in \{1, \ldots, 2^N\}$ such that the N-cell $I_{1,l}$ cannot be covered by a finite collection of sets from g. Let I_1 be such an N-cell. Notice that

   (a) $I \supseteq I_1$

   (b) I_1 cannot be covered by a finite collection of sets from g

   (c) Let $a_1 = (a_{11}, \ldots, a_{1N})$ and $b_1 = (b_{11}, \ldots, b_{1N})$ be such that

   $$I_1 = [a_{11}, b_{11}] \times \cdots \times [a_{1N}, b_{1N}]$$

   then if $x, y \in I_1$,

   $$\|x - y\| \leq \|b_1 - a_1\| = \sqrt{\sum_{i=1}^{N} (b_{1i} - a_{1i})^2} = \frac{\delta}{2}$$

2. **Step 2**:
   Induction. Suppose $n \geq 1$ is fixed and $I \supseteq I_1 \supseteq I_2 \supseteq \cdots \supseteq I_n \supseteq \cdots$ are N-cells where each I_l cannot be covered by a finite collection of sets from g, and if $x, y \in I_l$, we have $\|x - y\| \leq \delta/2^l$. Repeat the argument in step 1 to get an N-cell $I_{n+1} \subseteq I_n$ that cannot be covered by a finite collection of sets from g and $x, y \in I_{l+1}$, then $\|x - y\| \leq \delta/2^{n+1}$. We have proved the existence of a sequence $I, I_1, I_2, \ldots$ with the following properties:

   (a) $I \supseteq I_1 \supseteq I_2 \supseteq \cdots \supseteq I_n \supseteq \cdots$

   (b) Each I_n cannot be covered by a finite collection of sets from g

   (c) If $x, y \in I_n$, then $\|x - y\| \leq \delta/2^n$

   By Theorem 5.3 we can find $z \in \bigcap_{n=1}^{\infty} I_n$. Since $z \in I \subseteq \bigcup_{\alpha \in \Lambda} g_\alpha$, there exists some $\beta \in \Lambda$ such that $z \in g_\beta$. Because g_β is open, there exists $\varepsilon > 0$ such that $\mathcal{B}_\varepsilon(z) \subseteq g_\beta$. Let n be such that $\delta/2^n < \varepsilon$. We know that $z \in I_n$ and if $y \in I_n$, we have

   $$\|y - z\| \leq \frac{\delta}{2^n} < \varepsilon$$

   giving that $y \in \mathcal{B}_\varepsilon(z)$. This shows that

   $$I_n \subseteq \mathcal{B}_\varepsilon(z) \subseteq g_\beta$$

   which is a contradiction because I_n can be covered by the singleton $\{g_\beta\} \in g$.

□

6 Connected Sets

Intuitively, a set $S \subseteq \mathbb{R}^N$ is **connected** if any two points $x, y \in S$ can be connected by a continuous path that is completely contained in S.

We define connected sets using topology.

Definition 6.1: Disconnection and Connection

Let $S \subseteq \mathbb{R}^N$ be a set. We say that a pair of open set $\{U, V\} \in \mathbb{R}^N$ is a **disconnection** for S if

1. $S \subseteq U \cup V$
2. $S \cap U \neq \varnothing$ and $S \cap V \neq \varnothing$
3. $S \cap U \cap V = \varnothing$

If a disconnection exists, we say that S is **disconnected**. Otherwise we say S is **connected**.

Example 6.1

$\mathbb{Z}$ is not connected, set $U = (-\infty, 1/2)$ and $V = (1/2, +\infty)$.
$\mathbb{Q}$ is not connected, set $U = (-\infty, \sqrt{2})$ and $V = (\sqrt{2}, +\infty)$

6.1 Interval is Connected

Theorem 6.1

The interval $[0, 1]$ is connected.

Lecture 9 - Friday, May 24

Proof. SFAC that $\{U, V\}$ is a disconnection. WLOG we may assume $0 \in U$. Since U is open, there exists some $\varepsilon_0 > 0$ such that $(-\varepsilon_0, \varepsilon_0) \subseteq U$. We may assume $\varepsilon_0 < 1$. Then $[0, \varepsilon_0) \subseteq U$. It follows that

$$\{0 < \varepsilon < 1 : [0, \varepsilon) \subseteq U\}$$

is not empty. We let $t_0 = \sup\{0 < \varepsilon < 1 : [0, \varepsilon) \subseteq U\}$. Notice that $t_0 \leq 1$.

1. *Claim 1*: $[0, t_0) \subseteq U$.
   Indeed, for each $n \geq 1$, let $r_n > 0$ with $t_0 - 1/n < r_n < t_0$ such that $[0, r_n) \subseteq U$. We then have

$$[0, t_0) = \bigcup_{n=1}^{\infty} [0, r_n) \subseteq U$$

2. *Claim 2*: $t_0 \notin U$.
   SFAC $t_0 \in U$, thus we obtain that $t_0 \neq 1$ because if $t_0 = 1 \in U$, then

$$\begin{aligned} U &\supseteq [0, t_0) \cup \{t_0\} \\ &= [0, 1) \cup \{1\} \\ &= [0, 1] \end{aligned}$$

   which contradicts property (c) as we simultaneously have

$$U \cap [0, 1] \cap V = \varnothing \qquad [0, 1] \cap V \neq \varnothing$$

   Therefore, there exists $\delta > 0$ such that $(t_0 - \delta, t_0 + \delta) \subseteq U$. We may assume $t_0 + \delta < 1$. Then we know that

$$[0, t_0 + \delta) \subseteq [0, t_0) \cup [t_0, t_0 + \delta) \subseteq U$$

   contradicting the definition of t_0.

Therefore we deduce that $t_0 \in V$. Since V is open, we can find $\delta_V > 0$ such that $(t_0 - \delta_V, t_0 + \delta_V) \subseteq V$. But then take some $0 < r < t_0$, $r > t_0 - \delta_V$, then $r \in [0, 1]$, and $r \in U$ by claim 1, while $r \in V$. Contradiction (see theorem 6.1). ☐

6.2 Higher-Dimensional Examples

Definition 6.2: Convex

We say that $C \subseteq \mathbb{R}^N$ is **convex** if for all $x, y \in C$, we have

$$tx + (1 - t)y \in C \qquad \forall\, t \in [0, 1]$$

In other words, C contains the line segment between any two points in C.

6.3 Convex is Connected

Theorem 6.2

Any convex set $C \subseteq \mathbb{R}^N$ is connected.

Proof. SFAC $C \in \mathbb{R}^N$ is not connected. Let $\{U, V\}$ be a disconnection. Let $x \in C \cap U$ and let $y \in C \cap V$. Define

$$\begin{aligned} U_0 &:= \{t \in \mathbb{R} : tx + (1 - t)y \in U\} \\ V_0 &:= \{t \in \mathbb{R} : tx + (1 - t)y \in V\} \end{aligned}$$

we will show that $\{U_0, V_0\}$ gives a disconnection for $[0, 1]$.
Claim: U_0 and V_0 are open.

Let $t_0 \in U_0$, so $x_0 = t_0 x + (1 - t_0)y \in U$. Since U is open, there exists $\varepsilon > 0$ such that

$$\mathcal{B}_\varepsilon(x_0) \subseteq U$$

For each $t \in \mathbb{R}$, we set $z_t := tx + (1 - t)y$. Notice that

$$\begin{aligned} \|z_t - x_0\| &= \|tx + (1 - t)y - (t_0 x + (1 - t_0)y)\| \\ &\leq \|(t - t_0)x\| + \|(t_0 - t)y\| \\ &= |t - t_0| \, \|x\| + |t - t_0| \, \|y\| \end{aligned}$$

Let $\delta > 0$, $\delta = \frac{\varepsilon}{\|x\| + \|y\|}$, then if $t \in (t_0 - \delta, t_0 + \delta)$, we get $\|z_t - x_0\| < \varepsilon$, which suggests that

$$z_t \in \mathcal{B}_\varepsilon(x_0) \subseteq U$$

This shows that $(t_0 - \delta, t_0 + \delta) \subseteq U_0$, and hence U_0 is open. Similar argument could also show that V_0 is open. Then $\{U_0, V_0\}$ is a disconnection for $[0, 1]$ because

1. $[0, 1] \subseteq U_0 \cup V_0$.
   If $t \in [0, 1]$, $z_t = tx + (1 - t)y \in C$ because we know that C is convex, thus $z_t \in U$ or $z_t \in V$. So that $z_t \in U \cup V$.
2. $[0, 1] \cap U_0 \neq \varnothing$ because $1 \in U_0$, and $[0, 1] \cap V_0 \neq \varnothing$ because $0 \in V_0$.
3. $[0, 1] \cap U_0 \cap V_0 = \varnothing$.
   Indeed, if $t \in [0, 1] \cap U_0 \cap V_0$, then $z_t \in U \cap V \cap C$ (in C because C is convex). This cannot happen because $\{U, V\}$ is a disconnection for C. Hence $[0, 1] \cap U_0 \cap V_0 = \varnothing$.

Thus $\{U, V\}$ is a disconnection for $[0, 1]$. Contradiction. □

Corollary 6.1

The following subsets of $\mathbb{R}^N$ are connected:

1. $\mathbb{R}^N$
2. open balls
3. line segments
4. subspaces

6.4 Only $\mathbb{R}^N$ and $\varnothing$ are Clopen

Corollary 6.2

The only clopen sets in $\mathbb{R}^N$ are $\mathbb{R}^N$ and $\varnothing$.

Proof. Exercise.
My Attempt:

Suppose there exists $U \subseteq \mathbb{R}^N$ with $U \neq \varnothing$ and $U \neq \mathbb{R}^N$ such that U is clopen. Thus we can find that $V := \mathbb{R}^N \setminus U$ is also clopen. Notice that thus we have

1. $\mathbb{R}^N \subseteq U \cup V$
2. $U \cap \mathbb{R}^N \neq \varnothing$ and $V \cap \mathbb{R}^N \neq \varnothing$
3. $\mathbb{R}^N \cap U \cap V = \varnothing$

which implies that $\mathbb{R}^N$ is disconnected. Contradiction. □

Lecture 10 - Monday, May 27

7 Sequence and Limits in $\mathbb{R}^N$

Definition 7.1: Sequence

A **sequence** in $\mathbb{R}^N$ is a function $f : \mathbb{N} \to \mathbb{R}^N$.
Notation: we write $x_n = f(n)$, and we write (x_n), $(x_n)_{n=1}^{\infty}$, or $(x_n)_{n\in\mathbb{N}}$ for the sequence.

Definition 7.2: Limit

We say that a sequence (x_n) in $\mathbb{R}^N$ **converges** to $a \in \mathbb{R}^N$ if for every $\varepsilon > 0$, there exists $M \in \mathbb{R}$ such that for all $n \geq M$

$$\|x_n - a\| < \varepsilon$$

or equivalently,

$$x_n \in \mathcal{B}_\varepsilon(a)$$

We call a the **limit** of (x_n) and say that (x_n) is convergent.
Notation: we write $a = \lim_{n\to\infty} x_n$, or $x_n \to a$.

Discovery 7.1

Notice that (x_n) converges to a if and only if for every open $U \subseteq \mathbb{R}^N$ with $a \in U$, there exists $M_U \in \mathbb{N}$ such that $x_n \in U$ for all $n \geq M_U$.

Definition 7.3: Bounded

Let (x_n) be a sequence in $\mathbb{R}^N$, we say that (x_n) is **bounded** if its set of terms $\{x_n : n \in \mathbb{N}\}$ is a bounded set.

7.1 Bounded if (Cauchy iff Convergent)

Definition 7.4: Cauchy

We say (x_n) is **Cauchy** if for every $\varepsilon > 0$ there exists $M \in \mathbb{N}$ such that

$$\|x_n - x_m\| < \varepsilon \qquad \text{for all } n, m \geq M$$

Discovery 7.2

If (x_n) is a sequence in $\mathbb{R}^N$, then

$$(x_n) \text{ is convergent } \leftrightarrow (x_n) \text{ is cauchy } \rightarrow (x_n) \text{ is bounded}$$

Proposition 7.1

Let (x_n) be a sequence in $\mathbb{R}^N$, then

1. if (x_n) is convergent, then it is cauchy;
2. if (x_n) is cauchy, then it is bounded.

Proof.

1. Suppose (x_n) is convergent and let $a := \lim_{n\to\infty} x_n$. Let $\varepsilon > 0$ and let $M \in \mathbb{N}$ such that $\|x_n - a\| < \varepsilon/2$ for all $n > M$. For $m, n > M$, we have

$$\begin{aligned}\|x_n - x_m\| &\leq \|x_n - a\| + \|a - x_m\| \\ &= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\end{aligned}$$

thus the sequence (x_n) is cauchy.

2. Suppose (x_n) is cauchy. For $\varepsilon = 1$, let $M \in \mathbb{N}$ be such that $\|x_n - x_m\| = 1$ for all $m, n > M$, then

$$\|x_n\| = \|x_n - x_M + x_M\| \leq \|x_n - x_M\| + \|x_M\|$$

Take $R := \max\{\|x_1\|, \|x_2\|, \ldots, \|x_{M-1}\|, 1 + \|x_M\|\}$, then $\|x_n\| \leq R$ for all $n \in \mathbb{N}$, suggesting that (x_n) is bounded.

□

Proposition 7.2

A sequence (x_n) in $\mathbb{R}^N$ can have at most one limit.

Proof. Suppose (x_n) is convergent. SFAC that $a, b \in \mathbb{R}^N$, $a \neq b$ with $a = \lim_{n\to\infty} x_n = b$. Since $a \neq b$, we have $\|a - b\| \neq 0$, and we set $\varepsilon = \|a - b\|/2$. Then

$$\mathcal{B}_\varepsilon(a) \cap \mathcal{B}_\varepsilon(b) = \varnothing$$

Let $M_a \in \mathbb{N}$ be such that $x_n \in \mathcal{B}_\varepsilon(a)$ for all $n \geq M_a$, and let $M_b \in \mathbb{N}$ be such that $x_n \in \mathcal{B}_\varepsilon(b)$ for all $n \geq M_b$, then for $n \geq M := \max\{M_a, M_b\}$, we have

$$x \in \mathcal{B}_\varepsilon(a) \cap \mathcal{B}_\varepsilon(b) = \varnothing$$

which is a contradiction. □

8 Sequential Characterization of Compact Set

Proposition 8.1

Let $S \subseteq \mathbb{R}^N$ and $p \in \mathbb{R}^N$, then TFAE:

1. $p \in S'$;
2. There exists $(x_n) \in S$ with $x_n \neq x_m$ if $n \neq m$ such that $\lim_{n\to\infty} x_n = p$.

Proof. A2. □

Definition 8.1: Subsequence

A **subsequence** of a sequence (x_n) in $\mathbb{R}^N$ is a sequence of the form $(x_{n_k})_{k=1}^\infty$ with

$$n_1 < n_2 < n_3 < \cdots < n_k < \cdots$$

Example 8.1

Consider the sequence in $\mathbb{R}^3$ such that

$$x_n = \left((-1)^n, \cos\left(\frac{\pi n}{2}\right), \frac{1}{n}\right)$$

notice that it is not convergent, but it is bounded and has convergent subsequences. In particular, as for an instance, the following subsequences are convergent:

$$n_k = 2k + 1$$
$$n_k = 4k$$

Proposition 8.2

If (x_n) converges to $a \in \mathbb{R}^N$, then every subsequence also converges to a.

Proof. Let $a = \lim_{n\to\infty} x_n$ and let (x_{n_k}) be a subsequence. Let $\varepsilon > 0$ and let $M \in \mathbb{N}$ be such that

$$\|x_n - a\| < \varepsilon \qquad \text{for all } n > M$$

Let $k_0 \in \mathbb{N}$ be such that $n_{k_0} \geq M$. Then

$$k \geq k_0 \quad \Rightarrow \quad n_k \geq n_{k_0} \geq M$$

and so $\|x_{n_k} - a\| < \varepsilon$, which implies that (x_{n_k}) converges to a. □

8.1 Compact and Sequential Compact (in Metric Space $\mathbb{R}^N$)

Theorem 8.1

Let $K \subseteq \mathbb{R}^N$, TFAE:

1. K is compact;
2. Every sequence (x_n) in K has a subsequence that converges to a point in K.

Proof.

1. (1) $\Longrightarrow$ (2)
   Let (x_n) be a sequence in K, we need to consider two cases:

   (a) *Case 1*: $E := \{x_n : n \in \mathbb{N}\}$ is finite.
   Then there exists $a \in E$ such that the set $\{n \in \mathbb{N} : x_n = a\}$ is infinite. We build a subsequence (x_{n_k}) of (x_n) converging to $a \in K$ as following: We set

   $$A_1 = \{n \in \mathbb{N} : x_n = a\}$$

   then $A_1 \neq \varnothing$, and we set $n_1 = \min A_1$. Let

   $$A_2 = \{n \in \mathbb{N} : n > n_1, x_n = a\}$$

   then $A_2 \neq \varnothing$, and we set $n_2 = \min A_2$. Proceeding with the argument inductively we obtain

   $$n_1 < n_2 < \cdots < n_k < \cdots$$

   such that $x_{n_k} = a$ for all k. Thus (x_{n_k}) definitely converges to a.

   (b) *Case 2*: $E := \{x_n : n \in \mathbb{N}\}$ is infintie.
   In this case, since K is compact, then by Theorem (5.2), E has a cluster point $a \in K$. Then we build a subsequence (x_{n_k}) converging to a as following: For $\varepsilon_1 = 1$, take $x_{n_1} \in \mathcal{B}_{\varepsilon_1}(a)$; For $\varepsilon_2 = 1/2$, take $n_2 > n_1$ and $x_{n_2} \in \mathcal{B}_{\varepsilon_2}(a)$. Continue with the argument inductively, then for $\varepsilon_k = 1/n$, $n_k > n_{k-1}$ with $x_{n_k} \in \mathcal{B}_{\varepsilon_k}(a)$.

Lecture 11 - Wednesday, May 29

2. (2) $\Longrightarrow$ (1)
   SFAC K is not compact, then K is either not bounded or not closed.

   (a) if K is not bounded
   So for each $n \in \mathbb{N}$, we can find $x_n \in K$ with $\|x_n\| > n$. The sequence (x_n) has no bounded subsequence, which is hence not convergent. Hence we conclude that K must be bounded.

(b) if K is not closed

By the characterization of closed set 4.1, there exists $p \in K'$ with $p \notin K$. By A02-Q4, there exists (x_n), a sequence in K, converges to p. Then every subsequence also converge to $p \notin K$ by Proposition 8.2, contradicting 2), so k must be clsoed.

□

Theorem 8.2: Bolzano-Weierstrass Theorem in $\mathbb{R}^N$

Let (x_n) be a bounded sequence in $\mathbb{R}^N$, then (x_n) has a convergent subsequence.

Proof. Suppose (x_n) is bounded, say $(x_n) \subseteq \mathcal{B}_R[0]$. Since $\mathcal{B}_R[0]$ is closed and bounded, it is compact. Hence x_n has a convergent subsequence by Theorem 8.1. □

Proof. This is an alternative proof
Using BW 8.2 in $\mathbb{R}$, since

$$x_n = (x_{n,1}, x_{n,2}, \dots, x_{n,N})$$

For the first sequence, find a convergent subsequence $(x_{n_k,1})$, and take (x_{n_k}). Using this subsequence, at the second coordinate find a convergent subsequence of (x_{n_k}), denoted as $(x_{n_{k_j},2})$, to get $(x_{n_{k_j}})$. Continuing this argument for each coordinate.

Discovery 8.1

This proof is called the "Diagonal Argument".

□

Theorem 8.3: Completeness of $\mathbb{R}^N$

Every cauchy sequence in $\mathbb{R}^N$ is convergent.

Proof. We know that by Proposition 7.1 every cauchy sequence is bounded. Let (x_n) be a cauchy sequence in $\mathbb{R}^N$. It follows by BW Theorem (8.2) in $\mathbb{R}^N$ that (x_n) has a convergent subsequence (x_{n_k}). Let $a = \lim_{k\to\infty} x_{n_k}$. We will show that (x_n) converges to a. Let $\varepsilon > 0$ and let $k_0 \in \mathbb{N}$ be such that $\|x_{n_k} - a\| < \varepsilon/2$ for all $k \geq k_0$. Let $M \in \mathbb{N}$ be such that $\|x_n - x_m\| < \varepsilon/2$ for all $n, m \geq M$. Let $n \geq M$, let k be such that $k \geq k_0$ and $n_k \geq M$ (e.g. $k \geq \max\{k_0, M\}$). Then

$$\begin{aligned}
\|x_n - a\| &= \|x_n - x_{n_k} + x_{n_k} - a\| \\
&\leq \|x_n - x_{n_k}\| + \|x_{n_k} - a\| \\
&= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}$$

□

9 Limits of Function and Continuity

9.1 Limit

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ and $f : D \to \mathbb{R}^M$ a function, given $x_0 \in D'$, we wish to study the behaviour of f around x_0.

Definition 9.1: Limit

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ and $f : D \to \mathbb{R}^M$ a function, given $x_0 \in D'$. We say that $L \in \mathbb{R}^M$ is the **limit** of f as $x \to x_0$, written $L = \lim_{x \to x_0} f(x)$, if for every $\varepsilon > 0$, there exists $\delta > 0$ such that if $x \in D$ and $0 < \|x - x_0\| < \delta$, then $\|f(x) - L\| < \varepsilon$.
If there is no $L \in \mathbb{R}^M$ such that the above happens, then we say that the limit of f at x_0 does not exist.

Theorem 9.1

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ and $f : D \to \mathbb{R}^M$ a function, given $x_0 \in D'$. TFAE:

1. $L = \lim_{x \to x_0} f(x)$
2. For every sequence (x_n) in $D \backslash \{x_0\}$ with $x_n \to x_0$, the sequence $(f(x_n))$ converges to L
3. For every neighbourhood U of L, there exists an open neighbourhood V of x_0 such that

$$(V \cap D) \backslash \{x_0\} \subseteq f^{-1}(U) := \{x \in D : f(x) \in U\}$$

Definition 9.2: Neighbourhood

U is a **neighbourhood** of x_0 if there exists $\varepsilon > 0$ such that $\mathcal{B}_\varepsilon(x_0) \subset U$.

Proof. 1. $(1 \implies 2)$
Let (x_n) be a sequence in $D \backslash \{x_0\}$ converging to x_0. Let $\varepsilon > 0$ be given. Then there exists $\delta > 0$ such that if $x \in D$, $0 < \|x - x_0\| < \delta$, then $\|f(x) - L\| < \varepsilon$. Let $M \in \mathbb{N}$ be such that $x_0 \in \mathcal{B}_\delta(x_0)$ for all $n \geq M$. Then $\|f(x_n) - L\| < \varepsilon$, giving that $(f(x_n))$ converges to L.

2. $(2 \implies 1)$
SFAC $L \neq \lim_{x \to x_0} f(x)$, then there exists $\varepsilon > 0$ such that for every $\delta > 0$, we can find $x_\delta \in D$ with $0 < \|x_\delta - x_0\| < \delta$ such that

$$\|f(x_\delta) - L\| > \varepsilon$$

For $\delta = 1$, find $x_1 \in \mathcal{B}_1(x_0) \backslash \{x_0\}$, $x_1 \in D$ with $\|f(x_1) - L\| > \varepsilon$. For $\delta = 1/n$, find $x_n \in D$, $x_n \in \mathcal{B}_{1/n}(x_0) \backslash \{x_0\}$ with $\|f(x_n) - L\| > \varepsilon$. The corresponding sequence $(x_n) \subseteq D \backslash \{x_0\}$ converges to x_0, but $(f(x_n))$ does not converge to L. Contradiction.

Lecture 12 - Friday, May 31

3. $(1 \Longrightarrow 3)$
   Suppose (1) holds and let U be an open neighbourhood of L. Let $\varepsilon > 0$ such that $\mathcal{B}_\varepsilon(L) \subseteq U$. By (1), there exists $\delta > 0$ such that if $x \in D$ and $0 < \|x - x_0\| < \delta$, then $\|f(x) - L\| < \varepsilon$, which further implies that $f(x) \in \mathcal{B}_\varepsilon(L)$. Set $V := \mathcal{B}_\delta(x_0)$, then

$$(V \cap D)\backslash\{x_0\} \ \subset \ f^{-1}\left(\mathcal{B}_\varepsilon(L)\right) \ \subset \ f^{-1}(U)$$

4. $(3 \Longrightarrow 1)$
   Let $\varepsilon > 0$. Set $U := \mathcal{B}_\varepsilon(L)$. By (3) we can find an open neighbourhood V of x_0 such that

$$(V \cap D)\backslash\{x_0\} \ \subset \ f^{-1}(U)$$

   Let $\delta > 0$ be such that $\mathcal{B}_\delta(x_0) \subset V$, then if $x \in \mathcal{B}_\delta(x_0) \cap D$, $x \neq x_0$, then

$$x \in (V \cap D)\backslash\{x_0\} \quad \Rightarrow \quad x \in f^{-1}(U)$$

□

Notice: If $D \subset \mathbb{R}$, x approaches x_0 either from the left or from the right. In $\mathbb{R}^N$, $N \geq 2$, there are many different ways x can approach x_0.

Example 9.1

Consider $D = \mathbb{R}^2\backslash\{(0,0)\}$, $f : D \to \mathbb{R}$, $f(x,y) = xy/(x^2+y^2)$ and $x_0 = (0,0)$. Let (x_n) in $D\backslash\{x_0\}$, $x_n = (1/n, 1/n)$, then $x_n \to (0,0)$ and $f(x_n) \to 1/2$. Take $x_m = (1/m, 1/m^2)$, compute to find that $f(x_m) \to 0$. We conclude that by the Sequential Characterization ((2) of 9.1) that the limit of f at x_0 does not exist.

Example 9.2

Let $D = \mathbb{R}^2\backslash\{(0,0)\}$. Let $f : D \to \mathbb{R}$, $f(x,y) = x^4/(x^2+y^2)$ and $x_0 = (0,0)$. We claim that

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0$$

Assume $x \neq 0$, then $f(x,y) = \frac{x^2}{1+y^2/x^2}$. We have $1 + y^2/x^2 \geq 1$, hence $\frac{1}{1+y^2/x^2} \leq 1$, giving that $f(x,y) = \frac{x^2}{1+y^2/x^2} \leq x^2$. Thus given $\varepsilon > 0$, take $\delta = \sqrt{\varepsilon}$, thus if $\|(x,y)\| < \delta$, we have $x^2 < \varepsilon$.

9.2 Continuity

Definition 9.3: Continuous

Let $D \subseteq \mathbb{R}^N$, $f : D \to \mathbb{R}^M$ be a function. We say that f is **continuous** at $x_0 \in D$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $x \in D$ and $\|x - x_0\| < \delta$ we have $\|f(x) - f(x_0)\| < \varepsilon$. We say that f is **continuous on** D if f is continuous at every point $x_0 \in D$.

Discovery 9.1

1. Continuity only makes sense at a point $x_0 \in D$.
2. We say that a point $x_0 \in D$ is **isolated** if there exists $\delta > 0$ such that $\mathcal{B}_\delta(x_0) \cap D = \{x_0\}$ (e.g. $x_0 \in D \backslash D'$). If $x_0 \in D$ is an isolated point, then every function $f : D \to \mathbb{R}^M$ is continuous at x_0.

Theorem 9.2

Let $f : D \to \mathbb{R}^M$ be a function $x_0 \in D \cap D'$, then f is continuous at x_0 if and only if $\lim_{x \to x_0} f(x) = f(x_0)$.

9.3 Properties of Continuous Functions

Proposition 9.1

Let $D \subseteq \mathbb{R}^N$ and let $f, g : D \to \mathbb{R}^M$, $\phi : D \to \mathbb{R}$. Suppose f, g and ϕ are continuous at $x_0 \in D$, then

$$\begin{aligned} f + g : D &\to \mathbb{R}^M \\ x &\mapsto f(x) + g(x) \end{aligned} \qquad \begin{aligned} f \cdot g : D &\to \mathbb{R}^M \\ x &\mapsto f(x) \cdot g(x) \end{aligned} \qquad \begin{aligned} \phi f : D &\to \mathbb{R}^M \\ x &\mapsto \phi(x) \cdot f(x) \end{aligned}$$

where the second is dot product and the third is scalar multiplication, are continuous.

Proof. Exercise. (Use, for example, $f(x_n) \to f(x_0)$ if and only if $f(x_n)_j \to f(x_0)_j$ for $j = 1, \ldots, M$). □

Proposition 9.2

Let $f_1 : D_1 \to \mathbb{R}^K$, $D_1 \subseteq \mathbb{R}^N$ and $f_2 : D_2 \to \mathbb{R}^K$, $D_2 \subseteq \mathbb{R}^M$. Suppose $f_1(D_1) \subseteq D_2$. If f_1 is continuous at $x_0 \in D_1$ and f_2 is continuous at $f_1(x_0)$, then $f_2 \circ f_1 : D_1 \to \mathbb{R}^M$, $x \mapsto f_2(f_1(x))$ is continuous at x_0.

Proof. Let (x_n) be a sequence in D_1 converging to x_0. We need to show that

$$\lim_{n \to \infty} (f_2 \circ f_1)(x_n) = f_2(f_1(x_0))$$

Since f_1 is continuous at x_0, we have $(f_1(x_n))$ converges to $f_1(x_0)$. Because f_2 is continuous at $f_1(x_0)$ and $(f_1(x_n)) \to f_1(x_0)$, we get

$$\lim_{n \to \infty} f_2(f_1(x_n)) = f_2(f_1(x_0))$$

□

Proposition 9.3

Let $f : D \to \mathbb{R}^M$, $D \subseteq \mathbb{R}^N$, be a function. For each $j = 1, \ldots, M$, let $f_j : D \to \mathbb{R}$ be j^{th} component of f, so that

$$f(x) = (f_1(x), f_2(x), \ldots, f_M(x))$$

for all $x \in D$. Then f is continuous at x_0 if and only if f_j is continuous at x_0 for each j.

Proof. Exercise. $\square$

Example 9.3

For $j \in \{1, \ldots, N\}$, then the function $\pi_j : \mathbb{R}^N \to \mathbb{R}$, $(x_1, \ldots, x_N) \to x_j$ (projectino onto the j^{th} coordinate) is continuous. Then every function $f : \mathbb{R}^N \to \mathbb{R}$, $f(x_1, \ldots, x_N)x_1^{n_1} \cdots x_M^{n_N}$, $n_j \geq 0$, $j = 1, \ldots, N$ is continuous.

Example 9.4

The function $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = \frac{xy^2}{x^2+y^4+\pi}$ is continuous on $\mathbb{R}^2$. Indeed, $f = f_1 \cdot f_2$, for $f_1 = xy^2$ is continuous and $f_2 = \frac{1}{x^2+y^4+\pi}$ is continuous. f_2 is continuous because $f_2(x, y) = g_2 \circ g_1$ for $g_1(x, y) = x^2 + y^4 + \pi \subseteq \mathbb{R}\backslash\{0\}$ and $g_2 : \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $t \mapsto 1/t$ are continuous.

Example 9.5

The function $f : \mathbb{R}^2 \to \mathbb{R}^3$

$$f(x, y) = \left(\cos\left(\frac{xy^2}{x^2 + y^4 + \pi}\right), \sin\left(\frac{xy^2}{x^2 + y^4 + \pi}\right), e^{x+y}\right)$$

is continuous on $\mathbb{R}^2$ since each composition f_1, f_2, f_3 of f is continuous on $\mathbb{R}$.

Global Properties of Continuity

Theorem 9.3

Let $\varnothing \neq D \subseteq \mathbb{R}^N$, $f : D \to \mathbb{R}^M$ be a function, TFAE:

1. f is continuous on D;
2. For every $U \subseteq \mathbb{R}^M$ open, there exists $V \subseteq \mathbb{R}^N$ open such that $f^{-1}(U) = V \cap D$;
3. For every $F \subseteq \mathbb{R}^M$ closed, there exists $G \subseteq \mathbb{R}^N$ closed such that $f^{-1}(F) = G \cap D$.

Proof. 1. (1) $\implies$ (2)

Suppose f is continuous on D and let $U \subseteq \mathbb{R}^M$. We claim that for each $x \in f^{-1}(U)$, there exists an open neighbourhood V_x of x such that

$$V_x \cap D \subseteq f^{-1}(U)$$

Indeed, in case that $x \in D$ is an isolated point, let $\delta_x > 0$ be such that $\mathcal{B}_{\delta_x}(x) \cap D = \{x\}$, set $V_x = \mathcal{B}_{\delta_x}(x)$. If $x \in D \cap D'$, then $\lim_{y\to x} f(y) = f(x)$. By Theorem (9.1, (1) $\to$ (3)), there exists an open neighbourhood V_x of x such that

$$(V_x \cap D)\backslash\{x\} \subseteq f^{-1}(U)$$

and hence $V_x \cap D \subseteq f^{-1}(U)$. Set $V = \bigcup_{x\in f^{-1}(U)} V_x$, then V is open in $\mathbb{R}^M$ and

$$f^{-1}(U) \subseteq \bigcup_{x\in f^{-1}(U)} V_x \cap D \subseteq f^{-1}(U)$$

giving that $f^{-1}(U) = V \cap D$.

2. (2) $\implies$ (1)

Let $x_0 \in D \cap D'$, we apply Theorem (9.1, (3) $\to$ (1)). Let U be an open neighborhood of $f(x)$. We know that there exists $V \subseteq \mathbb{R}^N$ open such that $V \cap D = f^{-1}(U)$. Then V is open neighborhood of x since $x \in f^{-1}(U)$ and $(V \cap D)\backslash\{x\} \subseteq f^{-1}(U)$. By Theorem (9.1, (3) $\to$ (1)), $\lim_{y\to x} f(y) = f(x)$, and so f is continuous at x.

3. (2) $\implies$ (3)

Suppose $F \subseteq \mathbb{R}^M$ is closed. Then F^c is open. By assumption, there exists $V \subseteq \mathbb{R}^N$ open such that

$$f^{-1}(F^c) = V \cap D$$

Now we use that $f^{-1}(F^c) = f^{-1}(F)^c \cap D$. Hence $f^{-1}(F)^c \cap D = V \cap D$. Taking complement and then the intersection with D yields $f^{-1}(F) = V^c \cap D$. Setting $G := V^c$ gives the result.

4. (3) $\implies$ (2)

Follows a similar proof as above.

□

9.3.1 Example and Application

Example 9.6

Prove that the set $F \subseteq \mathbb{R}^4$,

$$F = \left\{(x, y, z, w) : e^{x+y} \sin(zw^2) \in [0, 2], x^2 + w^2 + z^3 - y^4 \in [0, 2024]\right\}$$

is clsoed

Proof. Let $f : \mathbb{R}^4 \to \mathbb{R}^2$:

$$f(x, y, z, w) = \left(e^{x+y} \sin(zw^2), x^2 + w^2 + z^3 - y^4\right)$$

then f is continuous on $\mathbb{R}^4$, we have

$$F = f^{-1}(F') \qquad \text{where} \quad F' = [0, 2] \times [0, 2024]$$

It follows from the above Theorem (9.3, $1 \to 3$) that F is closed. □

9.4 Continuity and Compactness

Theorem 9.4

Let $\varnothing \neq K \subseteq \mathbb{R}^N$ be compact and $f : K \to \mathbb{R}^M$ be continuous on K, then $f(K)$ is compact.

Proof. Let $U = \{U_\alpha\}_{\alpha\in\Lambda}$ be an open cover of $f(K)$. By Theorem (9.3) for each $\alpha \in \Lambda$, there exists $V_\alpha \subseteq \mathbb{R}^N$ open such that $V_\alpha \cap K = f^{-1}(U_\alpha)$. Set $V = \{V_\alpha\}_{\alpha\in\Lambda}$, then

$$\begin{aligned} K = f^{-1}(f(K)) = f^{-1}\left(\bigcup_{\alpha\in\Lambda} U_\alpha\right) &= \bigcup_{\alpha\in\Lambda} f^{-1}(U_\alpha) \\ &= \bigcup_{\alpha\in\Lambda} V_\alpha \cap K \subseteq \bigcup_{\alpha\in\Lambda} V_\alpha \end{aligned}$$

Hence V is an open cover for K. By compactness, V admists a finite subcover $V' = \{V_{\alpha_i} : i = 1, \ldots, l\}$. Then

$$\begin{aligned} f(K) &= f\left(\bigcup_{i=1}^{l} V_{\alpha_i} \cap K\right) \\ &= \bigcup_{i=1}^{l} f(V_{\alpha_i} \cap K) \\ &= \bigcup_{i=1}^{l} U_{\alpha_i} \cap f(K) \\ &\subseteq \bigcup_{i=1}^{l} U_{\alpha_i} \end{aligned}$$

Hence $U = \{U_{\alpha_i} : i = 1, \ldots, l\}$ is a finite subcover for $f(K)$. □

Corollary 9.1

If $\varnothing \neq K \subseteq \mathbb{R}^N$ is compact, $f : K \to \mathbb{R}^M$ be continuous, then $f(K)$ is is closed and bounded.

Proof. Theorem 5.6. □

Lecture 14 - Wednesday, Jun 5

9.4.1 Extreme Value Theorem

Theorem 9.5: Extreme Value Theorems

Suppose $\varnothing \neq K \subseteq \mathbb{R}^N$ is compact and $f : K \to R$ is continuous, then there are $x_{min}, x_{max} \in K$ such that

$$f(x_{min}) = \inf_{x \in K} f(x) \quad \text{and} \quad f(x_{max}) = \sup_{x \in K} f(x)$$

Proof. By Theorem (9.4) and Theorem (5.6), we know that $f(K)$ is closed and bounded. In particular, $\inf f(K) = \inf_{x \in K} f(x)$ and $\sup_{x \in K} f(x)$ exist. Since $f(K)$ is closed, we must have $\inf_{x \in K} \in f(K)$ and $\sup_{x \in K} \in f(K)$. □

9.5 Uniform Continuity

Definition 9.4: Uniformly continuous

Let $D \subseteq \mathbb{R}^N$ and $f : D \to \mathbb{R}^M$ be a function, we say that f is uniformly continuous if given $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in D$ satisfying $\|x - y\| < \delta$, we have $\|f(x) - f(y)\| < \varepsilon$.

Example 9.7

Let $D = [-d, d] \subseteq \mathbb{R}$ be closed and bounded. Let $f : D \to \mathbb{R}$ be defined as $f(x) = x^2$. Then f is uniformly continuous on D. (In fact, D only needs to be bounded.)

Proof. $\varepsilon > 0$, we have for $x, y \in D$,

$$|f(x) - f(y)| = |x + y||x - y|$$

hence we can easily take $\delta = \varepsilon/2d$. □

Example 9.8

Let $f : (0, 1) \to \mathbb{R}$ defined as $f(x) = 1/x$, then f is not uniformly continuous on $D = (0, 1)$.

Proof. Take $\varepsilon = 1$, given $\delta > 0$, let $n \in \mathbb{N}$ be such that $\frac{1}{n} < \frac{\delta}{2}$. Set $x = \frac{1}{n}$ and $y = \frac{1}{n+1}$. Now we have $|x - y| < \delta$, but $|f(x) - f(y)| = 1 \geq \varepsilon$. □

Example 9.9

The function $x \mapsto \sin 1/x$ $(x > 0)$ is not uniformly continuous on $(0, \infty)$ because $\lim_{x \to 0} \sin 1/x$ does not exist.

Theorem 9.6

Let $\varnothing \neq K \subseteq \mathbb{R}^N$ be compact and $f : K \to \mathbb{R}^M$ be continuous, then f is uniformly continuous on K.

Proof. SFAC that f is not. Then there exists $\varepsilon > 0$ such that for each $\delta_n = 1/n$, we can find $x_n, y_n \in K$ such that

$$\|x_n - y_n\| < \delta \qquad \|f(x_n) - f(y_n)\| \geq \varepsilon$$

Since K is compact, by Theroem (8.1), (x_n) has a subsequence (x_{n_k}) converging to a point $x \in K$. Notice that

$$\begin{aligned}\lim_{k\to\infty} y_{n_k} &= \lim_{k\to\infty} (y_{n_k} - x_{n_k} + x_{n_k}) \\ &= \underbrace{\lim_{k\to\infty} (y_{n_k} - x_{n_k})}_{\to 0} + \lim_{k\to\infty} x_{n_k} = x\end{aligned}$$

By continuity in Theorem (9.2),

$$f(x) = \lim_k f(x_{n_k}) = \lim_k f(y_{n_k})$$

then

$$\lim_k [f(x_{n_k}) - f(y_{n_k})] = 0$$

which is a contradiction. □

9.6 Continuity and Connectedness

Theorem 9.7

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ be connected and $f : D \to \mathbb{R}^M$ is continuous, then $f(D)$ is connected.

Proof. SFAC $\{U, V\}$ is a disconnection for $f(D)$. Since f is continuous, by Theorem (9.3), there are open sets $\tilde{U}$ and $\tilde{V} \subseteq \mathbb{R}^N$ such that

$$f^{-1}(U) = C \cap \tilde{U} \quad \text{and} \quad f^{-1}(V) = C \cap \tilde{V}$$

Then the pair $\{\tilde{U}, \tilde{V}\}$ is a disconnection for D. Contradiction. □

9.6.1 Intermediate Value Theorem

Corollary 9.2: Intermediate Value Theorem

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ be connected, $f : D \to \mathbb{R}$ be continuous. Then $f(D)$ is an interval. In particular, if $x_1, x_2 \in D$ such that $f(x_1) < c < f(x_2)$ for some $c \in \mathbb{R}$, then there exists $d \in D$ such that $f(d) = c$.

10 Differentiability on $\mathbb{R}^N$

We wish to introduce a notion of differentiability for functions $f : D \to \mathbb{R}^M$, $D \subseteq \mathbb{R}^N$ open extending the corresponding notion for real-valued functions in one variable.
Recall: If $f : (a, b) \to \mathbb{R}$ and $x_0 \in (a, b)$ then we say f is differentiable at x_0 if

$$\lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

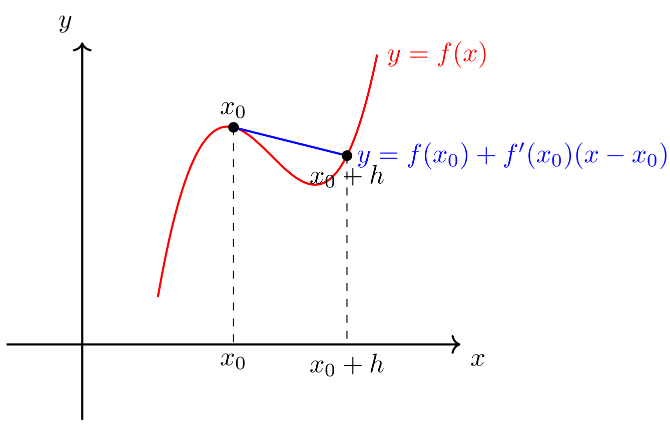
exists, and the derivative at x_0 is

$$f'(x_0) = \lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

The derivative $f'(x_0)$ gives us information such as:

- the minimum and maximum of the function,
- if the function is increasing or decreasing,
- and if $f'(x_0)$ exists then f is continuous at x_0.

The geometric intuition for a derivative is:



Here, $f'(x_0)$ is the slope of the line tangent to the graph of f at $(x_0, f(x_0))$.

Lecture 15 - Friday, Jun 7

Definition 10.1: Differentiable

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ be an open set, $f : D \to \mathbb{R}^M$ be a function. We say f is **differentiable** at $x_0 \in D$ if there exists a linear transformation $T : \mathbb{R}^N \to \mathbb{R}^M$ such that

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - T(h)\|}{\|h\|} = 0$$

Discovery 10.1

1. The numerator we have is a norm of a vector in $\mathbb{R}^M$, and the denominator is a norm of a vector in $\mathbb{R}^N$.
2. The linear transformation $T : \mathbb{R}^N \to \mathbb{R}^M$ is a nice approximation for $f(x_0 + h) - f(x_0)$. In particular, $T(0) = f(x_0 + 0) - f(x_0) = 0$. Additionally, not only

$$\lim_{h \to 0} (f(x_0 + h) - f(x_0) - T(h)) = 0$$

but also

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - T(h)\|}{\|h\|} = 0$$

10.1 Uniqueness of Derivative

Theorem 10.1: Uniqueness of Derivative

Let $\varnothing \neq D \subseteq \mathbb{R}^N$, $f : D \to \mathbb{R}^M$ be a function. Suppose $A_1, A_2 : \mathbb{R}^N \to \mathbb{R}^M$ are linear transformations such that

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - A_i(h)\|}{\|h\|} = 0 \qquad \text{for } i = 1, 2$$

then $A_1 = A_2$.

Proof. For h with $x_0 + h \in D$ we have

$$\|A_1 h - A_2 h\| \leq \|A_1 h - [f(x_0 + h) - f(x_0)]\| + \|[f(x_0 + h) - f(x_0)] - A_2 h\|$$

Hence we have

$$\lim_{h \to 0} \frac{\|A_1 h - A_2 h\|}{\|h\|} = 0$$

Fix $h \in \mathbb{R}^N$, $h \neq 0$, and $t \in \mathbb{R}$, $t > 0$. By linearity, we have

$$\frac{\|A_1(th) - A_2(th)\|}{\|th\|} = \frac{\|A_1 h - A_2 h\|}{\|h\|}$$

Taking the limit of $t \to 0$, we can get that

$$\frac{\|A_1 h - A_2 h\|}{\|h\|} = \lim_{t\to 0} \frac{\|A_1(th) - A_2(th)\|}{\|th\|} = 0$$

which suggests that $A_1(h) = A_2(h)$. □

Definition 10.2: Differential

If f is differentiable at $x_0 \in D$, we call the (unique) linear transformation $T : \mathbb{R}^N \to \mathbb{R}^M$ satisfying Definition (10.1) the **differential** of f at x_0. We denote it by $(Df)(x_0)$, also $(Df)_{x_0}$ or $f'(x_0)$. Thus $Df(x_0) : \mathbb{R}^N \to \mathbb{R}^M$ is a linear transformation. We say that f is differentiable in D if f is differentiable at all $x \in D$.

Result 10.1

$$f(x_0 + h) = f(x_0) + Df(x_0) \cdot h + \text{Error}(h)$$

where

$$\lim_{h\to 0} \frac{\|\text{Error}(h)\|}{\|h\|} = 0$$

Recall from Linear Algebra. Let $\{e_1, e_2, \ldots, e_N\}$ and $\{u_1, u_2, \ldots, u_M\}$ be the standard basis for $\mathbb{R}^N$ and $\mathbb{R}^M$ respectively. A linear transformation $T : \mathbb{R}^N \to \mathbb{R}^M$ is determined by a matrix $A \in \mathcal{M}_{M,N}(\mathbb{R})$, $A = (\alpha_{ij})$, where

$$A = \begin{bmatrix} | & | & & | \\ T(e_1) & T(e_2) & \cdots & T(e_N) \\ | & | & & | \end{bmatrix}$$

so that if we regard $v \in \mathbb{R}^N$ as a column vector, we have

$$T\mathbf{v} = A\mathbf{v} = A \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix}.$$

If $T : \mathbb{R}^N \to \mathbb{R}^M$ and $S : \mathbb{R}^M \to \mathbb{R}^K$, and $A \in M_{m\times n}(\mathbb{R})$ represents T and $B \in M_{m\times k}(\mathbb{R})$ represents S. Then $ST\mathbf{v} = BA\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^N$. That is, the matrix BA represents the linear transformation $ST : \mathbb{R}^N \to \mathbb{R}^K$. We have

$$\|T\| := \sup_{\|\mathbf{v}\|\le 1} \|T\mathbf{v}\| < \infty \quad \text{and} \quad \|T\mathbf{v}\| \le \|T\|\|\mathbf{v}\|$$

holds for every vector $\mathbf{v} \in \mathbb{R}^N$.

Example 10.1

Consider $N = 2$, $M = 1$ and let $D \subseteq \mathbb{R}^2$ open, $f : D \to \mathbb{R}$. Suppose that f is differentiable at $x_0 \in D$.

Then $(Df)(x_0)$ is determined by $(a, b) \in \mathcal{M}_{1,2}(\mathbb{R})$ for $a, b \in \mathbb{R}$:

$$f(x_0 + (h_1, h_2)) \approx f(x_0) + \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

$$f(x_0 + (h_1, h_2)) \approx \underbrace{f(x_0) + ah_1 + bh_2}_{\text{equation of a plane in } \mathbb{R}^3}$$

The graph of f is a surface in $\mathbb{R}^3$. Near the point $(x_0, f(x_0))$, the graph of f is approximated by the tangent plane at $(x_0, f(x_0))$.

Lecture 16 - Monday, Jun 10

Recall that if $T : \mathbb{R}^N \to \mathbb{R}^M$ is a linear transformation, then

$$\|T\| := \sup\{\|Tv\| : \|v\| \leq 1\} \ll \infty$$

Moreover,

1. $\|T\| = 0$ if and only if $T = 0$;
2. $\|\alpha T\| = |\alpha| \, \|T\|$;
3. $\|T + S\| = \|T\| + \|S\|$.

It follows that for all $h \in \mathbb{R}^N$,

$$\|T(h)\| \leq \|T\| \, \|h\|$$

because T is linear and if $\frac{h}{\|h\|}$ has norm 1, then

$$\left\| T\left(\frac{h}{\|h\|}\right) \right\| \leq \|T\| \quad \Rightarrow \quad \|T(h)\| \leq \|T\| \, \|h\|$$

Now we have the following theorem:

Theorem 10.2

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ be open, $f : D \to \mathbb{R}^M$ be differentiable at $x_0 \in D$, then f is continuous at x_0.

Proof. By the definition of differentiability (10.1), we have

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - (Df)(x_0)(h)\|}{\|h\|} = 0$$

Hence we have that

$$\lim_{h \to 0} \|f(x_0 + h) - f(x_0) - (Df)(x_0)(h)\| = 0$$

Then

$$\begin{aligned} 0 &\leq \|f(x_0 + h) - f(x_0)\| \\ &\leq \|f(x_0 + h) - f(x_0) - (Df)(x_0)(h)\| + \|(Df)(x_0)(h)\| \end{aligned}$$

Taking the limit as $h \to 0$ and using that $(Df)(x_0)$ is continuous (because it is linear) yields that

$$\lim_{h \to 0} \|f(x_0 + h) - f(x_0)\| = 0$$

which suggests that f is continuous at x_0. □

Example 10.2: What is the differential of a linear tranformation $T : \mathbb{R}^N \to \mathbb{R}^M$

Suppose $N = M = 1$, $T(x) = \alpha x$ for some $\alpha \in \mathbb{R}$ for all $x \in \mathbb{R}$. Then $T'(x) = T$ is linear transformation on $\mathbb{R}$ for every $x \in \mathbb{R}$. In general for $T : \mathbb{R}^N \to \mathbb{R}^M$, we have for all $h \in \mathbb{R}^N$ and $x_0 \in \mathbb{R}^N$, we have

$$T(x_0 + h) - T(x_0) - T(h) = 0$$

In particular, $(DT)(x_0) = T$.

Example 10.3

Let $f : \mathbb{R}^N \supseteq D \to \mathbb{R}^M$ be a function and write $f = (f_1, f_2, \ldots, f_M)$, where $f_j : D \to \mathbb{R}$ for all $j = 1, 2, \ldots, M$. A linear transformation $T : \mathbb{R}^N \to \mathbb{R}^M$ is determined by the vector

$$v := T(1)$$

Then T is the differential of f at $x_0 \in D$ if and only if

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - h \cdot v\|}{\|h\|} = 0$$

It follows that f is differentiable at x_0 if and only if each component f_j is, in which case

$$(Df)(x_0) = \begin{bmatrix} f_1'(x_0) \\ f_2'(x_0) \\ \vdots \\ f_M'(x_0) \end{bmatrix}$$

determined by the derivative of its components.

10.2 Chain Rule

Theorem 10.3: Chain Rule

Suppose $\varnothing \neq D \subseteq \mathbb{R}^N$ is open. If $f : D \to \mathbb{R}^M$, $f(D) \subseteq V$, $V \subseteq \mathbb{R}^M$ is open, $g : V \to \mathbb{R}^K$. If f is differentiable at $x_0 \in D$, g is differentiable at $f(x_0)$, then $g \circ f$ is differentiable at x_0 and

$$D(g \circ f)(x_0) = (Dg)(f(x_0))(Df)(x_0)$$

Note: On the right hand side, we have the product of linear transformation $\mathbb{R}^N \to \mathbb{R}^M$ and $\mathbb{R}^M \to \mathbb{R}^K$. On the left hand side we have a function $\mathbb{R}^N \to \mathbb{R}^K$.

Proof. Let us write $y_0 = f(x_0)$,

$$A = (Df)(x_0) \quad \text{and} \quad B = (Dg)(f(x_0))$$

we wish to show that

$$\lim_{h \to 0} \frac{\|g(f(x_0+h)) - g(f(x_0)) - BA(h)\|}{\|h\|} = 0$$

We have for $h \in \mathbb{R}^N$ such that $f(x_0 + h)$ is defined,

$$g(f(x_0+h)) - g(f(x_0)) - BA(h) = g(y_0 + k) - g(y_0) - BA(h)$$

where $k = f(x_0 + h) - f(x_0)$. Since $B = (Dg)(y_0)$, given $\varepsilon > 0$, there exists $\delta_1 > 0$ such that $g(y_0 + k')$ is defined and

$$\|g(y_0 + k') - g(y_0) - B(k')\| < \varepsilon \|k'\|$$

whenever $\|k'\| < \delta_1$. Since f is continuous at x_0, we can find $\delta_2 > 0$ such that of $h \in \mathbb{R}^N$ and $\|h\| < \delta_2$, then $f(x_0 + h)$ is defined and

$$\|k\| = \|f(x_0+h) - f(x_0)\| < \delta_1$$

Because $A = (Df)(x_0)$, we can find $\delta_3 > 0$ such that $f(x_0 + h)$ is defined and

$$\|k - A(h)\| < \varepsilon' \|h\|$$

where $\varepsilon' = \min\{\frac{\varepsilon}{\|B\|}, \varepsilon\}$. Take $\delta = \min\{\delta_2, \delta_3\}$, if $\|h\| < \delta$, then

$$\|B(k - A(h))\| \leq \|B\| \|k - A(h)\| < \varepsilon \|h\|$$

We also have

$$\|k\| < \|k - A(h)\| + \|A(h)\| < \varepsilon \|h\| + \|A\| \|h\| \tag{1}$$

and $\|k\| < \delta_1$. So we have

$$\|g(y_0 + k) - g(y_0) - BA(h)\| \leq \|g(y_0 + k) - g(y_0) - B(k)\| + \|B(k) - BA(h)\| < \varepsilon \|k\| + \varepsilon \|h\|$$

Then

$$\begin{aligned}
\frac{\|g(y_0 + k) - g(y_0) - BA(h)\|}{\|h\|} &< \frac{\varepsilon \|k\|}{\|h\|} + \varepsilon \\
&< \frac{\varepsilon(\varepsilon \|h\| + \|A\| \|h\|)}{\|h\|} + \varepsilon \\
&= \varepsilon^2 + (1 + \|A\|)\varepsilon && \text{by (1)}
\end{aligned}$$

This shows that

$$\lim_{h \to 0} \frac{\|g(y_0 + h) - g(y_0) - BA(h)\|}{\|h\|} = 0$$

□

10.3 Partial Derivative

Recall that $\{e_1, \ldots, e_N\}$ and $\{u_1, \ldots, u_M\}$ denote the standard basis of $\mathbb{R}^N$ and $\mathbb{R}^M$ respectively. For $f : D \to \mathbb{R}^M$, $\varnothing \neq D \subseteq \mathbb{R}^N$, $f = (f_1, \ldots, f_M)$ where $f_j : D \to \mathbb{R}$ is the j^{th} component of f.

Definition 10.3: Partial Derivative

For each $1 \leq i \leq N$ and $1 \leq j \leq M$, we define for $x_0 \in D$,

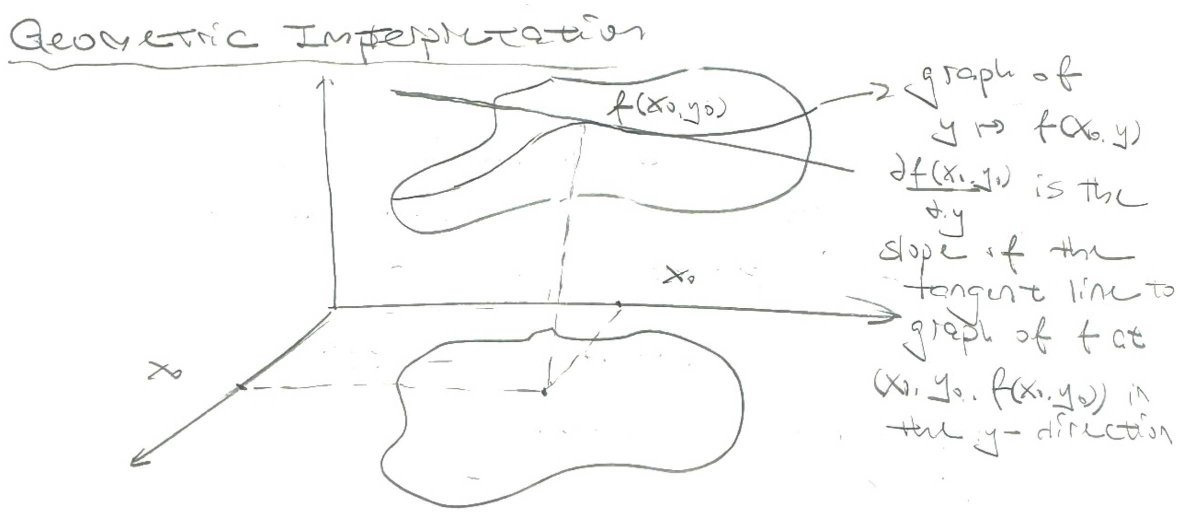
$$\frac{\partial f_j(x_0)}{\partial x_i} = \lim_{t \to 0} \frac{f_j(x_0 + te_i) - f_j(x_0)}{t}$$

provided that the limit exists. $\frac{\partial f_j(x_0)}{\partial x_i}$ is the derivative of f_j at x_0 in the x_i direction, and it is called **partial derivative** of f at x_0.
Further notation: $(D_i f_j)(x_0)$. If $M = 1$, we have $\frac{\partial f(x_0)}{\partial x_i}$, or $(D_i f)(x_0)$.

Discovery 10.2

It may happen that all partial derivative of f at x_0 exist, but f is not contimuous at x_0. But if f is differentiable at x_0, then its partial derivatives determine $(Df)(x_0)$.

10.3.1 Geometrix Interpretation



Algorithm 10.1: How do we calculate partial derivative?

We treat the variables $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N$ as constants.

Example 10.4

Let $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = e^x + x\cos(xy)$, then

$$\frac{\partial f}{\partial y}(x, y) = -x^2\cos(xy) \qquad \frac{\partial f}{\partial x}(x, y) = e^x + \cos(xy) - xy\cos(xy)$$

Example 10.5: This is related to discovery (10.2)

Let $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & (x, y) \neq (0, 0) \\ 0 & \text{otherwise} \end{cases}$. The partial derivatives of f at (x, y) exist if $(x, y) \neq (0, 0)$; If $(x, y) = (0, 0)$, we have

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

The partial derivatives of f exist at every point, but f is not continuous at $(0, 0)$.

Recall if $T : \mathbb{R}^N \to \mathbb{R}^M$, then the matrix of T with respect to the standard basis is given by

$$\begin{bmatrix} | & | & & | \\ T(e_1) & T(e_2) & \cdots & T(e_N) \\ | & | & & | \end{bmatrix} = (a_{ji})_{j,i}$$

where $T(e_i) = \sum_{j=1}^{M} a_{ji}u_j$.

Theorem 10.4

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ be open and $f : D \to \mathbb{R}^M$ be differentiable at $x_0 \in D$, then all the partial derivatives $\frac{\partial f_j(x_0)}{\partial x_i}$ of f at x_0 exist and

$$(Df)(x_0)(e_i) = \sum_{j=1}^{M} \frac{\partial f_j(x_0)}{\partial x_i}(u_j)$$

As a consequence, the matrix of $(Df)(x_0)$ with respect to the standard basis is given by

$$\begin{bmatrix} \frac{\partial f_1(x_0)}{\partial x_1} & \frac{\partial f_1(x_0)}{\partial x_2} & \cdots & \frac{\partial f_1(x_0)}{\partial x_N} \\ \frac{\partial f_2(x_0)}{\partial x_1} & \ddots & & \\ \vdots & & \ddots & \\ \frac{\partial f_M(x_0)}{\partial x_1} & & & \frac{\partial f_M(x_0)}{\partial x_N} \end{bmatrix} = \left(\frac{\partial f_j(x_0)}{\partial x_i}\right)_{j,i}$$

Proof. We know that

$$\lim_{t\to 0} \frac{\|f(x_0 + te_i) - f(x_0) - (Df)(x_0)(te_i)\|}{|t|} = 0$$

Using linearity of $(Df)(x_0)$, the above yields

$$\lim_{t\to 0} \frac{f(x_0 + te_i) - f(x_0)}{t} = (Df)(x_0)(e_i)$$

This implies that $\frac{\partial f_j(x_0)}{\partial x_i}$ exists for all $j = 1, \ldots, M$ and

$$(Df)(x_0)(e_i) = \left(\frac{\partial f_1(x_0)}{\partial x_i}, \ldots, \frac{\partial f_M(x_0)}{\partial x_i}\right) = \sum_{j=1}^{M} \frac{\partial f_j(x_0)}{\partial x_i}(u_j)$$

□

Definition 10.4: Jacobian Matrix

The matrix $\left[\frac{\partial f_j(x_0)}{\partial x_i}\right]_{j,i}$ is called the **Jacobian Matrix** of f at x_0 and denoted by $J_f(x_0)$.

Example 10.6

Let $\gamma : (a, b) \to D$ for $\varnothing \neq D \subseteq \mathbb{R}^N$ is open, suppose γ is differentiable in (a, b). Let $f : D \to \mathbb{R}$ be differentiable in D. Combining the chain rule (10.3) with the above theorem, we obtain that $g = f \circ \gamma$ is differentiable in (a, b) and

$$\begin{aligned} g'(t) &= (f \circ \gamma)'(t) \\ &= \left[\frac{\partial f(\gamma(t))}{\partial x_1} \cdots \frac{\partial f(\gamma(t))}{\partial x_N}\right] \begin{bmatrix} \gamma_1'(t) \\ \vdots \\ \gamma_N'(t) \end{bmatrix} = \sum_{i=1}^{N} \frac{\partial f(\gamma(t))}{\partial x_i} \gamma_i'(t) \end{aligned}$$

Definition 10.5: Gradient Notation

Let $f : D \to \mathbb{R}$ for $D \subseteq \mathbb{R}^N$ open, f differentiable at $x_0 \in D$, then $(Df)(x_0)$ is a $\mathcal{M}_{1,N}(\mathbb{R})$, $(Df)(x_0) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_N}\right)$ is called the **gradient** of f at x_0 and denoted as $\nabla f(x_0)$. Notice that if $f : D \to \mathbb{R}^M$, then

$$(Df)(x_0) = \begin{bmatrix} \nabla f_1(x_0) \\ \vdots \\ \nabla f_M(x_0) \end{bmatrix}$$

Lecture 18 - Friday, Jun 14

Definition 10.6: Directional Derivative

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ be open, $f : D \to \mathbb{R}^M$ be a function. Let $x_0 \in D$ and $v \in \mathbb{R}^N$ a unit (i.e, $\|v\| = 1$). The **directional derivative** of f in the direction of v at x_0 is given by

$$(D_v f)(x_0) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t}$$

provided that the limit exists.

Discovery 10.3

If $v = e_i$, then $(D_v f)(x_0) = \frac{\partial f}{\partial x_i}(x_0)$ is the partial derivative.

Theorem 10.5

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ be open, $f : D \to \mathbb{R}$ be a function differentiable at $x_0 \in D$. Then the directional derivative of f at x_0 exists for every unit vector $v \in \mathbb{R}^N$, and

$$(D_v f)(x_0) = \nabla f(x_0) \cdot v$$

Proof. Consider the function $\gamma : \mathbb{R} \to \mathbb{R}^N$, $\gamma(t) = x_0 + tv$. Then γ is differentiable in $\mathbb{R}$ and $\gamma'(t) = v$ for all $t \in \mathbb{R}$. We have $\gamma(0) = x_0$. Since D is open, we can find $\delta > 0$ such that

$$\gamma(t) \in D \qquad \text{for all } t \in (-\delta, \delta)$$

Now

$$\begin{aligned}(D_v f)(x_0) &= \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t} \\ &= \lim_{t \to 0} \frac{(f \circ \gamma)(t) - (f \circ \gamma)(0)}{t} \\ &= (f \circ \gamma)'(0)\end{aligned}$$

Example (10.6) yields

$$(f \circ \gamma)'(0) = \nabla f(\gamma(0)) \cdot \gamma'(0) = \nabla f(x_0) \cdot v$$

which is desired. □

Result 10.2

This allows for a geometric interpretation of the gradient vector. By Cauchy-Schwartz

$$\|(D_v f)(x_0)\| = \|\nabla f(x_0) \cdot v\| \leq \|\nabla f(x_0)\| \, \|v\| = \|\nabla f(x_0)\|$$

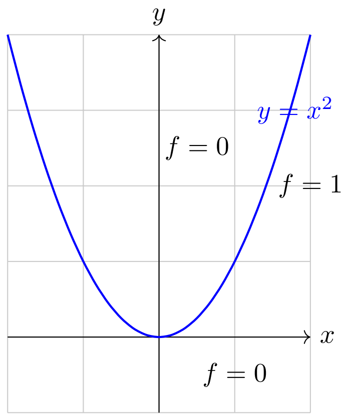
If $v = \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|}$, then $\|v\| = 1$ and

$$(D_v f)(x_0) = \|\nabla f(x_0)\|$$

So the gradient of f at x_0 points in the direction to which the slope of the tangent line to the graph of f at $(x_0, f(x_0))$ is maximal.

Example 10.7: Existence of directional derivative does not imply continuity

Consider $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = \begin{cases} 1 & 0 < y < x^2 \\ 0 & \text{otherwise} \end{cases}$,



We have $(D_v f)(0, 0) = 0$ for all unit vectors $v \in \mathbb{R}^2$, but f is not continuous at $(0, 0)$.

Recall Mean Value Theorem.

Exercise: *See more at HW3.* $f : \mathbb{R}^N \to \mathbb{R}^M$ is differentiable at $x_0 \in D$ if and only if the j^{th} component of f, $f_j = \mathbb{R}^N \to \mathbb{R}$ is differentiable at x_0 for all $j = 1, \ldots, M$.

Theorem 10.6: Sufficient Condition for Differentiability

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ be open, $f : D \to \mathbb{R}^M$, $x_0 \in D$. Suppose that all partial derivatives of f, $\frac{\partial f}{\partial x_i}$, exist in D and are continuous at x_0. Then f is differentiable at x_0.

Proof. We can assume $M = 1$. We know f is differentiable at x_0 if and only if

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - \nabla f(x_0) \cdot h\|}{h} = 0$$

Let $\varepsilon > 0$ be given. Since each $\frac{\partial f}{\partial x_i}$ is continuous at x_0, there exists $\delta > 0$ such that if $|z - x_0| < \delta$, then $z \in D$ and

$$\left| \frac{\partial f(z)}{\partial x_i} - \frac{\partial f(x_0)}{\partial x_i} \right| < \frac{\varepsilon}{N} \qquad i = 1, \ldots, N$$

Fix $h \in \mathbb{R}^N$ with $\|h\| < \delta$ and write $h = (h_1, \ldots, h_N)$. For each $k = 1, \ldots, N$, set

$$v_k = \sum_{i=1}^{k} h_i e_i = (h_1, \ldots, h_k, \ldots, 0_{N-k})$$

We also set $v_0 = 0$. Now $v_k = v_{k-1} + h_k e_k$ for $k = 1, \ldots, N$ and $\|v_k\| < \delta$ for all $k = 0, \ldots, N$. Now

$$f(x_0 + h) - f(x_0) - f(x_0 + v_{k-1}) + f(x_0 + v_{k-1}) = \sum_{k=1}^{N} f(x_0 + v_k) - f(x_0 + v_{k-1})$$

Fix $k = 1$, we have $x_0 + v_k, x_0 + v_{k-1} \subseteq \mathcal{B}_\delta(x_0)$. Since $\mathcal{B}_\delta(x_0)$ is convec, it follows that

$$t(x_0 + v_k) + (1-t)(x_0 + v_{k-1}) \in \mathcal{B}_\delta(x_0) \qquad \forall\, t \in [0,1]$$

For all $t \in [0,1]$,

$$x_0 + v_{k-1} + th_k e_k \in \mathcal{B}_\delta(x_0)$$

Hence the function

$$t \mapsto f(x_0 + v_{k-1} + th_k e_k)$$

is continuous on $[0,1]$ and differentiable in $(0,1)$ because $\frac{\partial f}{\partial x_k}$ exists in D. Set $g_k : [0,1] \to \mathbb{R}$, $g_k(t) = f(x_0 + v_{k-1} + th_k e_k)$, we have $g_k(1) = f(x_0 + v_k)$ and $g_k(0) = f(x_0 + v_{k-1})$. By Mean Value Theorem, there exists $c_k \in (0,1)$ such that

$$h_k \frac{\partial f}{\partial x_k}(x_0 + v_{k-1} + c_k h_k e_k) = g_k'(c_k) = f(x_0 + v_k) - f(x_0 + v_{k-1})$$

Thus

$$\begin{aligned} & f(x_0 + v_k) - f(x_0 + v_{k-1}) - \frac{\partial f(x_0)}{\partial x_k} h_k \\ =& h_k \frac{\partial f}{\partial x_k}(x_0 + v_{k-1} + c_k h_k e_k) - \frac{\partial f(x_0)}{\partial x_k} h_k \end{aligned}$$

and

$$\begin{aligned} & \left| f(x_0 + v_k) - f(x_0 + v_{k-1}) - \frac{\partial f(x_0)}{\partial x_k} h_k \right| \\ =& \left| h_k \frac{\partial f}{\partial x_k}(x_0 + v_{k-1} + c_k h_k e_k) - \frac{\partial f(x_0)}{\partial x_k} h_k \right| < h_k \cdot \frac{\varepsilon}{N} \le \|h\| \cdot \frac{\varepsilon}{N} \\ =& |h_k| \left| \frac{\partial f}{\partial x_k}(x_0 + v_{k-1} + c_k h_k e_k) - \frac{\partial f(x_0)}{\partial x_k} \right| \end{aligned}$$

now we have

$$\begin{aligned} \frac{\left| f(x_0 + h) - f(x_0) - \sum_{k=1}^N \frac{\partial f(x_0)}{\partial x_k} h_k \right|}{\|h\|} &= \frac{\left| \sum_{k=1}^N \left(f(x_0 + v_k) - f(x_0 + v_{k-1}) - \frac{\partial f(x_0)}{\partial x_k} h_k \right) \right|}{\|h\|} \\ &< \sum_{k=1}^N \frac{\|h\| \cdot \varepsilon}{\|h\| \cdot N} = \varepsilon \end{aligned}$$

□

Lecture 19 - Monday, Jun 17

Example 10.8

Let $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = \begin{cases} (x^2 + y^2) \sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) & (x, y) \neq (0, 0) \\ 0 & \text{otherwise} \end{cases}$. If $(x, y) \neq (0, 0)$, we have

$$\begin{aligned} \frac{\partial f(x, y)}{\partial x} &= 2x \sin\left(\frac{1}{\sqrt{x^2 + y^2}}\right) + (x^2 + y^2) \cos\left(\frac{1}{\sqrt{x^2 + y^2}}\right) \left(-\frac{1}{2}\right) \frac{1}{(x^2 + y^2)^{3/2}} (2x) \\ &= 2x \sin\left(\frac{1}{\sqrt{x^2 + y^2}}\right) - \frac{x}{\sqrt{x^2 + y^2}} \cos\left(\frac{1}{\sqrt{x^2 + y^2}}\right) \end{aligned}$$

At $(0, 0)$, we have

$$\begin{aligned} \frac{\partial f(0, 0)}{\partial x} &= \lim_{(h_1,h_2)\to(0,0)} \frac{\|(h_1, h_2)\|^2 \sin\left(\frac{1}{\|(h_1,h_2)\|}\right)}{\|(h_1, h_2)\|} \\ &= \lim_{(h_1,h_2)\to(0,0)} \|(h_1, h_2)\| \sin\left(\frac{1}{\|(h_1, h_2)\|}\right) = 0 \end{aligned}$$

by squeeze theorem. This suggests that $\frac{\partial f}{\partial x}$ is continuous at every point $(x, y) \in \mathbb{R}^2 \backslash \{(0, 0)\}$, but it is not continuous at $(0, 0)$ because, for example,

$$\lim_{n\to\infty} \frac{\partial f\left(\frac{1}{2n\pi}, 0\right)}{\partial x} = -1 \neq 0 = \frac{\partial f(0, 0)}{\partial x}$$

By Theorem (10.6), f is differentiable at every point $(x, y) \neq (0, 0)$. However, f is also differentiable at $(0, 0)$:

$$\frac{\partial f(0, 0)}{\partial x} = \lim_{t\to 0} \frac{f(t, 0) - f(0, 0)}{t} = \lim_{t\to 0} t \sin\left(\frac{1}{t}\right) = 0$$

Now we compute

$$\lim_{(h_1,h_2)\to(0,0)} \frac{|f(h_1, h_2) - f(0, 0) - 0(h_1, h_2)|}{\|(h_1, h_2)\|} = 0$$

which suggests that f is differentiable at $(0, 0)$.

10.4 Product Rule + Linearity

Proposition 10.1

Suppose $\varnothing \neq D \subseteq \mathbb{R}^N$ is open, $f, g : D \to \mathbb{R}^M$ are differentiable at $x_0 \in D$, then

$$\lambda f + g : D \to \mathbb{R}^M \qquad (\lambda f + g)(x) = \lambda f(x) + g(x)$$

is differentiable at x_0 for all $\lambda \in \mathbb{R}$, and

$$(D(\lambda f + g))(x_0) = \lambda(Df)(x_0) + (Dg)(x_0)$$

Proof. **Exercise.** □

Proposition 10.2: Product Rule

Suppose $\varnothing \neq D \subseteq \mathbb{R}^N$ is open, $f, g : D \to \mathbb{R}^M$ be functions. If f and g are differentiable at $x_0 \in D$, then

$$\underbrace{f \cdot g}_{\text{dot product}} \quad : D \to \mathbb{R} \qquad x \mapsto \underbrace{f(x) \cdot g(x)}_{\text{dot product}}$$

is differentiable at x_0, and

$$(D(f \cdot g))(x_0) = f(x_0)^T (Dg)(x_0) + g(x_0)^T (Df)(x_0)$$

In case of $M = 1$, this gives

$$\nabla(f \cdot g) = f \cdot \nabla g + g \cdot \nabla f$$

Proof. We write $v = f \cdot g = \sum_{j=1}^M f_j \cdot g_j$. If v is differentiable at x_0, then $(Dv)(x_0) = \left(\frac{\partial v}{\partial x_1}, \frac{\partial v}{\partial x_2}, \dots, \frac{\partial v}{\partial x_N}\right)$. Write

$$\frac{\partial v}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\sum_{j=1}^M f_j g_j\right) = \sum_{j=1}^M \left(\frac{\partial f_j}{\partial x_i} \cdot g_j + \frac{\partial g_j}{\partial x_i} \cdot f_j\right)$$

and this is exactly the i^{th} column of $(Dv)(x_0)$, so it suffices to show that v is differentiable at x_0. We have

$$\begin{aligned}
& v(x_0 + h) - v(x_0) - (f(x_0)^T (Dg)(x_0) + g(x_0)^T (Df)(x_0)h) \\
= \;& (f \cdot g)(x_0 + h) - (f \cdot g)(x_0) - f(x_0) \cdot g(x_0 + h) + f(x_0) \cdot g(x_0 + h) \\
& - g(x_0 + h)^T (Df)(x_0)h + g(x_0 + h)^T (Df)(x_0)h \\
& - f(x_0)^T (Dg)(x_0) - g(x_0)^T (Df)(x_0)h \\
= \;& s_1 + s_2 + s_3
\end{aligned}$$

where

$$\begin{aligned}
s_1 &= (f \cdot g)(x_0 + h) - f(x_0) \cdot g(x_0 + h) - g(x_0 + h)^T (Df)(x_0) \\
s_2 &= f(x_0)g(x_0 + h) - f(x_0)g(x_0) - f(x_0)^T (Dg)(x_0)h \\
s_3 &= (g(x_0 + h) - g(x_0))^T (Df)(x_0)h
\end{aligned}$$

Then by Cauchy-Schwartz (1.1), we have

$$\begin{aligned}
\frac{|s_1|}{\|h\|} &\leq \|g(x_0 + h)\| \cdot \frac{\|f(x_0 + h) - f(x_0) - (Df)(x_0)h\|}{\|h\|}, \\
\frac{|s_2|}{\|h\|} &\leq \|f(x_0)\| \cdot \frac{\|g(x_0 + h) - g(x_0) - (Dg)(x_0)h\|}{\|h\|}, \\
\frac{|s_3|}{\|h\|} &\leq \|g(x_0 + h) - g(x_0)\| \cdot \frac{\|(Df)(x_0)h\|}{\|h\|} \\
&\leq \|g(x_0 + h) - g(x_0)\| \cdot \|(Df)(x_0)h\|
\end{aligned}$$

Since g is continuous at 0, each summation goes to 0 as $h \to 0$. □

10.5 Higher Order Partial Derivatives

Suppose $\varnothing \neq D \subseteq \mathbb{R}^N$ open and $f : D \to \mathbb{R}$,

Definition 10.7: Second Order Partial Derivative

If $i \in \{1, \ldots, N\}$ is such that $\frac{\partial f}{\partial x_i}$ exists in D, then $\frac{\partial f}{\partial x_i}$ is a function on D. If the partial derivatives of $\frac{\partial f}{\partial x_i}$ exist, we define for $j = 1, \ldots, N$,

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

is called the **second order partial derivative of** f.

Definition 10.8

We say that $f \in C^0(D)$ if f is continuous on D, $f \in C^1(D)$ if $f \in C^0(D)$ and the partial derivatives of f exist in D and are continuous. If $f \in C^1(D)$, then f is continuously differentiable. In general, $f \in C^k(D)$ if $f \in C^{k-1}(D)$ and all $\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \cdots \partial x_{i_1}}$ are in $C^0(D)$.

Example 10.9

Suppose $f(x, y) = \frac{e^{xy}}{x}$, $(x \neq 0)$. then

$$f_x = \frac{ye^{xy}}{x} - \frac{e^{xy}}{x^2} = \left(\frac{y}{x} - \frac{1}{x^2}\right)e^{xy}$$
$$f_y = e^{xy}$$

The second order partial derivatives are

$$f_{xx} = y\left(\frac{y}{x} - \frac{1}{x^2}\right)e^{xy} + \left(\frac{-y}{x} + \frac{2}{x^3}\right)e^{xy}$$
$$f_{xy} = ye^{xy}$$
$$f_{yx} = ye^{xy}$$
$$f_{yy} = xe^{xy}$$

Discovery 10.4

Notice that $f_{xy} = f_{yx}$. In fact, partial derivatives are commutative. (See more in 10.8)

Discovery 10.5

Let $\varnothing \neq D \subseteq \mathbb{R}^N$, $N \geq 3$. Suppose $i, j \in \{1, \ldots, N\}$, $i < j$, and $\frac{\partial f}{\partial x_i}$, $\frac{\partial^2 f}{\partial x_i \partial x_j}$, $\frac{\partial f}{\partial x_j \partial x_i}$ all exist at $x_0 = (a_1, \ldots, a_N)$. We consider $g : \mathbb{R}^2 \supseteq U \to \mathbb{R}$ defined by

$$g(x, y) = f(a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_{j-1}, y_j, a_{j+1}, \ldots, a_N)$$

Then we have

$$\frac{\partial g(x, y)}{\partial x} = \frac{\partial f}{\partial x_i}(a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_{j-1}, y_j, a_{j+1}, \ldots, a_N)$$

This will allow us to assume $N = 2$ in the next theorem.

Theorem 10.7: Two Dimensional MVT

Let $\varnothing \neq D \subseteq \mathbb{R}^2$ be open, $f : D \to \mathbb{R}$ a function on D. Suppose $\frac{\partial f}{\partial x}$, $\frac{\partial^2 f}{\partial y \partial x}$ exist in D. Let $(a, b) \in D$, and let Q be a closed interval contained in D with opposite vertices (a, b) and $(a + h, b + k)$. Then there exists an interior point of Q, denoted as (x, y), such that

$$\Delta(f, Q) = hk \frac{\partial^2 f(x, y)}{\partial y \partial x}$$

where $\Delta(f, Q) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b)$.

Proof. Let $v(t) := f(t, b + k) - f(t, b)$ for $t \in [a, a + h]$ (or $[a + h, a]$). Then v is differentiable in the open interval and continuous in the closed interval. By MVT, we can find x between a and $a + h$ such that

$$v'(t) = \frac{v(a + h) - v(a)}{h} = \frac{\Delta(f, Q)}{h}$$

We know that

$$\frac{\partial f(x, b + k)}{\partial x} - \frac{\partial f(x, b)}{\partial x} = v'(x)$$

Now, the function $s \mapsto \frac{\partial f(x,s)}{\partial x}$ is continuous on the interval $[b, b + k]$ (or $[b + k, b]$) and is differentiable in the open interval because $\frac{\partial^2 f}{\partial y \partial x}$ exists in D. By MVT again, we can find y between b and $b + k$ such that

$$\frac{\partial^2 f(x, y)}{\partial y \partial x} = \frac{\frac{\partial f(x,b+k)}{\partial x} - \frac{\partial f(x,b)}{\partial x}}{k}$$

Replacing the above equation with the second one, we obtain

$$\frac{\partial^2 f(x, y)}{\partial y \partial x} = \frac{\Delta(f, Q)}{hk}$$

as desired. $\square$

Lecture 21 - Friday, Jun 21

10.5.1 Partial Derivatives are Commutative

Theorem 10.8: Partial Derivatives are Commutative

Let $\varnothing \neq D \in \mathbb{R}^2$ be open, $f : F \to \mathbb{R}$. Suppose that $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, and $\frac{\partial^2 f}{\partial y \partial x}$ all exist in D and that $\frac{\partial^2 f}{\partial y \partial x}$ is continuous at $(a,b) \in D$. Then $\frac{\partial^2 f}{\partial x \partial y}$ exists at (a,b) and

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

Proof. Set $A := \frac{\partial^2 f(a,b)}{\partial y \partial x}$, we need to show that

$$\lim_{h \to 0} \left(\frac{f_y(a+h,b) - f_y(a,b)}{h} - A \right) = 0$$

Let $\varepsilon > 0$, let $\delta' > 0$ be such that if $\mathcal{B}_{\delta'}((a,b)) \subset D$ and if $(x,y) \in \mathcal{B}_{\delta'}((a,b))$, then

$$|f_{xy}(x,y) - A| < \varepsilon$$

Let $\varepsilon > 0$ such that

$$[a-\delta, a+\delta] \times [b-\delta, b+\delta] \subset \mathcal{B}_{\delta'}((a,b))$$

Take $h, k \neq 0$ with $|h|, |k| < \delta$, then the closed rectangle Q with opposite vertices (a,b) and $(a+h, b+k)$ is contained in $\mathcal{B}_{\delta'}((a,b))$. Apply Theorem (10.7), there exists $(x,y) \in D^\circ$ such that

$$\Delta(f,Q) = hk \frac{\partial^2 f}{\partial y \partial x}(x,y)$$

Then

$$\left| \frac{\Delta(f,Q)}{hk} - A \right| < \varepsilon$$

Thus

$$\left| \frac{f(a+h,b+k) - f(a+h,b) - f(a,b+k) + f(a,b)}{hk} - A \right| < \varepsilon$$

Take limit as $k \to 0$, we get

$$\left| \frac{f_y(a+h,b) - f_y(a,b)}{h} - A \right| < \varepsilon$$

since $0 \neq h \in D$, $|h| < \delta$. This shows that $f_{yx}(a,b)$ exists and

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

□

Corollary 10.1: Clairaut's Theorem

Let $\varnothing \neq D \in \mathbb{R}^2$ be open, $f : F \to \mathbb{R}$ in $C^2(D)$. Then

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \qquad \forall\, 1 \leq i, j \leq N$$

Proof. This follows Theorem (10.8) and Discovery (10.5). □

11 Vector Fields

Definition 11.1: Vector Field

A vector field is simply a function $v : \mathbb{R}^N \supset D \to \mathbb{R}^N$.

Example 11.1: Important Example

Suppose $f : D \to \mathbb{R}$ is differentiable, then

$$\nabla f : D \to \mathbb{R}^N, \quad x \in D \mapsto \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_N}\right)$$

is a vector field called the **gradient field**.

Proposition 11.1

Suppose that $v : D \to \mathbb{R}^N$ for D open is a vector field of class 1 (in $C^1(D)$). Then a necassary condition for v to be a gradient field is that

$$\frac{\partial v_j}{\partial x_i} = \frac{\partial v_i}{\partial x_j} \qquad \forall\, 1 \leq i, j \leq N$$

Proof. Suppose $v = \nabla f$, then f must necessarily be class C^2. Then by Clairaut's Theorem (10.1),

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

□

11.1 Other Operations on a Vector Field

Definition 11.2: Divergence

Suppose $v : D \to \mathbb{R}^N$ is a differentiable vector field, then the divergence of v is

$$\begin{aligned} \operatorname{div}(v) &= \sum_{i=1}^{N} \frac{\partial v_i}{\partial x_i} = \nabla \cdot v \\ &= \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_N} \right) \cdot (v_1, \ldots, v_N) \end{aligned}$$

Remark: the div corresponds to taking the trace of the Jacobian of v.

Definition 11.3: Laplace Operator

If $f : D \to \mathbb{R}$ is of class C^2, the **Laplace Operator** is

$$\Delta f = \operatorname{div}(\underbrace{\text{grad } f}_{\nabla f}) = \sum_{i=1}^{N} \frac{\partial^2 f}{\partial x_i^2}$$

Definition 11.4: Harmonic

A function $f : D \to \mathbb{R}$ is said to be **Harmonic** if $\Delta f = 0$.

The Laplace operator appears in many partial differential equation:

Example 11.2: Heat Equation and Wave Equation

Let $D \subset \mathbb{R}^N$, $f : D \times (0, \infty) \to \mathbb{R}$, $f(x, t)$ for $x \in D$ and $t \in (0, \infty)$ (think of this as "time"). The **heat** equation is

$$\frac{\partial f}{\partial t} = k \Delta f$$

The **wave** equation is

$$\frac{\partial^2 f}{\partial t^2} = k \Delta f$$

11.2 Derivative as Linear Approximation

Suppose $N = 1$. Recall that $f'(x_0)$ is the derivative of f at x_0, and

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + R_{x_0}(h),$$

for some error function $R_{x_0}(h)$, where $h = x - x_0$ and $\lim_{h\to 0} \frac{R_{x_0}(h)}{h} = 0$. If $f : D \to \mathbb{R}, D \subseteq \mathbb{R}^N, N \geq 2$ and f is differentiable at x_0, then

$$f(x) = f(x_0) + (Df)(x_0)(x - x_0) + R_{x_0}(h)$$

where $h = x - x_0$ and $\lim_{h\to 0} \frac{\|R_{x_0}(h)\|}{\|h\|} = 0$. The function $L : \mathbb{R}^N \to \mathbb{R}$,

$$L(x) = f(x_0) + (Df)(x_0)(x - x_0)$$

is the linear approximation of f at x_0. If $N = 2$, then for $(x_0, y_0) \in D$,

$$\begin{aligned} L(x) &= f(x_0, y_0) + \nabla f(x_0, y_0) \cdot (x - x_0, y - y_0) \\ &= f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \end{aligned}$$

is the tangent plane to the graph of f.

Lecture 22 - Monday, Jun 24

12 Taylor's Theorem

12.1 Single Variable Taylor's Theorem

We wish to prove a version of Taylor's Theorem for functions of several variables.

Theorem 12.1: Taylor's Theorem (one variable case)

Let $n \geq 1$ and let $f : (a, b) \to \mathbb{R}$ be n-times differentiable in (a, b). Let $x_0 \in (a, b)$, then for each $x \in (a, b)$, $x \neq x_0$, there exists ξ lying between x_0 and x such that

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \frac{f^{(n)}(\xi)}{n!}(x - x_0)^n$$

Proof. We let $x \neq x_0$, we prove by induction on n:

1. *Base Case:*
   When $n = 1$, the statement is the MVT.

2. *Induction Step:*
   Suppose $n \geq 2$ and write

$$p(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(t)}{k!}(t - x_0)^k$$

for $t \in \mathbb{R}$. Set

$$M := \frac{f(x) - p(x)}{(x - x_0)^n}$$

such that $f(x) = p(x) + M(x - x_0)^n$. We need to show that $M = f^{(n)}(s)/n!$ for some s between x_0 and x. Or equivalently, $f^{(n)}(s) = n!M$. Consider $g(t) = f(t) - p(t) - M(t - x_0)^n$, then $g(x_0) = 0$. Moreover, for $k = 1, \ldots, n-1$, we have

$$g^{(k)}(x_0) = f^{(k)}(x_0) - p^{(k)}(x_0) = 0$$

because $p^{(k)}(x_0) \equiv f^{(k)}(x_0)$ for $k = 1, \ldots, n-1$. Now

$$g^{(n)}(t) = f^{(n)}(t) - n!M$$

So we need to find ξ between x_0 and x such that $g^{(n)}(\xi) = 0$. Since $g(x) = 0$ by our choice of M, by MVT, there exists x_1 between x_0 and x such that $g'(x_1) = 0$. Since $g'(x_0) = 0$ and $g'(x_1) = 0$, again, by MVT, there exists x_2 lying between x_0 and x_1 such that $g''(x_2) = 0$. Continuing with this process, after $n-1$ steps we obtain a point x_{n-1} between x_0 and x such that $g^{(n-1)}(x_{n-1}) = 0$. Since $g^{(n-1)}(x_0) = 0$, we apply MVT again and get x_n lying between x_0 and x_{n-1} such that $g^{(n)}(x_n) = 0$. Setting $\xi := x_n$, we get

$$\frac{f^{(n)}(\xi)}{n!} = M$$

□

Corollary 12.1: Second Derivative Test

Let $f \in C^2((a,b))$. Let $x_0 \in (a,b)$ be such that $f'(x_0) = 0$. Then

1. if $f''(x_0) < 0$, then x_0 is a local maximum of f;
2. if $f''(x_0) > 0$, then x_0 is a local minimum of f;

Proof. Since f'' is continuous, then there exists $\delta > 0$ such that $(x-\delta, x+\delta) \subset (a,b)$ and $f''(x) < 0$ whenever $|x - x_0| < \delta$. Now let x with $|x - x_0| < \delta$. By Taylor's Theorem, there exists ξ between x_0 and x such that

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(\xi)}{2}(x - x_0)^2 \\ &= f(x_0) + \frac{f''(\xi)}{2}(x - x_0)^2 \end{aligned}$$

Since $f''(\xi) < 0$, we get $f(x) < f(x_0)$, which implies that $f(x_0)$ is a local maximum. □

12.2 Multivariable Taylor's Theorem

Definition 12.1: Notation: Multiindex

For $n \geq 0$, we let $\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{N}_0^N$ (including 0) with $\alpha_1 + \cdots + \alpha_N = N$. For $\alpha \in \mathbb{N}_0^N$, we write

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_N^{\alpha_N}$$

for $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$. we define

$$|\alpha| := \alpha_1 + \cdots + \alpha_N \quad \text{and} \quad \alpha! := \alpha_1! \cdots \alpha_N!$$

For $\alpha \in \mathbb{N}_0^N$ a **multiindex**, we write

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_N^{\alpha_N}} \qquad \text{for } f \in C^{|\alpha|},\ |\alpha| \leq n$$

Example 12.1

For an example, we have

$$D^{(1,2,1)} f = \frac{\partial^4 f}{\partial x_1 \partial x_2^2 \partial x_3} \qquad \text{and} \qquad D^{(0,1,0)} = \frac{\partial f}{\partial x_2}$$

Let $(l_1, l_2, \ldots, l_n)$ be an n-tuple in $\{1, 2, \ldots, N\}^n$. For each $k = 1, \ldots, N$, we let α_k be the number of times k appears in $(l_1, \ldots, l_n)$. Then $\alpha = (\alpha_1, \ldots, \alpha_N)$ is a multiindex with $\alpha_1 + \cdots + \alpha_N = n$. If f is of class C^n, it follows from Clairaut's Theorem that

$$\frac{\partial^n f}{\partial x_{i_1} \cdots \partial x_{i_n}} = D^\alpha f$$

If $\alpha = (\alpha_1, \ldots, \alpha_N)$ be a multiindex of $\alpha_1 + \cdots + \alpha_N = n$, there are exactly $\frac{n!}{\alpha!}$ n-tuples whose associated multiindex as above is α. This follows from the multinomial theorem:

$$(x_1 + \cdots + x_N)^n = \sum_{\alpha_1 + \cdots + \alpha_N = n} \frac{n!}{\alpha!} x^\alpha$$

Lecture 23 - Wednesday, Jun 26

Theorem 12.2: Taylor's Theorem (N-variable)

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ be open, $f : D \to \mathbb{R}$, $f \in C^n(D)$ for $n \geq 1$. Let $x_0 \in D$ and let $\xi \in \mathbb{R}^N$ be such that $x_0 + t\xi \in D$ for all $t \in [0,1]$ (line segment between x_0 and $x_0 + \xi$). Then there exists $\theta \in (0,1)$ such that

$$f(x_0 + \xi) = \sum_{|\alpha| \leq n-1} \frac{D^\alpha f(x_0)}{\alpha!} \xi^\alpha + \sum_{|\alpha| = n} \frac{D^\alpha f(x_0 + \theta\xi)}{\alpha!} \xi^\alpha$$

Example 12.2

Suppose $n = 1$, then

$$f(x_0 + \xi) = f(x_0) + \sum_{i=1}^{N} \frac{\partial f(x_0 + \theta\xi)}{\partial x_i}\xi_i = f(x_0) + \nabla f(x_0 + \theta\xi) \cdot \xi$$

See more in A3.

Example 12.3

Suppose $n = 2$ and $N = 2$, then

$$\begin{aligned} f(x_0 + \xi) &= f(x_0) + \nabla f(x_0) \cdot \xi + \frac{f_{xx}(x_0 + \theta\xi)\xi_1^2}{2} + \frac{f_{yy}(x_0 + \theta\xi)\xi_2^2}{2} + f_{xy}(x_0 + \theta\xi) \cdot \xi_1\xi_2 \\ &= f(x_0) + \nabla f(x_0) \cdot \xi + \frac{1}{2}(A(x_0 + \theta\xi)\xi) \cdot \xi \end{aligned}$$

where

$$A(x_0 + \theta\xi) = \begin{bmatrix} f_{xx}(x_0 + \theta\xi) & f_{xy}(x_0 + \theta\xi) \\ f_{yx}(x_0 + \theta\xi) & f_{yy}(x_0 + \theta\xi) \end{bmatrix}$$

Before proving the Theorem, we first introduce a Lemma:

Lemma 12.1

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ be open, $f : D \to \mathbb{R}$, $f \in C^n(D)$ for $n \geq 1$. Let $x_0 \in D$ and let $\xi \in \mathbb{R}^N$ be such that $x_0 + t\xi \in D$ for all $t \in [0, 1]$. Then there exists an open interval (a, b) containing $[0, 1]$ such that $g : (a, b) \to \mathbb{R}$, $g(t) = f(x_0 + t\xi)$ is in $C^n(a, b)$ and

$$g^{(n)}(t) = \sum_{|\alpha|=n} \frac{n!}{\alpha!} D^\alpha f(x_0 + t\xi) \cdot \xi^\alpha$$

Proof. The existence of $(a, b) \supset [0, 1]$ with $x_0 + t\xi \in D$ follows because F is open and $x_0 + t\xi \in D$ for all $t \in [0, 1]$. Let us first prove by induction on n that

$$g^{(n)}(t) = \sum_{i_1,\ldots,i_n=1}^{N} \frac{\partial^n f(x_0 + t\xi)}{\partial x_{i_1} \cdots \partial x_{i_n}} \xi_{i_1} \cdots \xi_{i_n}$$

which is the sum over all n-tuples in $\{1, 2, \ldots, N\}^n$

1. For $n = 0$, there is nothing to prove.
2. For $n = 1$, since $g = f \circ \gamma$, for $\gamma : (a, b) \to \mathbb{R}^N$, $\gamma(a, b) \subset D$, and $\gamma(t) = x_0 + t\xi$, the Chain Rule (10.3) implies that g is differentiable at $t \in (a, b)$ and

$$g'(t) = \nabla f(x_0 + t\xi) \cdot \xi = \sum_{i=1}^{N} \frac{\partial f(x_0 + t\xi)}{\partial x_i}\xi_i$$

3. Now suppose $n \geq 2$ and

$$g^{(n-1)}(t) = \sum_{i_1,\ldots,i_{n-1}=1}^{N} \frac{\partial^{n-1} f(x_0 + t\xi)}{\partial x_{i_1} \cdots \partial x_{i_{n-1}}} \xi_{i_1} \cdots \xi_{i_{n-1}}$$

Then again by the Chain Rule (10.3), $g^{(n-1)}$ is differentiable at $t \in (a, b)$ and

$$\begin{aligned} g^{(n)}(t) &= \sum_{i_1,\ldots,i_{n-1}=1}^{N} \frac{d}{dt} \left(\frac{\partial^{n-1} f(x_0 + t\xi)}{\partial x_{i_1} \cdots \partial x_{i_{n-1}}} \xi_{i_1} \cdots \xi_{i_{n-1}} \right) \\ &= \sum_{i_1,\ldots,i_{n-1}=1}^{N} \frac{\partial^{n} f(x_0 + t\xi)}{\partial x_{i_1} \cdots \partial x_{i_n}} \xi_{i_1} \cdots \xi_{i_n} \end{aligned}$$

By Clairaut's Theorem (10.1), since there are exactly $\frac{n!}{\alpha!}$ n-tuples whose associated multiindex is $\alpha = (\alpha_1, \ldots, \alpha_N)$, we have

$$g^{(n)}(t) = \sum_{|\alpha|=n} \frac{n!}{\alpha!} D^{\alpha} f(x_0 + t\xi) \cdot \xi^{\alpha}$$

□

Proof. This is the prove of N-varaible Taylor's Theorem (12.2). We need to find $\theta \in (0, 1)$ such that

$$f(x_0 + \xi) = \sum_{|\alpha| \leq n-1} \frac{D^{\alpha} f(x_0)}{\alpha!} \xi^{\alpha} + \sum_{|\alpha|=n} \frac{D^{\alpha} f(x_0 + \theta\xi)}{\alpha!} \xi^{\alpha}$$

Let (a, b) and $g : (a, b) \to \mathbb{R}$, $g(t) = f(x_0 + t\xi)$ be as in Lemma above. By the one variable Taylor's Theorem (12.1), there exists $\theta \in (0, 1)$ such that

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

Since

$$\begin{aligned} \frac{g^{(k)}(0)}{k!} &= \frac{1}{k!} \left(\sum_{|\alpha|=k} \frac{k!}{\alpha!} D^{\alpha} f(x_0) \cdot \xi^{\alpha} \right) \quad (k \leq n-1) \\ \text{and} \quad \frac{g^{(n)}(0)}{n!} &= \frac{1}{n!} \left(\sum_{|\alpha|=n} \frac{n!}{\alpha!} D^{\alpha} f(x_0 + \theta\xi) \cdot \xi^{\alpha} \right) \end{aligned}$$

Substituting them in above equation we get the desired expression for $f(x_0 + \xi)$. □

12.3 Multivariate Polynomial

Definition 12.2: Multivariate Polynomial

A **multivariate polynomial** $p : \mathbb{R}^N \to \mathbb{R}$ (or N-variable) of degree n is given by

$$p(\xi) = \sum_{k=0}^{n} \left(\sum_{|\alpha|=k} C_\alpha \xi^\alpha \right)$$

where $C_\alpha \neq 0$ for some α with $|\alpha| = n$.

Discovery 12.1

Notice that

$$D^\alpha p(0) = \alpha! C_\alpha \quad \Rightarrow \quad C_\alpha = \frac{D^\alpha p(0)}{\alpha!}$$

Definition 12.3: Taylor Approximation

Suppose $f \in C^{n+1}(D)$, the n^{th} **order Taylor Approximation** of f is the polynomial

$$T_{n,x_0}(\xi) = \sum_{|\alpha| \leq n} \frac{D^\alpha f(x_0)}{\alpha!} \xi^\alpha$$

and the remainder term is $f(x_0 + \xi) - T_{n,x_0}(\xi) = \sum_{|\alpha|=n+1} \frac{D^\alpha f(x_0+\theta\xi)}{\alpha!} \xi^\alpha$.

Proposition 12.1

Let $f \in C^{n+1}(D)$, D open, $f : D \to \mathbb{R}$, let $x_0 \in D$, then

$$\lim_{\xi \to 0} \frac{|R_n(\xi)|}{\|\xi\|^n} = 0$$

Lecture 24 - Friday, Jun 28

Proof. Let $r > 0$ be such that $\mathcal{B}_r[x_0] \subset D$. Since $f \in C^{n+1}(D)$ and $\mathcal{B}_r[x_0]$ is compact, we can find $M \geq 0$ such that

$$|D^\alpha f(y)| \leq M \qquad \text{for all } y \in \mathcal{B}_r[x_0]$$

and all multiindex α with $|\alpha| = n + 1$. Then if $\|\xi\| \leq r$, we have

$$\frac{|R_n(\xi)|}{\|\xi\|^n} \leq \sum_{|\alpha|=n+1} \frac{|D^\alpha f(x_0 + \theta\xi)|}{\alpha!} \frac{|\xi^\alpha|}{\|\xi\|^n} \leq \sum_{|\alpha|=n+1} \frac{M}{\alpha!} \frac{\|\xi\|^{n+1}}{\|\xi\|} = \sum_{|\alpha|=n+1} \frac{M}{\alpha!} \|\xi\|$$

□

Example 12.4

et $f(x,y) = \cos(x+2y)$ defined on $\mathbb{R}^2$, find $T_{2,(0,0)}(\xi)$
We have $f(0,0) = 1$, also

$$\begin{aligned} f_x(x,y) &= -\sin(x+2y) & f_y(x,y) &= -2\sin(x+2y) \\ f_{xx}(x,y) &= -\cos(x+2y) & f_{yy}(x,y) &= -4\cos(x+2y) \\ f_{xy}(x,y) &= -2\cos(x+2y) \end{aligned}$$

Then

$$\begin{aligned} T_{2,(0,0)}(\xi_1,\xi_2) &= f(0,0) + f_x(0,0)\xi_1 + f_y(0,0)\xi_2 + \frac{f_{xx}(0,0)}{2}\xi_1^2 + \frac{f_{yy}(0,0)}{2}\xi_2^2 + f_{xy}(0,0)\xi_1\xi_2 \\ &= 1 - \frac{\xi_1^2}{2} - \frac{4\xi_2^2}{2} - 2\xi_1\xi_2 \\ &= 1 - \frac{1}{2}\left(\xi_1^2 + 4\xi_2^2 - 4\xi_1\xi_2\right) \end{aligned}$$

12.4 The Hessian

Definition 12.4: Hessian

Let $\varnothing \neq D \subseteq \mathbb{R}^N$ be open, $f : D \to \mathbb{R}$, $f \in C^2(D)$. The **Hessian of** f **at** $x \in D$ denoted by $(\text{Hess } f)(x)$, is $N \times N$ matrix whose i,j-entry is $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$, that is

$$(\text{Hess } f)(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_N} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \ddots & & \\ \vdots & & \ddots & \\ \frac{\partial^2 f(x)}{\partial x_N \partial x_1} & & & \frac{\partial^2 f(x)}{\partial x_N^2} \end{bmatrix}$$

Notice that $(\text{Hess } f)(x)$ is symmetric by Clairaut's Theorem (10.1).

Corollary 12.2

Let $f \in C^2(D)$, $D \subset \mathbb{R}^N$ be open. Let $x_0 \in D$ and $\xi \in \mathbb{R}^N$ be such that $x_0 + t\xi \in D$ for all $t \in [0,1]$, then there exists $\theta \in (0,1)$ such that

$$f(x_0 + t\xi) = f(x_0) + \nabla f(x_0) \cdot \xi + \frac{1}{2}\left[((\text{Hess } f)(x_0 + \theta\xi)\xi) \cdot \xi\right]$$

Proof. STP that for all $x \in D$ we have

$$\sum_{|\alpha|=2} \frac{(D^\alpha f)(x)}{\alpha!}\xi^\alpha = \frac{1}{2}\left[((\text{Hess } f)(x)\xi) \cdot \xi\right]$$

We compute,

$$\sum_{|\alpha|=2} \frac{(D^\alpha f)(x)}{\alpha!}\xi^\alpha = \sum_{i=1}^{N} \frac{f_{x_i x_j}(x_0)\xi_1^2}{2} + \sum_{i<j} f_{x_i x_j}(x)\xi_i\xi_j$$
$$= \frac{1}{2}\left(\sum_{i=1}^{N} f_{x_i x_j}(x)\xi_i^2 + \sum_{i\neq j} f_{x_i x_j}(x)\xi_i\xi_j\right)$$
$$= \frac{1}{2}\left[((\text{Hess } f)(x)\xi)\cdot\xi\right]$$

as desired. □

12.5 Critiacal Points

Definition 12.5: Stationary Point (Critiacal Point)

Let $f \in C^1(D)$, $f : D \to \mathbb{R}$,

1. we say that $x_0 \in D$ is a **stationary point** of f (or a **critical point** of f) if $\nabla f(x_0) = 0$.
2. x_0 is a local maximum if there exists $\delta > 0$ such that $f(x) \leq f(x_0)$ for all $x \in \mathcal{B}_\delta(x_0) \cap D$.
3. x_0 is a local minimum if there exists $\delta > 0$ such that $f(x) \geq f(x_0)$ for all $x \in \mathcal{B}_\delta(x_0) \cap D$.

Discovery 12.2

If x_0 is a local maximum (or a local minimum) of f, then x_0 is a critical point. This is becasue if $g(t) = f(x_0 + te_i)$ where $1 \leq i \leq N$, then 0 is a local maximum (or local minimum) of g and so

$$0 = g'(0) = \frac{\partial f(x_0)}{\partial x_i} \quad \Rightarrow \quad \nabla f(x_0) = 0$$

Example 12.5

Let $f(x, y) = x^2 - y^2$ defined on $\mathbb{R}^2$, then

$$\nabla f(x, y) = (2x, -2y)$$

hence $(0, 0)$ is a critial point of f, but it is neither a local maximum nor a local minimum.

Definition 12.6: Saddle Point

A critial point of f that is neither a local maximum nor a local minimum is called a **saddle point**.

In order to clarify stationary point we need more linear algebra.

Definition 12.7

Let $A \in \mathcal{M}_n(\mathbb{R})$ be a symmetric matrix, we say

1. A is **positive definite** if $(A\xi)\cdot\xi > 0$ for all $0 \neq \xi \in \mathbb{R}^N$;
2. A is **positive semidefinite** if $(A\xi)\cdot\xi \geq 0$ for all $\xi \in \mathbb{R}^N$;
3. A is **negative definite** if $(A\xi)\cdot\xi < 0$ for all $0 \neq \xi \in \mathbb{R}^N$;
4. A is **negative semidefinite** if $(A\xi)\cdot\xi \leq 0$ for all $\xi \in \mathbb{R}^N$;
5. A is **indefinite** if there are $x, y \in \mathbb{R}^N$ with $(Ax)\cdot x > 0$ and $(Ay)\cdot y < 0$.

Example 12.6

For an instance, $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ is positive definite, I is positive definite, and $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ is indefinite.

Lecture 25 - Wednesday, Jul 3

In order to prove the Theorem (12.3), we first prove the following Lemma:

Lemma 12.2

Suppose $f \in C^2(D)$, and $x_0 \in D$ be such that $(\text{Hess } f)(x_0)$ is positive definite (or negative definite). Then there exists $\delta > 0$ such that for $x \in D$ and $x \in \mathcal{B}_\delta(x_0)$, then $(\text{Hess } f)(x)$ is positive definite (or negative definite).

Proof. We will prove the statement for $(\text{Hess } f)(x_0)$ positive definite. Write $A_x = (\text{Hess } f)(x_0)$. Define $Q : \mathbb{R}^N \to \mathbb{R}$, $Q(\xi) = (A_{x_0}\xi)\cdot\xi$. Then Q is continuous because it is the dot product of continuous functions on $\mathbb{R}^N$. For all unit vectors $\xi \in S^{N-1} = \partial\mathcal{B}_1(0)$, we have $Q(\xi) > 0$. Since S^{N-1} is compact, by the Extreme Value Theorem, there exists $r > 0$ such that $Q(\xi) \geq r$ for all $\xi \in S^{N-1}$. Since $f \in C^2(D)$, we can find $\delta > 0$ such that $\mathcal{B}_\delta(x_0) \subset D$ and

$$\sum_{i=1}^{N} |f_{x_i x_i}(x) - f_{x_i x_i}(x_0)| + \sum_{i \neq j} \left|f_{x_i x_j}(x) - f_{x_i x_j}(x_0)\right| < \frac{r}{2}$$

Then if $x \in \mathcal{B}_\delta(x_0)$, we have for $\xi \in S^{N-1}$

$$\begin{aligned} |(A_x\xi)\cdot\xi - (A_{x_0}\xi)\cdot\xi| &= \left|\sum_{i=1}^{N}(f_{x_i x_i}(x) - f_{x_i x_i}(x_0))\xi_i^2 + \sum_{i \neq j}(f_{x_i x_j}(x) - f_{x_i x_j}(x_0))\xi_i\xi_j\right| \\ &\leq \sum_{i=1}^{N} |f_{x_i x_i}(x) - f_{x_i x_i}(x_0)| + \sum_{i \neq j} \left|f_{x_i x_j}(x) - f_{x_i x_j}(x_0)\right| < \frac{r}{2} \end{aligned}$$

This implies that for $\xi \in S^{N-1}$:

$$(A_x\xi)\cdot\xi > (A_{x_0}\xi)\cdot\xi - \frac{r}{2} \geq r - \frac{r}{2} = \frac{r}{2} > 0$$

so $x \in \mathcal{B}_\delta(x_0)$, and $\xi \in \mathbb{R}^N\backslash\{0\}$ and we get

$$(A_x\xi)\cdot\xi = \|\xi\|^2\left(A_x\left(\frac{\xi}{\|\xi\|}\right)\cdot\frac{\xi}{\|\xi\|}\right) > 0$$

Hence A_x is positive definite for all $x \in \mathcal{B}_\delta(x_0)$. □

Theorem 12.3: Second Derivative Test

Let $\varnothing \neq D \subset \mathbb{R}^N$ be open and $f : D \to \mathbb{R}$, $f \in C^2(D)$. Let $x_0 \in D$ be a critical point of f, then

1. If $(\text{Hess } f)(x_0)$ is positive definite, then f has an local minimum at x_0;
2. If $(\text{Hess } f)(x_0)$ is negative definite, then f has an local maximum at x_0;
3. If $(\text{Hess } f)(x_0)$ is indefinite, then f has an saddle point at x_0;

Discovery 12.3

For an example where the above Theorem (12.3) does not apple, see A4.

Proof. 1. Suppose $(\text{Hess } f)(x_0)$ is positive definite. Let $\delta > 0$ be such that $(\text{Hess } f)(\gamma)$ is positive definite for all $\gamma \in \mathcal{B}_\delta(x_0) \subset D$. Take $x \in \mathcal{B}_\delta(x_0)$. Write $\xi := x - x_0$, so that $\|\xi\| < \delta$. By Taylor's Theorem (12.2), there exists $\theta \in (0,1)$ such that

$$\begin{aligned} f(x_0+\xi) &= f(x_0) + \nabla f(x_0)\cdot\xi + \frac{1}{2}(\text{Hess } f)(x_0+\theta\xi)\cdot\xi \\ &= f(x_0) + \frac{1}{2}\left[(\text{Hess } f)(x_0+\theta\xi)\cdot\xi\right] \end{aligned}$$

Then

$$f(x) - f(x_0) = f(x_0+\theta\xi) - f(x_0) = \frac{1}{2}(\text{Hess } f(x_0+\theta\xi)\xi)\cdot\xi > 0$$

Hence x_0 is a local minimum for f;

2. Follows as in (1);

3. Suppose $(\text{Hess } f)(x_0)$ is indefinite, we want to show that given $\varepsilon > 0$, there are $x, y \in \mathcal{B}_\varepsilon(x_0) \cap D$ such that

$$f(x) < f(x_0) < f(y)$$

Let ξ_1, ξ_2 be unit vectors in $\mathbb{R}^N$ such that

$$(\text{Hess } f)(x_0)\xi_1\cdot\xi_1 < 0 \quad \text{and} \quad (\text{Hess } f)(x_0)\xi_2\cdot\xi_2 > 0$$

Arguing as in the proof of Lemma (12.2), we can find $\delta > 0$ such that $\mathcal{B}_\delta(x_0) \subset D$ and if $x \in \mathcal{B}_\delta(x_0)$,

$$(\text{Hess } f)(x)\xi_1 \cdot \xi_1 < 0 \quad \text{and} \quad (\text{Hess } f)(x)\xi_2 \cdot \xi_2 > 0$$

Then given $\varepsilon > 0$, set $\varepsilon' = \min\{\delta, \varepsilon\}$ and let $\xi_{\varepsilon'} := \frac{\varepsilon'}{2}\xi_1$ and $\eta_{\varepsilon'} := \frac{\varepsilon'}{2}\xi_2$. So $x_0 + \xi_{\varepsilon'}, x_0 + \eta_{\varepsilon'} \in \mathcal{B}_\delta(x_0)$. By Taylor's Theorem (12.2), there are $\theta_1, \theta_2 \in (0,1)$ such that

$$f(x_0 + \xi_{\varepsilon'}) = f(x_0) + \left(\frac{\varepsilon'}{2}\right) \cdot \frac{1}{2}(\text{Hess } f)(x_0 + \xi_{\varepsilon'})\xi_1 \cdot \xi_1$$
$$f(x_0 + \eta_{\varepsilon'}) = f(x_0) + \left(\frac{\varepsilon'}{2}\right) \cdot \frac{1}{2}(\text{Hess } f)(x_0 + \eta_{\varepsilon'})\xi_2 \cdot \xi_2$$

Setting $x = x_0 + \xi_{\varepsilon'}$ and $y = x_0 + \eta_{\varepsilon'}$ we see that $x, y \in \mathcal{B}_\varepsilon(x_0)$ and by (1), $f(x) < f(x_0) < f(y)$.

□

Theorem 12.4

Let $A = (\alpha_{ij})_{i,j} \in \mathcal{M}_n(\mathbb{R})$ be symmetric. TFAE:

1. A is positive definite (or negative definite);
2. All eigenvalues of A are positive (or negative);
3. $\det \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1k} \\ \alpha_{21} & \ddots & & \\ \vdots & & \ddots & \\ \alpha_{k1} & & & \alpha_{kk} \end{bmatrix} > 0 \left(\text{or } (-1)^k \det \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1k} \\ \alpha_{21} & \ddots & & \\ \vdots & & \ddots & \\ \alpha_{k1} & & & \alpha_{kk} \end{bmatrix} > 0 \right)$ for all $k = 1, \ldots, N$.

Corollary 12.3: Second Derivative Test in $\mathbb{R}^2$

Let $\varnothing \neq D \subset \mathbb{R}^2$ be open, $f : D \to \mathbb{R}$, $f \in C^2(D)$. Let $x_0 \in D$ be a critical point of f, then

1. If $f_{xx}(x_0) > 0$ and $f_{xx}(x_0)f_{yy}(x_0) - f_{xy}(x_0)^2 > 0$, then x_0 is a local minimum of f;
2. If $f_{xx}(x_0) < 0$ and $f_{xx}(x_0)f_{yy}(x_0) - f_{xy}(x_0)^2 > 0$, then x_0 is a local maximum of f;
3. If $f_{xx}(x_0)f_{xx}(x_0) - f_{xy}(x_0)^2 < 0$, then x_0 is a saddle point of f;

Proof. (1) and (2) are clear. For (3), let λ_1, λ_2 be the eigenvalues of $(\text{Hess } f)(x_0)$, then

$$f_{xx}(x_0)f_{xx}(x_0) - f_{xy}(x_0)^2 = \det((\text{Hess } f)(x_0)) = \lambda_1\lambda_2 \quad \Rightarrow \quad \lambda_1\lambda_2 < 0$$

So λ_1 and λ_2 have opposite signs. If ξ_1, ξ_2 are eigenvectors, we have $(\text{Hess } f)(x_0)\xi_1 \cdot \xi_1$ and $(\text{Hess } f)(x_0)\xi_2 \cdot \xi_2$ have opposite signs. Hence $(\text{Hess } f)(x_0)$ is indefinite. □

Lecture 26 - Friday, Jul 5

Example 12.7

Let $K = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ and let $f : K \to \mathbb{R}$, $f(x, y) = x^2 - xy + y^2$. Find the global maximum and minimum of f on K.

Proof. Since K is compact and f is continuous, we know from the Extreme Value Theorem that the problem has a solution. Let $D = K^\circ = \mathcal{B}_1((0, 0))$. We have $f_x = 2x - y$ and $f_y = 2y - x$. Then $(0, 0)$ is the only critial point of f in D. We have $f_{xx} = 2$, $f_{yy} = 2$, and $f_{xy} = -1$, so

$$(\text{Hess } f)(0, 0) = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Then $f_{xx} > 0$, and $f_{xx}f_{yy} - f_{xy}^2 > 0$, thus $(\text{Hess } f)(0, 0)$ is positive definite. By second derivative test, f has local minimum at $(0, 0)$. Now we want to verify

$$\partial K = \{(x, y) : x^2 + y^2 = 1\} = \{(\cos\theta, \sin\theta) : 0 \leq \theta \leq 2\pi\}$$

Consider $g(0) = f(\cos\theta, \sin\theta) = \cos^2\theta - \cos\theta\sin\theta + \sin^2\theta = 1 - \cos\theta\sin\theta = 1 - \frac{\sin(2\theta)}{2}$, we have $g(0) \geq \frac{1}{2}$. Hence f attains its minimum on K at $(0, 0)$ since $f(0, 0) = 0$. We have $g'(0) = -\cos(2\theta)$. Thus the crital points of g in $(0, 2\pi)$ are $\theta_1 = \frac{\pi}{4}$, $\theta_2 = \frac{3\pi}{4}$, $\theta_3 = \frac{5\pi}{4}$, and $\theta_4 = \frac{7\pi}{4}$. Now $g''(0) = 2\sin(2\theta)$ gives that

$$g''(\theta_1) = 2 = g''(\theta_3) \qquad \text{and} \qquad g''(\theta_2) = -2 = g''(\theta_4)$$

Also $g(0) = 1 = g(2\pi)$, so θ_2 and θ_4 are local maximum of g. Compute $g(\theta_2) = \frac{3}{2} = g(\theta_4)$. It follow that f attains its maximum at $(\cos(\theta_2), \sin(\theta_2)) = \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ and at $(\cos(\theta_4), \sin(\theta_4)) = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$ □

13 Local Properties of Continuously differentiable function

13.1 Inverse Function Theorem

Roughly, the IFT states that if $D \subset \mathbb{R}^N$, $f : D \to \mathbb{R}^N$, $f \in C^1(D, \mathbb{R}^N)$ and $(Df)(x_0)$ is invertible, then there exists an open neighborhood U of x_0 such that f is one-to-one on U, and $f^{-1} : f(U) \to \mathbb{R}^N$ is also continuously differentiable.

Definition 13.1: Contraction

Let $\varnothing \neq S \subset \mathbb{R}^N$ and $\varphi : S \to S$, we say that φ is a **contraction** if there exists $0 \leq c < 1$ such that

$$\|\varphi(x) - \varphi(y)\| \leq c \|x - y\| \qquad \forall\, x, y \in S$$

Theorem 13.1: Contradiction Mapping Principle

Let $\varnothing \neq F \subset \mathbb{R}^N$ be closed and $\varphi : F \to F$ be contraction. Then there exists a unique $x_* \in F$ such that $\varphi(x_*) = x_*$ (i.e. f has a unique fixed point $x_* \in F$).

Proof. For uniqueness, suppose x_*, y_* are fixed point of φ, then

$$\|x_* - y_*\| = \|\varphi(x_*) - \varphi(y_*)\| \leq c \|x_* - y_*\| < \|x_* - y_*\|$$

Hence we must have $x_* = y_*$. For existence of x_*, take $x_0 \in F$, define an sequence (x_n) in F recursively by setting $x_n = \varphi(x_{n-1})$ for $n \geq 1$, so we have for $n = 1$,

$$\begin{aligned} \|x_{n+1} - x_n\| = \|x_2 - x_1\| &= \|\varphi(x_1) - \varphi(x_0)\| \leq c \|x_1 - x_0\| \\ \|x_3 - x_2\| &= \|\varphi(x_2) - \varphi(x_1)\| \leq c \|x_2 - x_1\| \leq c^2 \|x_1 - x_0\| \end{aligned}$$

Continuing with this process by induction we obtain for all $n \geq 1$,

$$\|x_{n+1} - x_n\| = \|\varphi(x_n) - \varphi(x_{n-1})\| \leq c^n \|x_1 - x_0\|$$

Then if $m > n \geq 1$, we have

$$\|x_m - x_n\| = \left\| \sum_{k=n}^{m-1} (x_{k+1} - x_k) \right\| \leq \sum_{k=n}^{m-1} \|(x_{k+1} - x_k)\| \leq \sum_{k=n}^{m-1} c^k \|x_1 - x_0\|$$

Since the sum $\sum_{k=1}^{\infty} c^k \|x_1 - x_0\|$ converges because $0 \leq c < 1$, we deduce that (x_n) is a Cauchy Sequence. We let $x_* := \lim_{n\to\infty} x_n$, then $x_* \in F$ because F is closed. Since φ is continuous, we get

$$\varphi(x_*) = \lim_{n\to\infty} \varphi(x_n) = \lim_{n\to\infty} x_{n+1} = x_*$$

proving that x_* is a fixed point of φ. $\square$

Theorem 13.2

Let $\varnothing \neq D \subset \mathbb{R}^N$ be an open convex set. Let $f : D \to \mathbb{R}^M$ be differentiable and suppose there exists $R \in \mathbb{R}$ such that $\|Df(x)\| \leq R$ for all $x \in D$. Then for all $x, y \in D$, we have

$$\|f(x) - f(y)\| \leq R \|x - y\|$$

Proof. Fix $x, y \in D$, $x \neq y$ and consider $g : D \to \mathbb{R}$, $g(z) = (f(x) - f(y)) \cdot f(z)$. Then g is differentiable and $\nabla g(z) = (f(x) - f(y))^T (Df)(z)$ by Product Rule (10.2). By A3-Q3, there exists ξ in the line segment between x, y such that

$$g(x) - g(y) = \nabla g(\xi) \cdot (x - y)$$

Thus

$$\begin{aligned} & \|f(x) - f(y)\|^2 = (f(x) - f(y))^T (Df)(\xi)(x - y) \\ \Rightarrow \quad & \|f(x) - f(y)\|^2 \leq \|f(x) - f(y)\| \, R \, \|x - y\| \end{aligned}$$

giving us $\|f(x) - f(y)\| \leq R \|x - y\|$. □

Lecture 27 - Monday, Jul 8

Theorem 13.3: The Inverse Function Theorem

Let $\varnothing \neq D \subset \mathbb{R}^N$ be open and $f \in C^1(D, \mathbb{R}^N)$. Let $x_0 \in D$ be such that $(Df)(x_0)$ is invertible and set $y_0 := f(x_0)$, then

1. There exists an open set $U \subseteq D$, $V \subset \mathbb{R}^N$ with $x_0 \in U$, $y_0 \in V$, f is one-to-one on U and $V := f(U)$;
2. If $g : V \to \mathbb{R}^N$ is the inverse of f defined on V (i.e. $g(f(x)) = x$ for $x \in U$), then g is continuously differentiable and
$$(Dg)(y) = [(Df)(g(y))]^{-1}$$

Discovery 13.1

$(Df)(x_0)$ is invertible if and only if $\det(J_f(x_0)) \neq 0$.

If we write $f(x_1, \ldots, x_N) = (f_1(x_1, \ldots, x_N), \ldots, f_N(x_1, \ldots, x_N))$,

$$\begin{aligned} y_1 &= f_1(x_1, \ldots, x_N) \\ &\vdots \\ y_N &= f_N(x_1, \ldots, x_N) \end{aligned}$$

Result 13.1

Then the IFT (Inverse Function Theorem 13.3) tells us that the system given above can be solved for $x_1, \dots, x_N$ in terms of $y_1, \dots, y_N$ when we restrict to a small neighborhood of x_0 and y_0, and the solution is continuously differentiable.

Example 13.1

Let $u = \frac{x^4+y^4}{x}$ and $v = \sin x + \cos y$. Can we sovle the system above for x and y in terms of u and v? We have

$$J_f(x,y) = \begin{bmatrix} \frac{3x^4-y^4}{x^2} & \frac{4y^3}{x} \\ \cos x & -\sin y \end{bmatrix}$$

$$\Rightarrow \quad \det(J_f(x,y)) = -\sin y \left(\frac{3x^4 - y^4}{x^2} \right) - \cos x \cdot \frac{4y^3}{x}$$

If, for example, $x_0 = (x,y) = \left(\frac{\pi}{2}, \frac{\pi}{2}\right)$, then

$$\det(J_f(x_0)) = -\left[3\left(\frac{\pi}{2}\right)^2 - \left(\frac{\pi}{2}\right)^2 \right] = -2\left(\frac{\pi}{2}\right)^2 \neq 0$$

Hence the IFT (13.3) says that near x_0 we can solve the system for x and y in terms of u and v.

Proof. This is the proof for IFT (13.3)
The formula for $(Dg)(y)$ follows from $Q5c$ in the Midterm Exam.

1. For part 1:
   Set $A = (Df)(x_0)$. Let U be an open ball such that

   $$\|(Df)(x) - A\| < \lambda \qquad \text{where } \lambda = \frac{1}{2\,\|A^{-1}\|}$$

   This exists becasue f is continuously differentiable. We can also find that $(Df)(x)$ is invertible for all $x \in U$ (See A4Q5). For $y \in \mathbb{R}^N$ fixed, define $\varphi_y : D \to \mathbb{R}^N$ by

   $$\varphi_y(x) = x + A^{-1}(y - f(x))$$

   (a) *Claim 1*: $y = f(x)$ if and only if x is a fixed point of φ_y
   Indeed, $y = f(x)$ gives $\varphi_y(x) = x$ since $A^{-1}(y - f(x)) = 0$. Conversely, if $\varphi_y(x) = x$, then $A^{-1}(y - f(x)) = 0$, which implies that $y - f(x) = 0$ because A^{-1} is one-to-one.

   (b) *Claim 2*: $\|\varphi_y(x) - \varphi_y(z)\| \leq \frac{1}{2}\|x - z\|$ for all $x, z \in U$
   Notice that $\varphi_y(x) = Ix + A^{-1}y - A^{-1}f(x)$, so by the Chain Rule (10.3), φ_y is differentiable and

   $$(D\varphi_y)(x) = I - A^{-1}(Df)(x)$$

Then

$$\|(D\varphi_y)(x)\| = \left\|A^{-1}A - A^{-1}(Df)(x)\right\| = \left\|A^{-1}(A - (Df)(x))\right\|$$
$$\leq \left\|A^{-1}\right\| \|A - (Df)(x)\|$$
$$< \left\|A^{-1}\right\| \frac{1}{2\left\|A^{-1}\right\|} = \frac{1}{2}$$

Hence by Theorem (13.2), we have $\|\varphi_y(x) - \varphi_y(z)\| \leq \frac{1}{2}\|x - z\|$ for all $x, z \in U$.

This shows that φ_y has at most one fixed point in U, so f is one-to-one in U by Claim 1. Set $V = f(U)$, we will show that V is open. Let $w \in V$ and let $z \in U$ be such that $w = f(z)$. Let $r > 0$ be such that $\mathcal{B}_z = \mathcal{B}_r(z) \subset U$, we will find $\delta > 0$ such that if $\|y - w\| < \delta$, then $\varphi_y(\mathcal{B}_z) \subset \mathcal{B}_z$. First, notice that if $x \in \mathcal{B}_z$, then by Claim 2,

$$\|\varphi_y(x) - \varphi_y(z)\| \leq \frac{1}{2}\|x - z\| = \frac{r}{2}$$

Let $\delta := \lambda r$, and let $y \in \mathbb{R}^N$, $\|y - w\| < \delta$, then

$$\|\varphi_y(z) - z\| = \left\|z + A^{-1}(y - f(z)) - z\right\| = \left\|A^{-1}(y - w)\right\| \leq \left\|A^{-1}\right\| \|y - w\| < \left\|A^{-1}\right\| \cdot \frac{r}{2\left\|A^{-1}\right\|} = \frac{r}{2}$$

Then if $\|y - w\| < \delta$, and $x \in \mathcal{B}_z$, we have

$$\|\varphi_y(x) - z\| \leq \|\varphi_y(x) - \varphi_y(z)\| + \|\varphi_y(z) - z\|$$
$$\leq \frac{r}{2} + \frac{r}{2} = r$$

giving that $\varphi_y(\mathcal{B}_z) \subset \mathcal{B}_z$. By the Contraction Mapping Principle, φ_y has a unique fixed point $x_* \in \mathcal{B}_z$, so $y = f(x_*) \in f(U) = V$ by Claim 1. This shows that $f(U)$ is open.

2. For part 2:
Let $g : V \to \mathbb{R}^N$ be the inverse of f on U. Let $y \in V$, $y + k \in V$, and let $x, x + h \in U$ be such that

$$f(x) = y, \quad f(x + h) = y + k$$

Notice that h is uniquely determined by k.

Lecture 28 - Wednesday, Jul 10

Notice that

$$\varphi_y(x + h) - \varphi_y(x) = h + A^{-1}(y - f(x + h))$$
$$= h - A^{-1}k$$

Thus by Claim 2,

$$\left\|h - A^{-1}k\right\| \leq \frac{1}{2}\|x + h - x\| = \frac{\|h\|}{2} \quad \Rightarrow \quad \left\|A^{-1}k\right\| - \|h\| \leq \frac{\|h\|}{2}$$

giving that $\left\|A^{-1}k\right\| \geq \frac{\|h\|}{2}$. Hence

$$\|h\| \leq \left\|A^{-1}\right\| \cdot 2 \cdot \|k\| = \lambda^{-1}\|k\|$$

Let $T = [(Df)(x)]^{-1}$, then

$$\begin{aligned} g(y+k) - g(y) - Tk &= h - Tk \\ &= TT^{-1}h - Tk \\ &= T\left((Df)(x)h - (f(x+h) - f(x))\right) \end{aligned}$$

Now we have

$$\frac{\|g(y+k) - g(y) - Tk\|}{\|k\|} \leq \frac{\|T\|\,\|f(x+h) - f(x) - (Df)(x)h\|}{\lambda\,\|h\|}$$

Taking the limit of k approaches 0, then h approaches 0, and it follows that

$$\lim_{k \to 0} \frac{\|g(y+k) - g(y) - Tk\|}{\|k\|} = 0$$

proving that g is differentiable at y. Finally, we will show that $g \in C^1(V, \mathbb{R}^N)$, that is, $y \in V \mapsto J_g(y)$ is continuous. This follows because the map is the composition

$$V \longrightarrow^{g} U \longrightarrow^{J_f} \mathrm{GL}_N(\mathbb{R}) \longrightarrow^{-1} \mathrm{GL}_N(\mathbb{R})$$

All the maps are continuous (See A4Q5), hence $g \in C^1(V, \mathbb{R}^N)$.

□

Theorem 13.4: Open Mapping Theorem

Let $\varnothing \neq D \subset \mathbb{R}^N$ be open, $f \in C^1(D, \mathbb{R}^N)$. Suppose that $(Df)(x)$ is invertible for all $x \in D$, then for every $W \subset D$ open, $f(W) \subset \mathbb{R}^N$ is also open.

Proof. Exercise. □

13.2 Implicit Function Theorem

Definition 13.2: Level Curves

Let f be a function defined on $\mathbb{R}^2$, we write $z = f(x,y)$. The **level curve** of f determined by $c \in \mathbb{R}$ in the set of all points in $\mathbb{R}^2$ such that $f(x,y) = c$.

We wish to locally express the set of points $f(x,y) = 0$ as the graph of a function $y = g(x)$.

Example 13.2

$f(x,y) = x^2 - y$, so $f(x,y) = 0$ given $y = x^2$. Take $g(x) = x^2$.

Example 13.3

$f(x,y)=x^2+y^2-1$; Near $(1,0)$, we cannot express the set $f(x,y)=0$ as the graph of a function of $y=g(x)$.

Definition 13.3

We will write $(x,y)\in\mathbb{R}^{N+M}$ as

$$(x,y)=(x_1,\dots,x_N,y_1,\dots,y_M)$$

given a system of equations

$$\begin{aligned} f_1(x_1,\dots,x_N,y_1,\dots,y_M)&=0,\\ &\vdots\\ f_q(x_1,\dots,x_N,y_1,\dots,y_M)&=0 \end{aligned}$$

we want to locally express y in terms of x, so that $y_1=g_1(x_1,\dots,x_N),\dots,y_M=g_M(x_1,\dots,x_N)$.

13.2.1 The Linear Case

Suppose $f(x,y) = A\begin{bmatrix} x \\ y \end{bmatrix}$, $A \in \mathcal{M}_{M\times(N+M)}(\mathbb{R})$. In the case

$$A = \begin{bmatrix} A_x & A_y \end{bmatrix} \quad A_x \in \mathcal{M}_{M\times N}(\mathbb{R}),\ A_y \in \mathcal{M}_{M\times M}(\mathbb{R})$$

we have $f(x,y) = 0$ gives $A_x x + A_y y = 0$. From linear algebra we know that if A_y is invertible, then the equation $A_x x + A_y y = 0$ uniquely determines y in terms of x by

$$y = -A_y^{-1} A_x x$$

In general, given a linear transformation $A : \mathbb{R}^{N+M} \to \mathbb{R}^M$, we can split A into two linear transformations $A_x : \mathbb{R}^N \to \mathbb{R}^M$ and $A_y : \mathbb{R}^M \to \mathbb{R}^M$, where $A_x(x) = A(x,0)$ and $A_y(y) = A(0,y)$, so that

$$A(x,y) = A_x(x) + A_y(y)$$

If f is differentiable, $A = J_f(x_0)$, write $A_x = \frac{\partial f}{\partial x}$, $A_y = \frac{\partial f}{\partial y}$.

Theorem 13.5: Implicit Function Theorem

Let $\varnothing \neq D \subset \mathbb{R}^{N+M}$ be open and $f \in C^1(D, \mathbb{R}^M)$. Let $(x_0, y_0) \in \mathbb{R}^{N+M}$ be such that $f(x_0, y_0) = 0$ and let $A = (Df)(x_0, y_0)$. Suppose that A_y is invertible, i.e.

$$\det \begin{bmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_M} \\ \vdots & & \\ \frac{\partial f_M}{\partial y_1} & \cdots & \frac{\partial f_M}{\partial y_M} \end{bmatrix} \neq 0 \qquad \text{at } (x_0, y_0).$$

Then there exists an open neighbourhood $U \subset D$ of (x_0, y_0) and $W \subset \mathbb{R}^N$, open neighbourhood of x_0, such that

1. For every $x \in W$, there exists a unique y_x such that $(x, y_x) \in U$ such that $f(x, y_x) = 0$.
2. If we define $g : W \to \mathbb{R}^M$, $g(x) = y$, where y is as in part (a), then g is continuously differentiable ($g \in C^1(W, \mathbb{R}^M)$), $(x,y) \in U$ and $f(x,y) = 0$, $\forall x \in W$, and

$$(Dg)(x_0) = -A_y^{-1} A_x$$

Discovery 13.2

The function g is implicitly defined by the equation $f(x,y) = 0$.

Proof. Define $F := D \to \mathbb{R}^{N+M}$ by $F(x,y) = (x, f(x,y))$. Then F is continuously differentiable because f is. Our claim is that $(DF)(x_0, y_0)$ is invertible. Indeed, we have

$$J_F(x_0,y_0) = \begin{bmatrix} I_N & 0_{N\times M} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}$$

Then because A_y is invertible,

$$\det J_F(x_0,y_0) = \det I_N \cdot \det \frac{\partial f}{\partial y} \neq 0$$

Then the Inverse Function Theorem (13.3) gives us an open neighborhood $U \subset D$ of (x_0, y_0) such that $V := F(U)$ is open, F is one-to-one on U, and $G : V \to U \subset \mathbb{R}^{N+M}$ is also continuously differentiable.
We define $W \subset \mathbb{R}^N$ by $W := \{w \in \mathbb{R}^N : (x, 0) \in V\}$, then $x_0 \in W$ because (x_0, y_0) is in U and $F(x_0, y_0) = (x_0, 0_M)$. Also, W is open because V is open. If $x \in W$, then because $V = F(U)$, there exists $(x', y') \in U$ such that $F(x', y') = (x', f(x', y')) = (x, 0)$, which shows that $x' = x$ and $f(x, y') = 0$.
Now we wish to show uniqueness. Suppose $y_1, y_2 \in \mathbb{R}^M$ are such that $(x, y_1), (x, y_2) \in U$ and $f(x, y_1) = f(x, y_2) = 0$. It follows that $F(x, y_1) = (x, 0_M) = F(x, y_2)$. Because F is one-to-one on U, thus we must have $y_1 = y_2$, proving part (a).
For part (b), let $g : W \to \mathbb{R}^M$, $g(x) = y$. Consider $G(x, 0) = (x, g(x))$, since $G \in C^1(V, \mathbb{R}^{N+M})$ (is continuous differentiable), we must have that $g \in C^1(W, \mathbb{R}^M)$. Then we compute $(Dg)(x_0)$. Consider $\phi : W \to \mathbb{R}^{N+M}$, $\phi(x) = (x, g(x))$, then $\phi \in C^1(W, \mathbb{R}^N + M)$, $\phi(x_0) = (x_0, y_0)$. Also, for all $x \in W$ and $h \in \mathbb{R}^N$

$$(D\phi)(x)h = (h, Dg(x)h)$$

In terms of the Jacobian Matrix of ϕ at x,

$$J_\phi(x) = \begin{bmatrix} I_N \\ J_g(x) \end{bmatrix}$$

Now $f(\phi(x)) = 0$ for all $x \in W$. Applying the Chain Rule (10.3) we get

$$(Df)(\phi(x))(D\phi)(x) = 0 \qquad \forall x \in W$$

Thus for $x = x_0$ and $h \in \mathbb{R}^N$, $(Df)(x_0, y_0)(D\phi)(x_0) = 0$, add

$$\begin{aligned} (Df)(x_0,y_0)(D\phi)(x_0)h &= 0 \\ (Df)(x_0,y_0)(h, (Dg)(x_0)h) &= 0 \\ \Rightarrow \quad A_x h + A_y (Dg)(x_0)h &= 0 \end{aligned}$$

Since $A = (Df)(x_0, y_0)$, this yields

$$(Dg)(x_0)h = -A_y^{-1} A_x h$$

because A_y is invertible. Hence $(Dg)(x_0) = -A_y^{-1} A_x$ as needed. □

Discovery 13.3

Above we only needed A_y invertible to obtain $(Dg)(x_0) = -A_y^{-1}A_x$. Since the set of invertible linear tranformations is open, we can assume that $\frac{\partial f}{\partial y}$ is invertible for all $(x, y) \in U$ and hence

$$(Dg)(x) = -\left(\frac{\partial f}{\partial y}\right)^{-1}\frac{\partial f}{\partial x} \qquad \forall x \in W$$

Example 13.4

Consider the system of equations,

$$\begin{cases} 2e^{y_1} + y_2x_1 - 4x_2 + 3 & = 0 \\ y_2 \cos y_1 - 6y_1 + 2x_1 - x_3 & = 0 \end{cases}$$

where there are five variables and two equations:

$$N + M = 5, \qquad M = 2$$

It is easy to check that $(3, 2, 7, 0, 1)$ is a solution. Can we solve the solution near $(3, 2, 7, 0, 1)$ by $(x, g(x))$ where $g : W \to \mathbb{R}^2$, $W \subset \mathbb{R}^3$.
Let $f : \mathbb{R}^3 \to \mathbb{R}^2$, $f(x_1, x_2, x_3, y_1, y_2) = (f_1(x, y), f_2(x, y))$ where

$$f_1(x, y) = 2e^{y_1} + y_2x_1 - 4x_2 + 3$$
$$f_2(x, y) = y_2 \cos y_1 - 6y_1 + 2x_1 - x_3$$

We have $f \in C^1(\mathbb{R}^5, \mathbb{R}^2)$ and

$$J_f(x, y) = \begin{bmatrix} y_2 & -4 & 0 & 2e^{y_1} & x_1 \\ 2 & 0 & -1 & -y_2 \sin y_1 - 6 & \cos y_1 \end{bmatrix}$$

At $(3, 2, 7, 0, 1)$

$$J_f(3, 2, 7, 0, 1) = \begin{bmatrix} 1 & -4 & 0 & 2 & 3 \\ 2 & 0 & -1 & -6 & 1 \end{bmatrix}$$

Hence

$$A_x = \begin{bmatrix} 1 & -4 & 0 \\ 2 & 0 & -1 \end{bmatrix}, \qquad A_y = \begin{bmatrix} 2 & 3 \\ -6 & 1 \end{bmatrix}$$

Now $\det A_y = 2 + 18 = 20 \neq 0$, so A_y is invertible. Thus by the Implicit Function Theorem (13.5), there exists an open neighborhood $W \subset \mathbb{R}^3$, of $(3, 2, 7)$, and $g : W \to \mathbb{R}^2$, continuously differentiable with $g(3, 2, 7) = (0, 1)$. Also

$$f(x, g(x)) = 0 \qquad \forall x \in W$$

We have $(Dg)(3,2,7) = -A_y^{-1}A_x$, where $A_y^{-1} = \frac{1}{20}\begin{bmatrix}1 & -3\\ 6 & 2\end{bmatrix}$, thus

$$(Dg)(3,2,7) = \begin{bmatrix}\frac{1}{4} & \frac{1}{5} & \frac{3}{20}\\ -\frac{1}{2} & \frac{6}{5} & \frac{1}{10}\end{bmatrix}$$

This does not give the partial derivative of g at $(3,2,7)$.

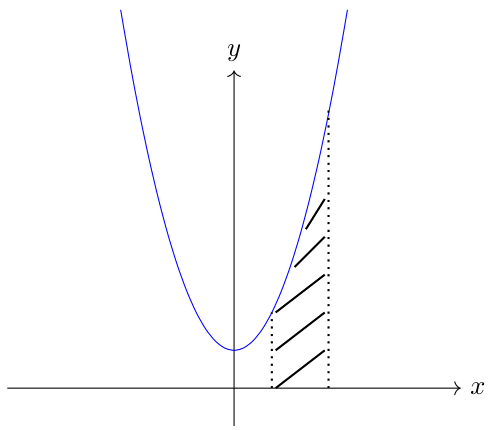
Lecture 30 - Monday, Jul 15

14 Integration on $\mathbb{R}^N$

Suppose $f : [a,b] \to \mathbb{R}$, $f \geq 0$, f is Riemann Integrable. Then

$$\int_a^b f\ dx$$

represents the area under the graph of f:



$\displaystyle\int f\,\mathrm{d}x$ is defined as the limit of Riemann Sums, so that

$$\int f\,\mathrm{d}x \approx \sum_{i=1}^{n} f(x_i)(x_i - x_{i-1})$$

Suppose $f : [a,b]\times[c,d] \to \mathbb{R}$, $f(x) = e$, $e \geq 0$, then we expect the $\displaystyle\int f$ to be the "volume" under the graph of f, so that

$$\int f = e\cdot(b-a)\cdot(d-c)$$

We wish to define th Riemann integral of $f : A \to \mathbb{R}$, $f \geq 0$ via a limit process.

We start by considering function defined on rectangles

$$I = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_N, b_N] \subset \mathbb{R}^N$$

Definition 14.1: Volumn (Content)

We define the **volume** of I (also called the **content** of I) by

$$\mu(I) = Vol(I) = \prod_{i=1}^{N} (b_i - a_i)$$

Definition 14.2: Partition

For each $j = 1, \ldots, N$, let $a = t_{j,0} < t_{j,1} < \cdots < t_{j,n_j} = b_j$ be a partition of the closed interval $[a_j, b_j]$, and define

$$P_j = \{t_{j,l} : l = 0, \ldots, n_j\}$$

Then the Cartesian Product $P = P_1 \times \cdots \times P_N$ is called a **partition of** I. A partition P of I gives the subdivision of I into $n_1 \times \cdots \times n_N$ subrectangles, which are called the subrectangles corresponding to P. So for each j and $1 \leq k_j \leq N$, we have a subrectangle

$$I = [t_{1,k_1-1}, t_{1,k_1}] \times [t_{2,k_2-1}, t_{2,k_2}] \times \cdots \times [t_{N,k_N-1}, t_{N,k_N}]$$

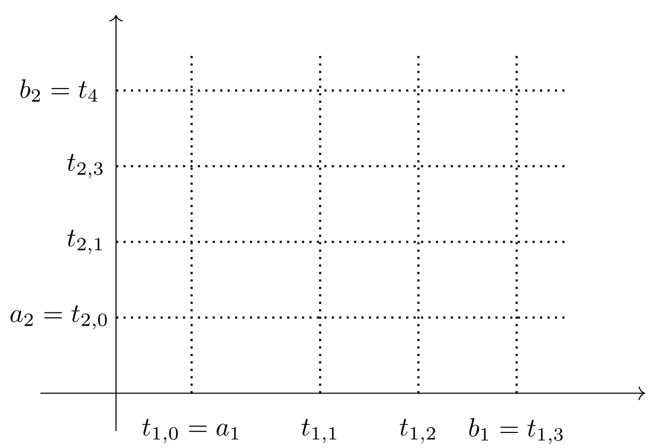


Figure 1: Subdivision generated by a partition

14.1 Riemann Sum

Definition 14.3: Riemann Sum

Let $I = [a_1, b_1] \times \cdots \times [a_N, b_N] \subset \mathbb{R}^N$ be a rectangle and $f : I \to \mathbb{R}^M$ be a function. Let P be a partition of I. For each rectangle I_α in the subdivision of I corresponding to P choose $x_\alpha \in I_\alpha$, then the sum

$$S(f, P) := \sum_{\alpha \in P} f(x_\alpha)\mu(I_\alpha)$$

is called the **Riemann Sum** of f corresponding to P.

Discovery 14.1

Notice that the sum $S(f, P)$ depends on the partition P and also on the choice of points $x_\alpha \in I_\alpha$.

Definition 14.4: Refinement

Let $P = P_1 \times \cdots \times P_N$ be a partition of I, we say that a partition Q is a **refinement** of P if $P_j \subset Q_j$ for all $j = 1, \ldots, N$.

Discovery 14.2

Suppose P is a partition of I, then

$$I = \bigcup_{\alpha \in P} I_\alpha \quad \text{and} \quad \mu(I) = \sum_{\alpha \in P} \mu(I_\alpha)$$

Proof. Prove this by induction on N. The result holds because the rectangles I_α's may overlap at most along their boundaries, so Q is a refinement of P, then for each $\alpha \in P$,

$$I_\alpha = \bigcup_{\substack{\beta \in Q \\ J_\beta \subset I_\alpha}} \quad \text{and so} \quad \mu(I_\alpha) = \sum_{\substack{\beta \in Q \\ J_\beta \subset I_\alpha}} \mu(J_\beta)$$

□

Discovery 14.3

Suppose P and Q are partitions of I, then there is always a **common** refinement R of P and Q. For example,

$$R = R_1 \times \cdots \times R_N$$

where $R_j := P_j \cup Q_j$ for $j = 1, \ldots, N$.

14.2 Riemann Integrable

Definition 14.5: Riemann integrable

Let $I \subset \mathbb{R}^N$ be a rectangle and $f : I \to \mathbb{R}^M$ be a function. Suppose that there exists $y \in \mathbb{R}^M$ such that for every $\varepsilon > 0$, there exists a partition P_ε of I such that for each refinement P of P_ε and all Riemann sums $S(f, P)$ corresponding to P, we have

$$\|S(f,P) - y\| < \varepsilon$$

Then we say that f is **Riemann integrable** and y is the Riemann integral of f.

Notation:

$$y = \int_I f \qquad \int f \, \mathrm{d}\mu \qquad \int_I f(x_1, \ldots, x_N) \, \mathrm{d}\mu(x_1, \ldots, x_N)$$

Proposition 14.1

Suppose $f : I \to \mathbb{R}^M$ is Riemann integrable, then $\int_I f$ is unique.

Proof. **Exercise.** (The proof uses the uniqueness of limit). □

14.2.1 Cauchy Criterion for Riemann Integrable

Theorem 14.1: Cauchy Criterion for Riemann integrable

Let $I \subset \mathbb{R}^N$ be a rectangle and $f : I \to \mathbb{R}^M$, TFAE:

1. f is Riemann integrable;

2. For every $\varepsilon > 0$, there exists a partition P_ε such that for all refinement P and Q of P_ε and all Riemann sum $S(f, P)$ and $S(f, Q)$ corresponding to P and Q respectively, we have

$$\|S(f,P) - S(f,Q)\| < \varepsilon$$

Lecture 31 - Wednesday, Jul 17

Proof. 1. ($\Longrightarrow$)

Given $\varepsilon > 0$, let P_ε be a partition of I such that

$$\left\| S(f,P) - \int_I f \right\| < \frac{\varepsilon}{2}$$

for all refinements P of P_ε and Riemann sums $S(f, P)$. Thus if P and Q are refinements of P_ε and $S(f, P)$ and $S(f, Q)$ are Riemann sums corresponding to P and Q respectively, we have

$$\|S(f,P) - S(f,Q)\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

2. $(\Longleftarrow)$
Suppose 2. holds. Then for every $\varepsilon = \frac{1}{2^n}$ there exists a partition P_n of I such that

$$\|S(f,P) - S(f,Q)\| < \frac{1}{2^n}$$

for all refinements P and Q of P_n, and all Riemann sums $S(f,P)$ and $S(f,Q)$. By taking common refinements if necessary, we may assume that P_{n+1} is a refinement of P_n and in particular

$$\|S(f,P_{n+1}) - S(f,P_n)\| < \frac{1}{2^n}$$

for all Riemann sums corresponding to P_n and P_{n+1} respectively. For each n let y_n be a Riemann sum corresponding to the subdivision of I given by P_n. Thus $\|y_{n+1} - t_n\| < \frac{1}{2^n}$ for all n. It follows that (y_n) is a Cauchy sequence. Set $y := \lim_{n\to\infty} y_n$. We will show that $y = \int_I f$. Let $\varepsilon > 0$ be given. Choose k such that $\|y - y_n\| < \frac{\varepsilon}{2}$ for all $n \geq k$. Let $n \geq k$ such that $\frac{1}{2^n} < \frac{\varepsilon}{2}$. Set $P_\varepsilon := P_n$. Let P be a refinement of P_n and $\mathcal{S}(f,P)$ be a Riemann sum. By (10), $\|\mathcal{S}(f,P) - y_n\| < \frac{1}{2^n} < \frac{\varepsilon}{2}$. Thus

$$\|\mathcal{S}(f,P) - y\| < \|\mathcal{S}(f,P) - y_n\| + \|y_n - y\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

giving that $y = \int_I f$, and f is Riemann integrable. □

Discovery 14.4

Let $I \subset \mathbb{R}^N$ be a rectangle and $f : I \to \mathbb{R}^M$ be a function. Then f is Riemann integrable if and only if each component $f_j : I \to \mathbb{R}$, $j = 1, \dots, M$ of f is Riemann integrable (see A5).

Corollary 14.1

Let $I \subset \mathbb{R}^N$ be a rectangle and $f : I \to \mathbb{R}^M$ be a function. TFAE:

1. f is Riemann integrable;
2. For every $\varepsilon > 0$, there exists a partition P_ε of I such that

$$\|S_1(f,P_\varepsilon) - S_2(f,P_\varepsilon)\| < \varepsilon$$

   for all Riemann sums $S_1(f,P_\varepsilon)$ and $S_2(f,P_\varepsilon)$ corresponding to P_ε.

Proof. 1. $\Longrightarrow$ 2. is by *Theorem 5.7.*
2. $\Longrightarrow$ 1. Suppose 2. holds. By the preceding remark, we may assume $M = 1$. Let $\varepsilon > 0$ be given and let P_ε be a partition of I as in 2. Let P and Q be refinements of P_ε and let

$$\mathcal{S}(f,P) = \sum_{\beta \in P} f(x_\beta)\mu(J_\beta) \quad \text{and} \quad \mathcal{S}(f,Q) = \sum_{\gamma \in Q} f(x_\gamma)\mu(K_\gamma)$$

be Riemann sums associated to P and Q respectively. Then for each $\alpha \in P_\varepsilon$ we have

$$I_\alpha = \bigcup_{\beta \in P, J_\beta \subseteq I_\alpha} J_\beta = \bigcup_{\gamma \in Q, K_\gamma \subseteq I_\alpha} K_\gamma$$

and

$$\mu(I_\alpha) = \sum_{\beta \in P, J_\beta \subseteq I_\alpha} \mu(J_\beta) = \sum_{\gamma \in Q, K_\gamma \subseteq I_\alpha} \mu(K_\gamma)$$

by Discovery (14.2). For each $\alpha \in P_\varepsilon$ let

$$B_\alpha = \{f(x_\beta) \mid \beta \in P, J_\beta \subseteq I_\alpha\} \cup \{f(x_\gamma) \mid \gamma \in Q, K_\gamma \subseteq I_\alpha\}$$

Then B_α is finite and we let $z_\alpha, w_\alpha \in I_\alpha$ such that

$$f(z_\alpha) = \max B_\alpha, \quad f(w_\alpha) = \min B_\alpha.$$

Then

$$\begin{aligned} f(w_\alpha) \leq f(x_\beta) \leq f(z_\alpha), \quad \forall \beta \in P, J_\beta \subseteq I_\alpha \\ f(w_\alpha) \leq f(x_\gamma) \leq f(z_\alpha), \quad \forall \gamma \in P, K_\gamma \subseteq I_\alpha. \end{aligned}$$

We have

$$\begin{aligned} \mathcal{S}(f, P) - \mathcal{S}(f, Q) &= \sum_{\beta \in P} f(x_\beta)\mu(J_\beta) - \sum_{\gamma \in Q} f(x_\gamma)\mu(K_\gamma) \\ &= \sum_{\alpha \in P_\varepsilon} \left(\sum_{\beta \in P, J_\beta \subseteq I_\alpha} f(x_\beta)\mu(J_\beta) - \sum_{\gamma \in Q, K_\gamma \subseteq I_\alpha} f(x_\gamma)\mu(K_\gamma) \right) \\ &\leq \sum_{\alpha \in P_\varepsilon} \left(f(z_\alpha) \sum_{\beta \in P, J_\beta \subseteq I_\alpha} \mu(J_\beta) - f(w_\alpha) \sum_{\gamma \in Q, K_\gamma \subseteq I_\alpha} \mu(K_\gamma) \right) \\ &= \sum_{\alpha \in P_\varepsilon} f(z_\alpha)\mu(I_\alpha) - \sum_{\alpha \in P_\varepsilon} f(w_\alpha)\mu(I_\alpha) \\ &= S_1(f, P_\varepsilon) - S_2(f, P_\varepsilon) < \varepsilon. \end{aligned}$$

Similarly,

$$S(f, P) - S(f, Q) \leq S_1(f, P_\varepsilon) - S_2(f, P_\varepsilon) > -\varepsilon \implies \|S(f, P) - S(f, Q)\| < \varepsilon$$

by Theorem (14.1) (2. $\implies$ 1.), f is Riemann integrable. □

Theorem 14.2

Let $I \subset \mathbb{R}^N$ be a rectangle and $f : I \to \mathbb{R}^M$ be continuous. Then f is Riemann integrable.

Proof. Since I is compact and f is continuous, then f is uniformly continuous on I. Given $\varepsilon > 0$, let $\delta > 0$ be such that

$$\|f(x) - f(y)\| < \frac{\varepsilon}{\mu(I)}$$

for all $x, y \in I$, $\|x - y\| < \delta$. Choose a partition P_ε of I such that $x, y \in I_\alpha$, $\|x - y\| < \delta$ for all $\alpha \in P_\varepsilon$. Let

$$S_1(f, P_\varepsilon) = \sum_{\alpha \in P_\varepsilon} f(x_\alpha)\mu(I_\alpha), \quad S_2(f, P_\varepsilon) = \sum_{\alpha \in P_\varepsilon} f(y_\alpha)\mu(I_\alpha)$$

be Riemann sums corresponding to P_ε. Then

$$\begin{aligned}
\|S_1(f, P_\varepsilon) - S_2(f, P_\varepsilon)\| &= \left\| \sum_{\alpha \in P_\varepsilon} (f(x_\alpha) - f(y_\alpha))\mu(I_\alpha) \right\| \\
&\le \sum_{\alpha \in P_\varepsilon} \|f(x_\alpha) - f(y_\alpha)\| \mu(I_\alpha) \\
&< \sum_{\alpha \in P_\varepsilon} \frac{\varepsilon}{\mu(I)} \mu(I_\alpha) \\
&= \varepsilon
\end{aligned}$$

since $x_\alpha, y_\alpha \in I_\alpha \implies \|x_\alpha - y_\alpha\| < \delta$. By Corollary (14.1), f is Riemann integrable. □

Lecture 32 - Friday, Jul 19

14.3 Content Zero

Definition 14.6: Content Zero

We say that a set $A \subset \mathbb{R}^N$ has **content zero**, write $\mu(A) = 0$, if for every $\varepsilon > 0$, the rectangle $I_1, \ldots, I_n$ (may overlap, finitely many) with

$$A \subset \bigcup_{j=1}^{n} I_j \qquad \text{and} \qquad \sum_{j=1}^{n} \mu(I_j) < \varepsilon$$

Note: if $A \subset B$ and B has a content zero, then A has content zero.

Example 14.1: Examples of content zero

1. Finite set;
2. If $A_1, \ldots, A_m$ have content zero, then their union has content zero;
3. If $I \subset \mathbb{R}^N$ is a rectangle, then ∂I has content zero. This is because ∂I is a finite union of sets of the form $[a, b] \times \cdots \times [a_{i-1}, b_{i-1}] \times \{c_i\} \times [a_{i+1}, b_{i+1}] \times [a_n, b_n]$, where $c_i \in [a_i, b_i]$.

Proposition 14.2

Suppose $K \subset \mathbb{R}^N$ is compact and $f : K \to \mathbb{R}$ is continuous, then $\text{graph}(f) = \{(x, f(x)) : x \in K\} \subset \mathbb{R}^{N+1}$ has content zero.

Proof. See A5. □

Example 14.2: Examples of non-content zero

1. $\mathbb{Z}$;
2. $\mathbb{Q}$;
3. $\mathbb{Q} \cap [0, 1]$.

14.4 Measure Zero

Definition 14.7: Measure Zero

Let $A \subset \mathbb{R}^N$, we say that A has **measure zero** if for every $\varepsilon > 0$, there are countably many (possibly infinite) rectangles $I_1, I_2, \ldots$ in $\mathbb{R}^N$ such that

$$A \subset \bigcup_{j=1}^{\infty} I_j \qquad \text{and} \qquad \sum_{j=1}^{\infty} \mu(I_j) < \varepsilon$$

Discovery 14.5

1. $A \subset B$ and B has measure zero implies that A has measure zero;
2. A has content zero implies A has measure zero; (How does this work? Choose all the subsequent rectangles to be $\varnothing$, iykyk :3).

Proposition 14.3

Suppose $A_1, A_2, \ldots, A_n, \ldots$ are subsets of $\mathbb{R}^N$ with measure zero, then $A = \bigcup_{i=1}^{\infty} A_i$ has measure zero.

Proof. Let $\varepsilon > 0$. For each $i = 1, \ldots$, let $I_{i,1}, I_{i,2}, \ldots$ be a coutable collection of ractangles such that

$$A_i \subset \bigcup_{j=1}^{\infty} I_{i,j} \qquad \text{and} \qquad \sum_{j=1}^{\infty} \mu[I_{i,j}] < \frac{\varepsilon}{2^i}$$

Then

$$A \subset \bigcup_{i=1}^{\infty} \left(\bigcup_{j=1}^{\infty} I_{i,j} \right) \qquad \text{and} \qquad \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mu[I_{i,j}] < \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon$$

Since $\mathbb{N} \times \mathbb{N}$ is countable, we get A has measure zero. □

Example 14.3

Countable set have measure zero (e.g. $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{Q} \cap [0, 1]$), while $[0, 1] \backslash \mathbb{Q}$ does not have measure zero.

Theorem 14.3

Suppose $K \subset \mathbb{R}^N$ is compact and has measure zero, then K has content zero.

Proof. Let $\varepsilon > 0$ and let $I_1, I_2, \ldots$ be rectangles with

$$K \subset \bigcup_{j=1}^{\infty} I_j \qquad \text{and} \qquad \sum_{j=1}^{\infty} \mu[I_j] < \frac{\varepsilon}{2}$$

For each j choose I_j' a rectangle with $I_j'^{\circ} \supset I_j$ and

$$\mu(I_j') < \mu(I_j) + \frac{\varepsilon}{2^j}$$

By compactness, there are rectangles $I_{j1}', \ldots, I_{jn}'$ such that

$$K \subset \bigcup_{i=1}^{n} I_{ji}'^{\circ} \subset \bigcup_{i=1}^{n} I_{ji}'$$

$$\sum_{i=1}^{n} \mu(I_{ji}') \leq \sum_{j=1}^{\infty} \mu(I_j)' \leq \sum_{j=1}^{\infty} \left(\mu(I_j) + \frac{\varepsilon}{2^j}\right) < \varepsilon$$

□

Definition 14.8: "Has Content"

1. Let $\varnothing \neq D \subset \mathbb{R}^N$ be bounded and let $I \subset \mathbb{R}^N$ be rectangles containing D. We say that a function $f : D \to \mathbb{R}^M$ is Riemann integrable on D if the $\overline{f} : I \to \mathbb{R}^M$ given by $\overline{f}(x) = \{f(x) : x \in D \text{ or } 0 : x \text{ otherwise}\}$ is Riemann integrable, in which case we define the integral of f on D by

$$\int_D f = \int_I \overline{f}$$

2. Let $\varnothing \neq D \subset \mathbb{R}^N$ be bounded, we say that D **Has Content** if the **Characteristic Function** on D is integrable, where

$$\mathcal{X}_D : \mathbb{R}^N \to \mathbb{R}^M, \;\; \mathcal{X}_D(x) = \begin{cases} 1 & \text{if } x \in D \\ 0 & \text{otherwise} \end{cases}$$

We define the content of D (the volume) by

$$\mu(D) = \int_D \mathcal{X}_D = \int_D 1$$

Discovery 14.6

If $D = I$ is rectangle, then it coincides with the volume of I.

14.5 Lebesgue Theorem

Theorem 14.4: Lebesgue Theorem

Let $I \subset \mathbb{R}^N$ be a rectangle and let $f : I \to \mathbb{R}^M$ be bounded, then f is Riemann integrable if and only if the set $B_f = \{x \in I : f \text{ is not countinuous at } x\}$ has measure zero.

Proof. Notice that we may assume $M = 1$ because

$$B_f = \bigcup_{j=1}^{M} B_{f_j}$$

where $B_{f_j} = \{x \in I : f_j \text{ is not countinuous at } x\}$, f_j is component of f.

1. $(\Longleftarrow)$
   We define for $x \in I$ the **ocsillation** of f at x by

   $$\mathfrak{o}(f,x) = \lim_{\delta \to 0} [M(x,f,\delta) - m(x,f,\delta)]$$

   where $M(x,f,\delta) = \sup\{f(y) : y \in \mathcal{B}_\delta(x)\}$ and $m(x,f,\delta) = \inf\{f(y) : y \in \mathcal{B}_\delta(x)\}$. The limit above exists because the function

   $$\delta \mapsto M(x,f,\delta) - m(x,f,\delta)$$

   is decreasing. Notice also $\mathfrak{o}(f,x) \geq 0$.

   (a) *Claim 1*: f is continuous at x if and only if $\mathfrak{o}(f,x) = 0$;

   (b) *Claim 2*: For every $\varepsilon > 0$ the set $B_\varepsilon = \{x \in I : \mathfrak{o}(f,x) \geq \varepsilon\}$ is closed (in particular, B_ε is compact).

   Proof. We will prove that $B_\varepsilon^c \cap I$ is relatively open in I. Let $x \in I$ with $\mathfrak{o}(f,x) < \varepsilon$. Let $\delta > 0$ be such that $M(x,f,\delta) - m(x,f,\delta) < \varepsilon$. Let $y \in \mathcal{B}_\delta(x)$ and take $\delta_y > 0$ such that $\mathcal{B}_{\delta_y}(y) \subset \mathcal{B}_\delta(x)$, then

   $$M(y,f,\delta_y) - m(y,f,\delta_y) \leq M(x,f,\delta) - m(x,f,\delta) < \varepsilon$$

   giving that $\mathfrak{o}(f,y) < \varepsilon$. Thus B_ε is relatively closed in I, so B_ε is closed. □

   Lecture 33 - Monday, Jul 22

   Notice that $B_\varepsilon \subset B_f$ by claim 1, hence B_ε has measure zero. Thus B_ε has content zero by Theorem (14.3). Let $\varepsilon > 0$ be given, let $U_1, U_2, \ldots, U_n$ be rectangles such that $B_\varepsilon \subset \bigcup_{j=1}^n U_j^\circ$ (union of intervals) and $\sum_{j=1}^n \mu(I_j) < \varepsilon$. Let P'_ε be a partition of I such that for each $\alpha \in P'_\varepsilon$, the rangles I_α has one of the following properties:

   (a) $I_\alpha \subset U_j$ for some $j = 1, 2, \ldots, n$, or;

   (b) $I_\alpha \cap B_\varepsilon \neq \varnothing$.

This can be done by considering the rectangles $U_j \cap I$, and because if

$$I_\alpha \cap \left[\bigcup_{j=1}^{n} (U_j \cap I)^\circ\right] = \varnothing$$

then $I_\alpha \cap B_\varepsilon = \varnothing$. Let $M \geq 0$ be such that $|f(x)| \leq M$ for all $x \in I$, then

$$|f(x_\alpha) - f(y_\alpha) \leq 2M \qquad \forall\, x_\alpha, y_\alpha \in I_\alpha|$$

Now we get

$$\begin{aligned}\left|\sum_{I_\alpha \subset U_j \text{ for some j}} [f(x_\alpha) - f(y_\alpha)]\, \mu(I_\alpha)\right| &\leq \sum_{I_\alpha \subset U_j \text{ for some j}} |f(x_\alpha) - f(y_\alpha)|\, \mu(I_\alpha) \\ &\leq 2M \sum_{I_\alpha \subset U_j \text{ for some j}} \mu(I_\alpha) \\ &\leq 2M \sum_{j=1}^{n} \mu(U_j) = 2M\varepsilon\end{aligned}$$

(a) *Claim 3*: If $\alpha \in P'_\varepsilon$ and $I_\alpha \cap B_\varepsilon = \varnothing$, then there exists a partition P_α of I_α such that

$$|f(x_\beta) - f(y_\beta)| \leq 2\varepsilon \qquad \forall\, x_\beta, y_\beta \in J_{\alpha,\beta}$$

where $J_{\alpha,\beta}$ is a subrectangle in the in the subdivision corresponding to P_α.

Proof. Since $I_\alpha \cap B_\varepsilon = \varnothing$, we have $ø(f, x) < \varepsilon$ for all $x \in I_\alpha$. For each $x \in I_\alpha$, let $\delta_x > 0$ be such that

$$|f(y) - f(z)| < \varepsilon \qquad \forall\, y, z \in \mathcal{B}_{\delta_x}(x)$$

then

$$I_\alpha \subset \bigcup_{x \in I_\alpha} \mathcal{B}_{\delta_x/2}(x)$$

Let $\{x_1, x_2, \ldots, x_\ell\}$ be such that

$$I_\alpha \subset \bigcup_{i=1}^{\ell} \mathcal{B}_{\delta_{x_i}/2}(x_i)$$

Take $\delta = \min\{\delta_{x_i}/2 : i = 1, \ldots, \ell\}$. Let P_α be a partition of I_α such that x, y belong to the subrectangles, we have $\|x - y\| < \delta$. It follows that if $x_\beta, y_\beta \in I_{\alpha,\beta}$, then taking i such that $x_\beta \in \mathcal{B}_{\delta_{x_i}/2}(x_i)$, we have $y_\beta \in \mathcal{B}_{\delta_{x_i}}(x_i)$. This gives $|f(x_\beta) - f(y_\beta)| < 2\varepsilon$. □

It follows by Claim 3 that we can find a refinement P_ε of P'_ε with the properties above and also with the additional property that $|f(x_\alpha) - f(y_\alpha)| < 2\varepsilon$, where $\alpha \in P_\varepsilon$ and $I_\alpha \cap B_\varepsilon = \varnothing$. Let $S_1(f, P_\varepsilon)$ and

$S_2(f, P_\varepsilon)$ be Riemann sums corresponding to P_ε, then

$$\begin{aligned}\left|\sum_{\alpha\in P_\varepsilon}[f(x_\alpha)-f(y_\alpha)]\,\mu(I_\alpha)\right| &\le \sum_{\alpha\in P_\varepsilon, I_\alpha\subset U_j}|f(x_\alpha)-f(y_\alpha)|\,\mu(I_\alpha)+\sum_{\alpha\in P_\varepsilon, I_\alpha\cap U_j=\varepsilon}|f(x_\alpha)-f(y_\alpha)|\,\mu(I_\alpha)\\ &\le 2M\varepsilon+2\varepsilon\mu(I)\end{aligned}$$

Thus by Corollary 14.1, f is Riemann Integrable.

2. ($\Longrightarrow$)
Suppose f is Riemann integrable, for each n, let

$$B_{1/n}=\{x\in I:\o(f,x)\ge\frac{1}{n}\}$$

By Claim 1,

$$B_f=\bigcup_{n=1}^{\infty}B_{1/n}$$

Thus STP that each $B_{1/n}$ has measure zero (in fact, content zero). Fix n and let $\varepsilon>0$. Let P_ε be a partition of I such that

$$S_1(f,P_\varepsilon)-S_2(f,P_\varepsilon)<\frac{\varepsilon}{n}$$

for all Riemann sums $S_1(f, P_\varepsilon)$ and $S_2(f, P_\varepsilon)$. Write

$$B_{1/n}=C_1\cup C_2 \qquad \text{where } C_1=\{x\in B_{1/n}:x\in\partial I_\alpha \text{ for some } \alpha\}$$
$$C_2=\{x\in B_{1/n}:x\in I_\alpha^\circ \text{ for some } \alpha\}$$

Then C_1 has content zero because each I_α does. Let

$$\mathbb{S}=\{I_\alpha:I_\alpha^\circ\cap C_2\ne\varnothing\}$$

Then $C_2\subset\bigcup_{I_\alpha\in\mathbb{S}}I_\alpha$. Given $\varepsilon'>0$, $\varepsilon'<1/n$, for each $I_\alpha\in\mathbb{S}$, we can find $x_\alpha, y_\alpha\in I_\alpha$ such that

$$f(x_\alpha)-f(y_\alpha)>\frac{1}{n}-\varepsilon'$$

since $I_\alpha^\circ\cap C_2\ne\varnothing$. It follows that

$$\begin{aligned}0\le\sum_{I_\alpha\in\mathbb{S}}\left(\frac{1}{n}-\varepsilon'\right)\mu(I_\alpha)&\le\sum_{I_\alpha\in\mathbb{S}}\left(f(x_\alpha)-f(y_\alpha)\right)\mu(I_\alpha)\\ &=S_1(f,P_\varepsilon)-S_2(f,P_\varepsilon)<\frac{\varepsilon}{n}\end{aligned}$$

Since $\varepsilon'>0$, this yields that

$$\sum_{I_\alpha\in\mathbb{S}}\frac{\mu(I_\alpha)}{2}\le\frac{\varepsilon}{2}\ \Rightarrow\ \sum_{I_\alpha\in\mathbb{S}}\mu(I_\alpha)\le\varepsilon$$

so C_2 has content zero as needed.

□

Corollary 14.2

Let $\varnothing \neq D \subset \mathbb{R}^N$ be bounded, TFAE:

1. D has content;
2. ∂D has content zero.

Corollary 14.3

Let $\varnothing \neq D \subset \mathbb{R}^N$ be bounded and ∂D has content zero. If $f : D \to \mathbb{R}^M$ is continuous, then f is Riemann integrable.

Corollary 14.4

Let $f : I \to \mathbb{R}^M$ and suppose the set of points at which f is discontinuous is countable, then f is Riemann integrable.

Proposition 14.4: Properties of the Riemann integrable

Let $\varnothing \neq D \subset \mathbb{R}^N$ be bounded, let $f, g; D \to \mathbb{R}^M$ be Riemann integrable, then

1. $f + g$ is Riemann integrable, and;
$$\int (f+g) = \int f + \int g$$
2. $\|f\| : D \to \mathbb{R}$, $x \mapsto \|f(x)\|$ is Riemann integrable, and;
3. If $M = 1$, $f \leq g$, then
$$\int f \leq \int g$$
4. If $M = 1$, D has content and $r \leq f \leq R$, then
$$r\mu(D) \leq \int f \leq R\mu(D)$$

Lecture 34 - Wednesday, Jul 24

14.5.1 Mean Value Theorem for Integration

Theorem 14.5: Mean Value Theorem for Integration

Let $\varnothing \neq D \subset \mathbb{R}^N$ and $f : D \to \mathbb{R}$ continuous on D. Suppose that D is compact, connected and has content. Then there exists $x_0 \in D$ such that

$$\int_D f = f(x_0)\mu(D)$$

Proof. Since D has content, and f is continuous, then f is Riemann integrable by Corollary (14.3). Let $r, R \in \mathbb{R}$ such that

$$r \leq f \leq R$$

By extreme value theorem, there are $p, q \in D$ such that

$$f(p) = r \qquad \text{and} \qquad f(q) = R$$

We have

$$r\mu(D) \leq \int f \leq R\mu(D)$$

so if $\mu(D) = 0$, $\displaystyle\int f = 0$ and any $x_0 \in D$ satisfies the result. Assume $\mu(D) \neq 0$ and let

$$\lambda := \frac{\int f}{\mu(D)}$$

so $f(p) \leq \lambda \leq f(q)$ and since D is connected, there exists by the intermediate value theorem $x_0 \in D$ such that

$$f(x_0) = \lambda = \frac{\int f}{\mu(D)}$$

□

14.6 Fubini's Theorem

How do we actually calculate the integral, $\displaystyle\int_D f$?

Example 14.4

Using a simple exmaple to show the idea: Suppose $I = [a, b] \times [c, d] \subset \mathbb{R}^2$ and $f : I \to \mathbb{R}$ continuous, $f \geq 0$. Hence $\displaystyle\int f$ is the volume of the region under the graph of f. In particular, we have

$$\int_I f = \int_a^b \left(\int_c^d f(x, y)\ dy \right)\ dx = \int_c^d \left(\int_a^b f(x, y)\ dx \right)\ dy$$

It could happen in general that for some x, the function $y \mapsto f(x, y)$ is not Riemann Integrable.

Theorem 14.6: Fubini's Theorem

Let $I \subset \mathbb{R}^N$ and $J \subset \mathbb{R}^M$ be rectangles and $f : I \times J \to \mathbb{R}^K$ be Riemann Integrable. Suppose that for each $x \in I$ the function $y \in J \mapsto f(x,y) \in \mathbb{R}^K$ is Riemann integrable and let

$$h(x) = \int_J f(x,y)\, dy \qquad (x \in I)$$

Then h is integrable and

$$\int_I \left(\int_J f(x,y)\, dy \right)\, dx = \int_I h(x)\, dx = \int_{I\times J} f$$

Discovery 14.7

A similar statement holds if $x \mapsto f(x,y)$ is integrable for each $y \in J$ and we let $g(y) = \int_I f(x,y)\, dx$.

Proof. We may assume that K is 1 by A5Q2. Let $\varepsilon > 0$ be given and P_ε be a partition of $I \times J$ such that

$$\left| S(f,P) - \int f \right| < \frac{\varepsilon}{2}$$

for all refinement P of P_ε and all Riemann sum corresponding to P. Let P_ε^I and P_ε^J be partitions of I and J respectively, so that

$$P_\varepsilon = P_\varepsilon^I \times P_\varepsilon^J$$

Let P^I and P^J be refinements of P_ε^I and P_ε^J respectively and for each $\alpha \in P^I$ and $\beta \in P^J$ chosoe $x_\alpha \in I_\alpha$ and $y_\beta \in J_\beta$, then the above inequality yields

$$\left| \sum_{(\alpha,\beta)\in P^I\times P^J} f(x_\alpha, y_\beta)\mu(I_\alpha \times J_\beta) - \int_{I\times J} f \right| < \frac{\varepsilon}{2}$$

Then since $\mu(I_\alpha \times J_\beta) = \mu(I_\alpha)\mu(J_\beta)$, we get

$$\left| \sum_{\alpha\in P^I} \left(\sum_{\beta\in P^J} f(x_\alpha, y_\beta)\mu(J_\beta) \right) \mu(I_\alpha) - \int_{I\times J} f \right| < \frac{\varepsilon}{2}$$

Fix P^I and $x_\alpha \in I_\alpha$, let Q_ε^J be a refinement of P_ε^J such that

$$\left| \sum_{\beta\in Q_\varepsilon^J} f(x_\alpha, y_\beta)\mu(J_\beta) - h(x_\alpha) \right| < \frac{\varepsilon}{2\mu(I)}$$

for all $\alpha \in P_I$. Then combining this with the previous inequality, we have

$$\begin{aligned} &\left| \sum_{\alpha \in P^I} \left(\sum_{\beta \in Q_\varepsilon^J} f(x_\alpha, y_\beta)\mu(J_\beta) \right) \mu(I_\alpha) - \sum_{\alpha \in P_J} h(x_\alpha)\mu(I_\alpha) \right| \\ &\le \sum_{\alpha \in P^I} \left| \sum_{\beta \in Q_\varepsilon^I} f(x_\alpha, y_\beta)\mu(I_\beta)\mu(I_\alpha) - h(x_\alpha)\mu(I_\alpha) \right| \\ &< \sum_{\alpha \in P_I} \frac{\varepsilon}{2\mu(I)} \cdot \mu(I) = \frac{\varepsilon}{2} \end{aligned}$$

Thus we know that

$$\left| \sum_{\alpha \in P^I} h(x_\alpha)\mu(I_\alpha) - \int_{I \times J} f \right| < \varepsilon$$

which implies that h is integrable and $\displaystyle\int h(x)\, dx = \int f$. $\square$

Corollary 14.5

Let $I = [a, b] \times [c, d] \subset \mathbb{R}^2$ and $f : I \to \mathbb{R}$ be integrable. Suppose that the function

$$y \mapsto f(x, y) \qquad \text{and} \qquad x \mapsto f(x, y)$$

are integrable for all $x \in [a, b]$ and $y \in [c, d]$, then

$$\int_a^b \int_c^d f(x, y)\, dy\, dx = \int_I f = \int_c^d \int_a^b f(x, y)\, dx\, dy$$

Example 14.5

Let $I = [0, 1] \times [0, 1]$ and let $f(x, y) = y^3 e^{xy^2}$. Then

$$\begin{aligned} \int_0^1 \left(\int_0^1 y^3 e^{xy^2}\, dy \right) dx &= \int_0^1 \left(\int_0^1 y^3 e^{xy^2}\, dx \right) dy \\ &= \int_0^1 \frac{y^3 e^{xy^2}}{y^2} \bigg|_0^1 dy \\ &= \int_0^1 y \left(e^{y^2} - 1 \right) dy = \frac{e}{2} - 1 \end{aligned}$$

Corollary 14.6

Let $\varphi, \psi : [a, b] \to \mathbb{R}$ be continuous and let $D = \{(x, y) : x \in [a, b],\ \text{and}\ \varphi(x) \le y \le \psi(x)\} \subset \mathbb{R}^2$.

Suppose that $F : D \to \mathbb{R}$ is continuous, then

$$\int_D f = \int_a^b \left(\int_{\varphi(x)}^{\psi(x)} f(x,y)\, dy \right) dx$$

Proof. Notice that ∂D has content zero because it is the finite union of graphs of continuous functions on compact set. By Corollary (14.3) f is integrable. Let $I = [a,b] \times [c,d]$ containing D and $\tilde{f}$ the extension of f to I by $\tilde{f}(x) = 0$ for $x \notin D$. For $x \in [a,b]$ fixed, the function $y \mapsto \tilde{f}(x,y)$ is continuous on $[c,d]$ at $\varphi(x)$ and $\psi(x)$. By Fubini

$$\begin{aligned}\int_D f = \int_I \tilde{f} &= \int_a^b \int_c^d \tilde{f}(x,y)\, dy\, dx \\ &= \int_a^b \int_{\varphi(x)}^{\psi(x)} f(x,y)\, dy\, dx\end{aligned}$$

as desired. □

Lecture 35 - Friday, Jul 26

Example 14.6

Let $D = \{(x,y) : 1 \le x \le 3, x^2 \le y \le x^2 + 1\}$. Compute the content (the area) of D.

Proof. We have by the above Corollary that

$$\int_D 1 = \int_1^3 \int_{x^2}^{x^2+1} 1\, dy\, dx = 2$$

□

Example 14.7

Compute $\int_D f$ where $f(x,y,z) = y$ and D is the region bounded by the plane $z = 0$, $x = 0$, $y = 0$ and $x + y + z = 1$.

Proof. We can desciribe D as following:

$$0 \le x \le 1 \quad 0 \le y \le 1 - x \quad 0 \le z \le 1 - x - y$$

Thus by Fubini's Theorem and the above Corollary we have

$$\int_D f = \int_{[0,1]^3} \tilde{f} = \int_{[0,1]} \int_{[0,1]^2} \tilde{f} = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} y\, dz\, dy\, dx = \frac{1}{24}$$

Note: other ways to describe D could be, for example

$$0 \le z \le 1 \quad 0 \le x \le 1 - z \quad 0 \le y \le 1 - z - x$$

□

14.7 Change of Variables

Consider the function $f(x,y) = \dfrac{1}{(x^2+y^2)^{3/2}}$ defined on D where $D = \{(x,y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 \leq 4\}$. We wish to compute $\int_D f$.

Discovery 14.8

The idea is to use polar coordinates.

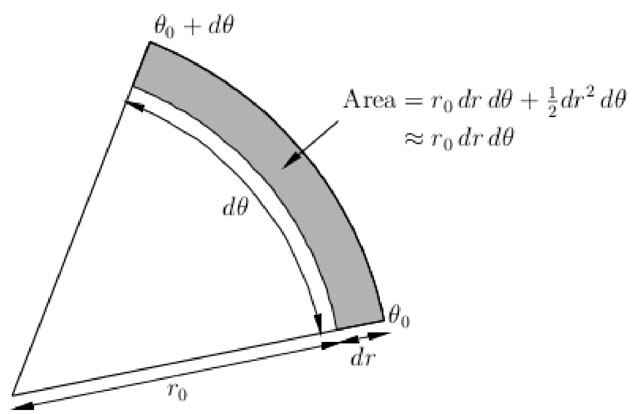
Suppose we have $g(r,\theta) = (r\cos\theta, r\sin\theta)$, then

$$D = g(A) \quad \text{where } A = \{(r,\theta) : 1 \leq r \leq 2, 0 \leq \theta < 2\pi\}$$

Hence D is replaced by a rectangle.
Also

$$(f \circ g)(r,\theta) = \frac{1}{r^3}$$

so everything looks simple. Can we compute $\int_D f$ in terms of $f \circ g$?
Consider an infinitesimal pizza-like box in polar coordinate:



The area of the shaded region would be

$$\frac{r^2\, d\theta}{2} - \frac{(r^2 - dr)\, d\theta}{2} \approx r\, dr\, d\theta \quad \text{if } dr \approx 0$$

so

$$\int_D f = \int_A f \circ g\, dA = \int_A f(r\cos\theta, r\sin\theta) r\, dr\, d\theta$$

Theorem 14.7: Change of Variable Theorem

Let $\varnothing \neq U \subset \mathbb{R}^N$ be open and let $\varnothing \neq K \subset U$ be compact with content. Suppose $g : U \to \mathbb{R}^N$ is continuously differentiable and suppose that there exists $Z \subset K$ with content zero such that

1. g is one-to-one on $K \backslash Z$;

2. $\det J_g(x) \neq 0$ for all $x \in K \backslash Z$,

then $g(K)$ has content and for every $f : g(K) \to \mathbb{R}$ continuous we have

$$\int_{g(K)} f = \int_K (f \circ g)\,|\det J_g|$$

where $\det J_g : K \to \mathbb{R}$ is defined as $x \mapsto \det J_g(x)$.

Example 14.8

Back to the example we had at the start. Consider $g(r,\theta) = (r\cos\theta, r\sin\theta)$, then $g \in C^1(\mathbb{R}^2, \mathbb{R}^2)$. We have

$$J_g(r,\theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} \quad \Rightarrow \quad \det J_g(r,\theta) = r$$

Notice that if $A = [1,2] \times [0, 2\pi]$, then $\det J_g(r,\theta) \neq 0$ on A and g is injective on $[1,2] \times [0, 2\pi)$. Since $[1,2] \times \{2\pi\}$ has content zero, we apply the Change of Variable Theorem:

$$\int_D f = \int_A f(r\cos\theta, r\sin\theta) r\; dr\; d\theta = \int_1^2 \int_0^{2\pi} \frac{1}{r^2}\; d\theta\; dr = \int_1^2 \frac{2\pi}{r^2}\; dr = \pi$$

14.8 Integration with Cylindrical Coordinates

The sylindrical coordinates in $\mathbb{R}^3$ are

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

Thus

$$g : \mathbb{R}^3 \to \mathbb{R}^3 \qquad g(r,\theta,z) = (r\cos\theta, r\sin\theta, z)$$

Then g is continuously differentiable and

$$J_g(r,\theta,z) = \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Thus $\det J_g(r,\theta,z) = r$.

Example 14.9

Find the volume of the region D in $\mathbb{R}^3$ above the paraboloid $z = x^2 + y^2$, and inside the sphere $x^2 + y^2 + z^2 = 12$.

Lecture 36 - Monday, Jul 29

Proof. Write that $x = r\cos\theta$, $y = r\sin\theta$ and $z = z$. On the paraboloid, we have $z = r^2$ while on the sphere, we have $r = \sqrt{12 - r^2}$. We now want to find the the value of r where the paraboloid and the sphere meets:

we have

$$r_{\max}^2 + r_{\max}^4 = 12 \quad \Rightarrow \quad r_{\max} = \sqrt{3}$$

Hence $D = g(K)$ where $g(r, \theta, z) = (r\cos\theta, r\sin\theta, z)$ and

$$K = \{(r, \theta, z) : 0 \le r \le \sqrt{3}, 0 \le \theta \le 2\pi, r^2 \le z \le \sqrt{12 - r^2}\}$$

By the Change of Variable Theorem,

$$\mu(D) = \int_D 1 = \int_K r\, dz\, d\theta\, dr = \int_0^{\sqrt{3}} \int_0^{2\pi} \int_{r^2}^{\sqrt{12-r^2}} r\, dz\, d\theta\, dr = 2\pi\left[-\frac{45}{4} + \frac{12^{3/2}}{3}\right]$$

□

14.9 Spherical Coordinates

In the system of spherical coordinates, we have the following coordinate axes:

1. ρ: the distance to the origin, so that $x^2 + y^2 + z^2 = \rho^2$, $(\rho \ge 0)$;
2. θ: "longitude" angle from the positive x-axis, $(0 \le \theta \le 2\pi)$;
3. φ: "latitude" angle from the positive z-axis, $(0 \le \varphi \le \pi)$.

Definition 14.9

We wish to denote (x, y, z) in terms of (ρ, θ, φ):

$$z = \rho\cos\varphi \qquad x = \rho\sin\varphi\cos\theta \qquad y = \rho\sin\varphi\sin\theta$$

Discovery 14.9

Consider $g : \mathbb{R}^3 \to \mathbb{R}^3$ where

$$g(\rho, \theta, \varphi) = (\rho\cos\varphi, \rho\sin\varphi\cos\theta, \rho\sin\varphi\sin\theta)$$

so $g \in C1(\mathbb{R}^3, \mathbb{R}^3)$ and g is injective on

$$\{(\rho, \theta, \varphi) : \rho > 0, 0 \le \theta < 2\pi, 0 \le \varphi \le \pi\}$$

Moreover

$$J_g(\rho, \theta, \varphi) = \begin{bmatrix} \cos\theta\sin\varphi & -\rho\sin\theta\sin\varphi & \rho\cos\theta\cos\varphi \\ \sin\theta\sin\rho & \rho\cos\theta\sin\varphi & \rho\sin\theta\cos\varphi \\ \cos\varphi & 0 & -\rho\sin\varphi \end{bmatrix}$$

so

$$\det J_g(\rho, \theta, \varphi) = -\rho^2\sin\varphi$$

Hence $\det J_g(\rho, \theta, \varphi) \neq 0$ if $\rho \neq 0$ and $\varphi \neq 0, \pi$.

Example 14.10: Example in Spherical Coordinates

Suppose $\rho = r$ is a non-zero constant and φ is also a constant not equal to 0 or π, then we get a cone with vertex at the origin.

Example 14.11

Compute the volume of the sphere with radius r using spherical coordinates:

Proof. We have

$$D = \{(x, y, z) : x^2 + y^2 + z^2 = r^2\}$$

hence $D = g(K)$, where $g(\rho, \theta, \varphi) = (\rho\cos\varphi, \rho\sin\varphi\cos\theta, \rho\sin\varphi\sin\theta)$ and

$$K = \{(\rho, \theta, \varphi) : 0 < \rho < r, 0 \le \theta < 2\pi, 0 \le \varphi \le \pi\}$$

Then

$$\mu(D) = \int_D 1 = \int_K |\det J_g| = \int_0^r \int_0^\pi \int_0^{2\pi} \rho^2 \sin\varphi \, d\theta \, d\varphi \, dr = \frac{4\pi}{3} r^3$$

□

Result 14.1: Idea of the proof for the Change of Variable Theorem

Suppose $I = [a_1, b_1] \times \cdots \times [a_N, b_N]$ and $a = (a_1, \ldots, a_N)$, then

$$I = \{a_1 + h_1 e_1 + \cdots a_N + h_N e_N : 0 \le h_k \le \ell_k \text{ for } 0 \le k \le N\}$$

where $\ell_k = b_k - a_k$. If I is very small,

$$g(I) \approx \left\{ g(a) + D_g(a) \begin{bmatrix} h_1 \\ \vdots \\ h_N \end{bmatrix}, 0 \le h_k \le \ell_k \text{ for } k = 1, \ldots, N \right\}$$

where

$$D_g(a) = \begin{bmatrix} | & & | \\ D_g(a)e_1 & \cdots & D_g(a)e_N \\ | & & | \end{bmatrix}$$

Then column vectors are linearly independent, and they form a parallelepiped. We observe that

$$\mu(\text{par}) = \left| \det \begin{bmatrix} | & & | \\ \ell_1 D_g(a)e_1 & \cdots & \ell_N D_g(a)e_N \\ | & & | \end{bmatrix} \right| = \mu(I) |\det J_g(a)|$$

Thus

$$\mu(g(I)) \approx \mu(I) |\det J_g(a)| \quad \Rightarrow \quad \int_I |\det J_g|$$

In general, take partition P of I,

$$\int g(K)I \approx \sum_{\alpha\in P}\int_{g(I_\alpha)} f = \sum_{\alpha\in P} f(y_\alpha)\int_{g(I_\alpha)} 1 = \sum_{\alpha\in P} f(y_\alpha)\mu(g(I_\alpha)) = \sum_{\alpha\in P} f(y_\alpha)\int_{I_\alpha} \det J_g$$

which yields $\int_I f \circ g|\det J_g|$.

Index

N-cell, 20

Angle, 8

Boundary, 14
Bounded, 18

Cauchy, 27
Characteristic Function, 87
Clopen, 9
Closed Ball, 8
Closed Set, 9
Closure, 13
Cluster Point, 12
Compact, 17
Connection, 23
Content, 80
Content Zero, 85
Continuous, 32
Contraction, 70
Convex, 24

Differentiable, 40
Differential, 41
Directional Derivative, 47
Disconnection, 23
Distance, 7
Divergence, 57

Euclidean Inner Product, 5

Gradient Notation, 47

Has Content, 87
Hessian, 64

Interior, 14

Jacobian Matrix, 47

Level Curves, 74
Limit, 26

Measure Zero, 86
Multindex, 60
Multivariate Polynomial, 63

Neighbourhood, 31
Norm, 6

Open ball, 8
Open Cover, 17
Open Set, 9
Orthogonal, 8

Partial Derivative, 45
Partition, 80

Refinement, 81
Riemann integrable, 82
Riemann Sum, 81

Saddle Point, 65
Sequence, 26
Stationary Point (Critiacal Point), 65
Subsequence, 28

Taylor Approximation, 63

Uniformly Continuous, 37

Vector Field, 56
Volumn, 80