
Analysis II: Multiple Variables

Spring Semester 2026

Lecture Notes

Contents

1 Metric Spaces	1
1.1 Topology of metric spaces	3
1.2 Normed Vector Spaces	7
2 Multidimensional Differentiation	9
2.1 Higher Order Derivatives	12
3 Optimization	15
3.1 First and Second Order Optimality Conditions	15

References

[1] Serra, Joaquim: Skript zur Vorlesung Analysis II. ETH Zürich. 2026

Analysis II: Multiple Variables

Fynn Krebsler—fkrebsler@student.ethz.ch

Spring Semester 2026

Lec 1

1 Metric Spaces

In analysis II, will be working mostly in \mathbb{R}^n , which is a vector space defined as

$$\mathbb{R}^n := \{x = (x_1, \dots, x_n), x_i \in \mathbb{R}\}.$$

Here, we call $n \in \mathbb{N}$ the **DIMENSION**. We want to add some extra structure in particular the **EUCLIDEAN STRUCTURE**

Definition 1.1: Euclidean structure

On \mathbb{R}^n we define the **EUCLIDEAN NORM** as

$$\|x\| := \sqrt{x_1^2 + \dots + x_n^2},$$

which describes the length of the vector x . We also define the **SCALAR PRODUCT** as

$$x \cdot y = \langle x, y \rangle := \sum_{i=1}^n x_i y_i,$$

furthermore the **EUCLIDEAN DISTANCE** is defined as

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

Lemma 1.2: Triangle inequality

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

$$\|z - x\| \leq \|y - x\| + \|z - y\|.$$

The lemma states that the shortest path between two points is a straight line.

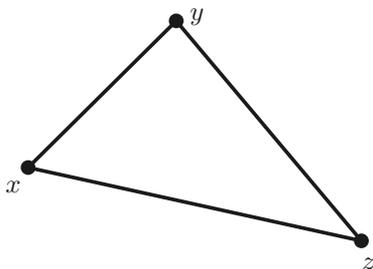


Figure 1: Triangle Inequality

Proof. Equivalently, $\|a + b\| \leq \|a\| + \|b\|$ for all $a, b \in \mathbb{R}^n$. This is because if we let $a = y - x$ and $b = z - y$, then $a + b = z - x$. Equivalently, squaring both sides, we have

$$\|a + b\|^2 \leq \|a\|^2 + \|b\|^2 + 2\|a\|\|b\|. \quad (1.1)$$

By definition of the norm, we have

$$\sum_{i=1}^n (a_i + b_i)^2 = \sum_{i=1}^n a_i^2 + b_i^2 + 2a_i b_i = \|a\|^2 + \|b\|^2 + 2a \cdot b.$$

Together with (1.1), we have that our statement is equivalent to $a \cdot b \leq \|a\|\|b\|$, which is the Cauchy-Schwarz inequality. \square

Lemma 1.3: Cauchy-Schwarz inequality

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

$$x \cdot y \leq \|x\|\|y\|.$$

Proof. If either $x = 0$ or $y = 0$, then the statement becomes $0 \leq 0$, which is true.

Let now $x, y \in \mathbb{R}^n$ be nonzero. Now for every $\lambda > 0$, we have

$$2x \cdot y = 2 \sum_{i=1}^n \lambda x_i \frac{y_i}{\lambda} \leq \sum_{i=1}^n \lambda^2 x_i^2 + \frac{y_i^2}{\lambda^2} = \lambda^2 \|x\|^2 + \frac{1}{\lambda^2} \|y\|^2,$$

Since $2ab \leq a^2 + b^2$. Since x, y are nonzero we can take $\lambda^2 = \frac{\|y\|}{\|x\|}$, which gives us

$$2x \cdot y \leq 2\|x\|\|y\|.$$

\square

In order to not always define convergence and continuity in other subjects, we will now introduce the notion of a **METRIC SPACE**, which is a set with a distance function defined on it.

Definition 1.4: Metric space

A metric space is a pair (X, d) , where X is a set and $d : X \times X \rightarrow [0, \infty)$ is a function such that for all $x, y, z \in X$ we have

- $d(x, y) = 0$ if and only if $x = y$,
- $d(x, y) = d(y, x)$, and
- $d(x, z) \leq d(x, y) + d(y, z)$.

Example 1.5:

1. $(\mathbb{R}^n, d_{\text{Euclidean}})$ is a metric space.
2. $(\mathbb{R}^2, d_{\text{NY}}(x, y))$ is a metric space, where the **MANHATTAN METRIC** is defined as

$$d_{\text{NY}}(x, y) = |x_1 - y_1| + |x_2 - y_2|.$$

3. Given (X, d) a metric space and $Y \subseteq X$, then $(Y, d|_{Y \times Y})$ is a metric space.

For example if we take $X = \mathbb{R}^2$ and Y to be the unit circle, then we also have the metric space $(Y, d_{\text{Euclidean}}|_{Y \times Y})$.

In general it is useful to first think about the proof in \mathbb{R}^n , and then to see if it can be adapted to a more general metric space. This is often easier as we tend to have a better intuition for \mathbb{R}^n .

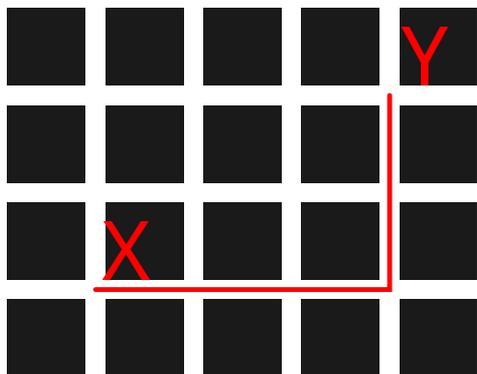


Figure 2: Manhattan (Glarus) metric visualized

Given $a < b \in \mathbb{R}$, and

$$X = \{f : [a, b] \rightarrow \mathbb{R}, f \text{ continuous}\},$$

then how can we find the distance between two functions $f, g \in X$? It can for example be defined as

$$d(f, g) = \max_{[a, b]} |f(x) - g(x)|.$$

Another distance would be

$$d(f, g) = \sqrt{\int_a^b |f(x) - g(x)|^2 dx}.$$

Lec 2

As an exercise, proof that both of these are metrics on X .

In the following, let (X, d) be a metric space. We want to define the notion of convergence and continuity in X .

Definition 1.6: Sequence

Given a set X , we call a map $x : \mathbb{N} \rightarrow X$ a **SEQUENCE** in X . Instead of writing $x(n)$, we write x_n as the n -th term of the sequence $\in X$. To denote the full sequence we write $(x_n)_{n \geq 0}$.

Definition 1.7: Convergent Sequence

We say $(x_n)_{n \geq 0}$ has limit $x \in X$ if $d(x_n, x) \rightarrow 0$ as real numbers. x is called the **LIMIT** of the sequence, and we write $\lim_{n \rightarrow \infty} x_n = x$.

Another way to phrase this is that for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have $d(x_n, x) < \varepsilon$.

Notice that the limit has to be in the set unlike in Analysis I, as otherwise the metric is not defined.

Lemma 1.8: Uniqueness of the limit

Given a sequence $(x_n)_{n \geq 0}$ such that $x_n \rightarrow x$ and $x_n \rightarrow y$, then $x = y$.

Proof. Assume by contradiction that $x \neq y$ and let $3\varepsilon = d(x, y) > 0$. By the definition of the limit, $\exists N_x$ such that $d(x_n, x) < \varepsilon \forall n \geq N_x$, and $\exists N_y$ such that $d(x_n, y) < \varepsilon \forall n \geq N_y$.

Let $N = \max(N_x, N_y)$, then for all $n \geq N$ we have

$$3\varepsilon = d(x, y) \leq d(x, x_n) + d(x_n, y) < 2\varepsilon,$$

by triangle inequality, which is a contradiction. \square

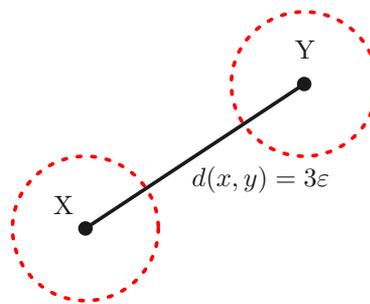


Figure 3: Uniqueness of the limit

Definition 1.9: Subsequence

Given a sequence $(x_n)_{n \geq 0}$, we define a **SUBSEQUENCE** as a sequence of the form

$$(x_{f(k)})_{k \geq 0},$$

where $f : \mathbb{N} \rightarrow \mathbb{N}$ is a strictly increasing function.

Usually, we will write x_{n_k} instead of $x_{f(k)}$ for the subsequence.

Definition 1.10: Accumulation Point

Given $Y \subset X$, we say that $y \in X$ is an **ACCUMULATION POINT** of Y if there exists a sequence $(y_n)_{n \geq 0}$ in Y such that $y_n \rightarrow y$.

Given $(x_n)_{n \geq 0}$ as a sequence of X we say that x is an **ACCUMULATION POINT** of $(x_n)_{n \geq 0}$ if there exists a subsequence $(x_{n_k})_{k \geq 0}$ such that $x_{n_k} \rightarrow x$.

Lemma 1.11:

Given a sequence $(x_n)_{n \geq 0}$. The sequence converges to x if and only if $\forall (x_{n_k})_{k \geq 0}$ we have $x_{n_k} \rightarrow x$.

Proof. We rewrite the statement to $x_n \rightarrow x \Leftrightarrow (x_{n_k}) \rightarrow y \Rightarrow y = x$. \Leftarrow : Taking the particular subsequence $x_{n_k} = x_n$, we have $x_n \rightarrow y$, and thus $y = x$.

\Rightarrow : Let $f : \mathbb{N} \rightarrow \mathbb{N}$ be a strictly increasing function. Since $x_n \rightarrow x$, for every $\varepsilon > 0$ there exists N such that for all $n \geq N$ we have $d(x_n, x) < \varepsilon$. Since $f(n)$ is increasing, $f(n) \geq n$, thus for all $n \geq N$ we have $d(x_{f(n)}, x) < \varepsilon$, which means that $x_{f(n)} \rightarrow x$. \square

Lemma 1.12:

Given a sequence $(x_n)_{n \geq 0}$. Then $x_n \rightarrow x$ if and only if every subsequence $(x_{n_k})_{k \geq 0}$ has a subsequence $(x_{n_{k_l}})_{l \geq 0}$ such that $x_{n_{k_l}} \rightarrow x$.

Proof. See later... We will do it later try first at home. \square

Definition 1.13: Cauchy Sequence

We say that a sequence $(x_n)_{n \geq 0}$ is a **CAUCHY SEQUENCE** if $\forall \varepsilon > 0, \exists N$ such that $d(x_n, x_m) < \varepsilon$ whenever $n, m \geq N$.

Definition 1.14: Complete Metric Space

We say that a metric space (X, d) is **COMPLETE** if every Cauchy sequence in X converges (to a limit in X).

Example 1.15:

1. $(\mathbb{R}, d_{\text{Euclidean}})$ is a complete metric space.

Theorem 1.16:

$(\mathbb{R}^n, d_{\text{Euclidean}})$ is complete.

Lemma 1.17:

Given a sequence $(x_m)_{m \geq 0} \subset \mathbb{R}^n$, then $x_m \rightarrow x \in \mathbb{R}^n$ if and only if $x_{m,i} \rightarrow x_i$ for all $i = 1, \dots, n$, where $x_m = (x_{m,1}, \dots, x_{m,n})$ and $x = (x_1, \dots, x_n)$.

Proof. \Rightarrow : Since $x_m \rightarrow x$, $\forall \varepsilon > 0 \exists N$ such that $\|x_n - x\| < \varepsilon$ for all $n \geq N$. In particular,

$$\|x_{m,i} - x_i\| \leq \sqrt{\sum_{j=1}^n |x_{m,j} - x_j|^2} = \|x_m - x\| < \varepsilon.$$

\Leftarrow : Since $x_{m,i} \rightarrow x_i$ for all $i = 1, \dots, n$, $\forall \varepsilon > 0 \exists N_i$ such that $|x_{m,i} - x_i| < \frac{\varepsilon}{\sqrt{n}}$ for all $m \geq N_i$. Let $N = \max(N_1, \dots, N_n)$, then for all $m \geq N$ we have

$$\|x_m - x\| = \sqrt{\sum_{i=1}^n |x_{m,i} - x_i|^2} < \sqrt{\sum_{i=1}^n \frac{\varepsilon^2}{n}} = \varepsilon.$$

□

Proof. [Proof of Theorem 1.16] Given $(x_m)_{m \geq 0}$ Cauchy, let us show $\exists x$ such that $x_m \rightarrow x$.

By Lemma 1.17, it suffices to show that $\forall i = 1, \dots, n$ we have that $(x_{m,i})_{m \geq 0}$ is Cauchy in \mathbb{R} .

But since Cauchy sequences in \mathbb{R} converge, $\exists x_i$ such that $x_{m,i} \rightarrow x_i$ for all $i = 1, \dots, n$. Thus by Lemma 1.17, we have $x_m \rightarrow x$, where $x = (x_1, \dots, x_n)$.

□

Tip 1.18:

An often seen counter example is the discrete metric space, where $d(x, y) = 1$ if $x \neq y$ and 0 otherwise.

Lec 3

1.1 Topology of metric spaces

We would like to define the notion of open sets in a metric space, which will allow us to define continuity and other topological properties of metric spaces.

Definition 1.19: Open Ball

Define the **OPEN BALL** of radius $r > 0$ centered at $x \in X$ as

$$B_r(x) = B(x, r) := \{y \in X : d(x, y) < r\}.$$

Definition 1.20: Open and Closed Sets

We say that $U \subset X$ is **OPEN** if for every $x \in U$, $\exists r > 0$ such that $B(x, r) \subset U$.

$A \subset X$ is **CLOSED**, if $X \setminus A$ is open.

Exercise 1.21:

Show that $(0, 1) \times (0, 1) \subset \mathbb{R}^2$ is open.

Show that $[0, 1] \times [0, 1] \subset \mathbb{R}^2$ is closed.

The topology associated to d is

$$\mathcal{T} = \{U \subset X : U \text{ is open}\}.$$

Lemma 1.22:

Arbitrary unions of open sets are open.

$$\{U_i\}_{i \in I}, U_i \subset X \text{ open} \Rightarrow \bigcup_{i \in I} U_i \text{ is open.}$$

Finite intersections of open sets are open.

$$U_1, \dots, U_k \subset X \text{ open} \Rightarrow \bigcap_{i=1}^k U_i \text{ is open.}$$

Proof. Take $x \in \bigcup_{i \in I} U_i$, then $x \in U_j$ for some $j \in I$. Since U_j is open, $\exists r > 0$ such that $B(x, r) \subset U_j \subset \bigcup_{i \in I} U_i$, thus $\bigcup_{i \in I} U_i$ is open.

Take $x \in \bigcap_{i=1}^k U_i$, then $x \in U_i$ for all $i = 1, \dots, k$. Since U_i is open, $\exists r_i > 0$ such that $B(x, r_i) \subset U_i$ for all $i = 1, \dots, k$. Let $r = \min(r_1, \dots, r_k)$, then $B(x, r) \subset B(x, r_i) \subset U_i$ for all $i = 1, \dots, k$, thus $\bigcap_{i=1}^k U_i$ is open. □

Lemma 1.23:

Finite unions of closed sets are closed.

Arbitrary intersections of closed sets are closed.

Proof. Notice that $X \setminus \bigcup_{i=1}^k A_i = \bigcap_{i=1}^k (X \setminus A_i)$, and $X \setminus A_i$ is open for all $i = 1, \dots, k$, thus $\bigcup_{i=1}^k A_i$ is closed. □

Example 1.24: Finite is Important

Take $(\mathbb{R}, d_{\text{Euclidean}})$ and $U_k = (-\frac{1}{k}, \frac{1}{k})$. Then

$$\bigcap_{k=1}^{\infty} U_k = \{0\}.$$

Which is not open.

Definition 1.25: Interior, Closure and Boundary

Given $\Omega \subset X$, we define

- $\text{int}\Omega = \Omega^\circ = \{U \subset \Omega \mid U \text{ is open}\}$ as the **INTERIOR** of Ω ,
- $\bar{\Omega} = \{x \in X \mid \exists (x_n)_{n \geq 0} \subset \Omega, x_n \rightarrow x\}$ as the **CLOSURE** of Ω , and
- $\partial\Omega = \bar{\Omega} \setminus \Omega^\circ$ as the **BOUNDARY** of Ω .

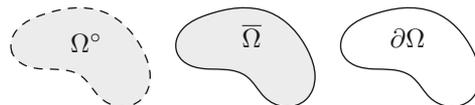


Figure 4: Interior, closure and boundary of a set Ω

Lemma 1.26:

$U \subset X$ is open if and only if whenever $(x_n)_{n \geq 0} \subset U$ such that $x_n \rightarrow x \in U$, then $x_n \in U$ eventually.

$A \subset X$ is closed if and only if whenever $(x_n)_{n \geq 0} \subset A$ such that $x_n \rightarrow x$, then $x \in A$.

Proof. 1, \Rightarrow : Take $x \in U$. By definition of open set, $\exists r > 0$ such that $B(x, r) \subset U$. Since $x_n \rightarrow x$, $\exists N$ such that $x_n \in B(x, r) \forall n \geq N$. (Since this is equivalent to $d(x_n, x) < r$ for all $n \geq N$). Thus $x_n \in U$ eventually.

1, \Leftarrow : We argue by contraposition. Since U is not open, $\exists x \in U$ such that $\forall r > 0, B(x, r) \not\subset U$. This is the same as saying $\exists x_r \in B(x, r) \cap X \setminus U$.

Taking $r = \frac{1}{n}$ and calling $x_n = x_{\frac{1}{n}} \rightarrow x$. Since $x_n \in X \setminus U$ for all $n \geq 0$, we have that $x_n \rightarrow x \in U$ but $x_n \notin U$ for all $n \geq 0$, contradicting that $x_n \in U$ eventually.

2. Exercise □

Exercise 1.27:

Given (X, d) complete, $A \subset X$ closed, show that (A, d) is complete.

Definition 1.28: Continuity

Given $f : X \rightarrow Y$ with (X, d_X) and (Y, d_Y) metric spaces, we say that f is **CONTINUOUS** if one of the following 3 equivalent properties hold:

- Epsilon-Delta Continuity: $\forall x \in X, \forall \varepsilon > 0, \exists \delta > 0$ such that:

$$\forall x' \in X, d_X(x, x') < \delta \Rightarrow d_Y(f(x), f(x')) < \varepsilon.$$

Which is equivalent to saying that

$$f(B(x, \delta)) \subset B(f(x), \varepsilon).$$

- Sequential Continuity: $(x_n)_{n \geq 0} \subset X$,

$$x_n \rightarrow x \Rightarrow f(x_n) \rightarrow f(x).$$

- topological continuity: $\forall U \subset Y$ open, $f^{-1}(U)$ is open.

Proposition 1.29:

The three definitions of continuity are equivalent.

Proof. 1 \Rightarrow 2: Assume $f : X \rightarrow Y$ is continuous in the epsilon-delta sense. Let $(x_n)_{n \geq 0} \subset X$ such that $x_n \rightarrow x$. Given $\varepsilon > 0$, by continuity, $\exists \delta > 0$ such that $f(B(x, \delta)) \subset B(f(x), \varepsilon)$. Since $x_n \rightarrow x$, $\exists N$ such that $x_n \in B(x, \delta)$ for all $n \geq N$. Thus $f(x_n) \in B(f(x), \varepsilon)$ for all $n \geq N$, which means that $f(x_n) \rightarrow f(x)$.

2 \Rightarrow 3: Assume f is not topologically continuous. Then $\exists U \subset Y$ open such that $f^{-1}(U)$ is not open. Hence $\exists x \in f^{-1}(U)$ and a sequence $(x_n)_{n \geq 0} \subset X \setminus f^{-1}(U)$ such that $x_n \rightarrow x$. Then $f(x) \in U$ but $f(x_n) \notin U$ for all n . But U is open so $f(x_n) \rightarrow f(x)$, contradicting sequential continuity.

3 \Rightarrow 1: Let $x \in X$ and $\varepsilon > 0$. The preimage $f^{-1}(B(f(x), \varepsilon))$ contains the point x and is open as $B(f(x), \varepsilon)$ is open and f is topologically continuous. Thus, $\exists \delta > 0$ such that $B(x, \delta) \subset f^{-1}(B(f(x), \varepsilon))$. Hence, f is $\varepsilon - \delta$ continuous. □

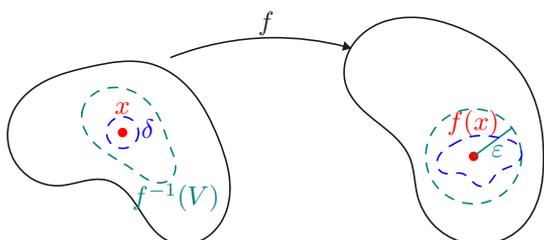


Figure 5: Topological continuity implies $\varepsilon - \delta$ continuity

Definition 1.30: Uniform and Lipschitz Continuity

Let (X, d_X) and (Y, d_Y) be metric spaces. We say that $f : X \rightarrow Y$ is **UNIFORMLY CONTINUOUS** if $\forall \varepsilon > 0, \exists \delta > 0$ such that

$$d_X(x, x') < \delta \Rightarrow d_Y(f(x), f(x')) < \varepsilon.$$

We say that f is **LIPSCHITZ CONTINUOUS** if $\exists L > 0$ such that

$$d_Y(f(x), f(x')) \leq L d_X(x, x') \forall x, x' \in X.$$

The constant L is called the **LIPSCHITZ CONSTANT** of f .

Example 1.31:

Fix $x_0 \in X$ and define $f := d(x, x_0)$. Then f is 1-Lipschitz.

Proof. Notice that $Y = [0, \infty)$ with the Euclidean metric. Then, for all $x, y \in X$ we have

$$d(f(x), f(y)) = |f(x) - f(y)| = |d(x, x_0) - d(y, x_0)| \leq d(x, y).$$

□

Theorem 1.32: Banach Fixed Point Theorem

Let (X, d) be a complete metric space and $T : X \rightarrow X$ λ -Lipschitz with $\lambda \in (0, 1)$ (sometimes called a **CONTRACTION**). Then, T has a unique **FIXED POINT** ($\exists! x \in X$ such that $T(x) = x$).

Proof. Fix $x_0 \in X$ and define $x_1 = T(x_0), x_2 = T(x_1), \dots, x_n = T(x_{n-1}), \dots$. We will show that $(x_n)_{n \geq 0}$ is a Cauchy sequence, and thus converges to some $x \in X$.

Since T is a contraction,

$$\begin{aligned} d(x_{n+1}, x_n) &= d(T(x_n), T(x_{n-1})) \\ &\leq \lambda \cdot d(x_n, x_{n-1}) \leq \dots \\ &\leq \lambda^n \cdot d(x_1, x_0). \end{aligned}$$

Since the distance is symmetric, w.l.o.g assume $m < n$. Then, by triangle inequality,

$$\begin{aligned} d(x_n, x_m) &\leq \sum_{k=m}^{n-1} d(x_{k+1}, x_k) \\ &\leq \sum_{k=m}^{n-1} \lambda^k \cdot d(x_1, x_0) \\ &\leq d(x_1, x_0) \cdot \lambda^m \cdot \frac{1}{1 - \lambda}. \end{aligned}$$

All terms besides λ^m are constants, and $\lambda^m \rightarrow 0$ as $m \rightarrow \infty$, thus $(x_n)_{n \geq 0}$ is a Cauchy sequence.

Since X is complete, $\exists x \in X$ such that $x = \lim_{n \rightarrow \infty} x_n$. Since T is continuous,

$$T(x) = \lim_{n \rightarrow \infty} T(x_n) = \lim_{n \rightarrow \infty} x_{n+1} = x.$$

So x is indeed a fixed point.

For uniqueness, suppose x, y are two fixed points. Then,

$$d(x, y) = d(T(x), T(y)) \leq \lambda \cdot d(x, y) < d(x, y).$$

□

We now like to extend our definition for compactness from Analysis I to metric spaces. To that extent, we need the following definition.

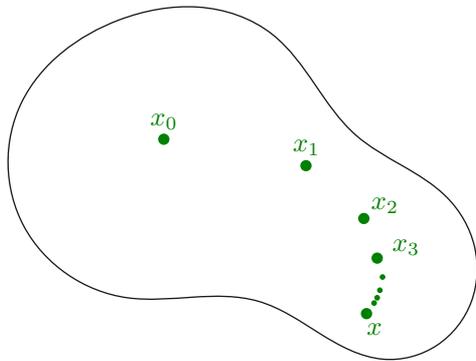


Figure 6: Proof of the Banach Fixed Point Theorem

Definition 1.33: Cover

Given a set X and $E \subset X$, we say that $\mathcal{U} = \{U_i, i \in I\}$ is a **COVER** of E if $E \subset \bigcup \mathcal{U} = \bigcup_{i \in I} U_i$.

If $V \subset \mathcal{U}$ is still a cover of E , we call V a **SUBCOVER**.

If \mathcal{U} is a collection of open sets, we call it an **OPEN COVER**.

Definition 1.34: Compactness

A set $K \subset X$ is called

(1) **SEQUENTIALLY COMPACT** if for every sequence $(x_n)_{n \geq 0} \subset K$, there exists a convergent subsequence $(x_{n_k})_{k \geq 0}$ such that $x_{n_k} \rightarrow x \in K$.

(2) **TOPOLOGICALLY COMPACT** if for every open cover \mathcal{U} of K , there exists a finite subcover $V \subset \mathcal{U}$ of K .

Example 1.35: Analysis I

Bolzano-Weierstrass: Any closed interval $[a, b] \subset \mathbb{R}$ is sequentially compact.

$[0, 1] \cap \mathbb{Q}$ is not sequentially compact. It is also not topologically compact, as $\mathbb{Q} \subset \{x_n | n \geq 0\}$ and we consider the open balls $\mathcal{U} = \{B(x_n, 2^{-n-10^3})\}$ then the total length of the intervals is $4 \cdot 2^{-10^3}$, which is less than 1, thus we cannot cover $[0, 1] \cap \mathbb{Q}$ with a finite number of intervals.

Proposition 1.36:

The two definitions of compactness are equivalent.

We will now show that (1) \Rightarrow (2), and we will show the converse later.

Proof. Assume $K \subset X$ is sequentially compact. Let $\{U_i\}$ be an open cover. Thus $\forall x \in K, \exists U_i$ open, such that $x \in U_i$. Given $x \in K$, let

$$r(x) = \min\{\sup\{r > 0 : B(x, r) \subset U_i \in \mathcal{U}\}, 1\}.$$

Given $x \in K$, select U_i , such that $B(x, \frac{r(x)}{2}) \subset U_i$ ¹

Pick any $x_0 \in K$ and define

$$\mathcal{V} := \{\underbrace{U_i(x_0)}_{=U_0}, \underbrace{U_i(x_1)}_{=U_1}, \dots\},$$

where $x_1 \in K \setminus U_0, x_2 \in K \setminus (U_0 \cup U_1)$, and so on.

¹We select $\frac{r}{2}$ in case the Supremum is not attained.

Doing so, unless I find a finite subcover, I will produce a sequence $x_n \in K \setminus \bigcup_{k=0}^{n-1} U_i(x_k)$. By sequential compactness, this sequence has a subsequence x_{n_l} such that $x_{n_l} \rightarrow x \in K$ with $r(x) > 0$. By construction $r(x_n) \rightarrow 0$! But $B_{\frac{r(x)}{2}} \subset U_i(x)$. Thus, for l large enough, $x_{n_l} \in U_i(x)$, contradicting the construction of the sequence. \square

For the converse, the proof is a bit shorter.

Proof. Given $(x_n)_{n \geq 0} \subset K$, we want to show that there exists a subsequence $(x_{n_k})_{k \geq 0}$ such that $x_{n_k} \rightarrow x \in K$.

Assume by contradiction, $\forall x \in K, x$ is not an accumulation point of $(x_n)_{n \geq 0}$ so $\forall x \in K, \exists \epsilon(x) > 0$ such that (x_n) visits $B(x, \epsilon(x))$ only finitely many times. Thus, $x_n \in K \setminus B(x, \epsilon(x)) \forall n \geq N(x)$.

Define now

$$\mathcal{U} = \{B(x, \epsilon(x)) | x \in K\}.$$

Since K is topologically compact,

$$K \subset \bigcup_{i=1}^N B(x_i, \epsilon(x_i)).$$

This would imply that our sequence only has finitely many terms, contradicting the fact that it is a sequence. \square

Corollary 1.37:

If $K \subset X$ is compact, then

1. K is closed,
2. K is complete,
3. If $A \subset X$ is closed, $K \cap A$ is compact.

Proposition 1.38:

If $f : X \rightarrow Y$ is continuous and $K \subset X$ is compact, then $f(K)$ is compact.

Proof. The goal is to show that $f(K)$ is topologically compact. Given \mathcal{V} as an open cover of $f(K)$, we have that $\mathcal{U} = \{f^{-1}(V) | V \in \mathcal{V}\}$ is an open cover of K . Since K is compact, there exists a finite subcover $f^{-1}(V_1), \dots, f^{-1}(V_n)$ of K . Thus V_1, \dots, V_n is a finite subcover of $f(K)$, and thus $f(K)$ is compact. \square

Theorem 1.39:

Given $f : X \rightarrow \mathbb{R}$ continuous, $K \subset X$ compact, such that $\sup\{f(x) | x \in K\}$ and $\inf\{f(x) | x \in K\}$ are finite, then $\exists t \in K$ such that $f(t) = \sup\{f(x) | x \in K\}$

Proof. By definition of the supremum, $\exists (x_n)_{n \geq 0} \subset K$ such that $f(x_n) \rightarrow \sup\{f(x) | x \in K\}$ as $n \rightarrow \infty$. Since K is compact, $\exists (x_{n_k})_{k \geq 0}$ such that $x_{n_k} \rightarrow t \in K$. By continuity of $f, f(x_{n_k}) \rightarrow f(t)$ as $k \rightarrow \infty$. Thus, $f(t) = \sup\{f(x) | x \in K\}$. \square

Since in this course, we will mostly work in \mathbb{R}^n , we will now show the following theorem.

Theorem 1.40: Heine-Borel

$K \subset \mathbb{R}^n$ is compact if and only if K is closed and bounded.

Note

We call a set $K \subset \mathbb{R}^n$ **BOUNDED** if $\exists M > 0$ such that

$$B(0, M) \supset K.$$

Proof. \Rightarrow : K is closed by corollary 1.37.

If K were unbounded, then $\exists(x_N)_{N \geq 0} \subset K$ such that $\|x_N\| \geq N \forall N$. Any subsequence of $(x_N)_{N \geq 0}$ is also unbounded. Indeed, by triangle inequality

$$x_{N_k} \rightarrow x \Leftrightarrow d(x_{N_k}, x) \rightarrow 0 \Rightarrow \|x_{N_k}\| < \|x\| + \|x_{N_k} - x\| \rightarrow \|x\|.$$

Thus, K is not sequentially compact, contradicting the fact that K is compact.

\Leftarrow : Our goal will be to show, that given $N \in \mathbb{N}$, $[-N, N]^n \subset \mathbb{R}^n$ is compact. This is sufficient, since if $K \subset \mathbb{R}^n$ is closed and bounded, then $K \subset B(0, N) \subset [-N, N]^n$ for some $N \in \mathbb{N}$, and thus K is a closed subset of a compact set, and thus compact by corollary 1.37.

We want to reduce the problem to Bolzano-Weierstrass in \mathbb{R} . Given $(x_k) \subset [-N, N]^n$, we can write $x_k = (x_{k,1}, \dots, x_{k,n})$ where $x_{k,i} \in [-N, N]$.

1. Look at the sequence $(x_{k,1})_{k \geq 0} \subset [-N, N]$. By Bolzano-Weierstrass², there exists a increasing sequence $(k_m^{(1)})_{m \geq 0}$ in \mathbb{N} such that

$$x_{k_m^{(1)},1} \rightarrow x_1 \in [-N, N].$$

2. Look at the sequence $(x_{k_m^{(1)},2})_{m \geq 0} \subset [-N, N]$. By Bolzano-Weierstrass, there exists a increasing sequence $(k_m^{(2)})_{m \geq 0}$ in \mathbb{N} such that

$$x_{k_m^{(2)},2} \rightarrow x_2 \in [-N, N].$$

Since $(k_m^{(2)})_{m \geq 0}$ is a subsequence of $(k_m^{(1)})_{m \geq 0}$, we also have

$$x_{k_m^{(2)},1} \rightarrow x_1 \in [-N, N].$$

We continue this process, in each step making one new index converge, and keeping the previous ones converging. After n steps, we have an increasing sequence $(k_m^{(n)})_{m \geq 0}$ in \mathbb{N} such that

$$x_{k_m^{(n)},i} \rightarrow x_i \in [-N, N] \quad \forall i = 1, \dots, n.$$

Let $x = (x_1, \dots, x_n)$, then by Lemma 1.17,

$$x_{k_m^{(n)}} \rightarrow x \in [-N, N]^n.$$

So K is sequentially compact. \square

Tip 1.41:

The idea to split the proof into n times applying an argument from \mathbb{R} is a very typical idea in Analysis II.

We now want to talk about **CONNECTEDNESS**. As a preliminary, in any metric space, X, \emptyset are always open and closed (sometimes called clopen).

Question

Given $X = \mathcal{S}^2 = \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$, equipped with the Euclidean metric restricted to \mathcal{S}^2 .

Is it possible to write $X = U \cup V$ such that $U \cap V = \emptyset$ and U, V are open in X and non-empty?

Definition 1.42: Connectedness

Given a metric space (X, d) , we say that $A \subset X$ (nonempty) is **DISCONNECTED** if $\exists U, V$ open, disjoint, such that $A \subset U \cup V$ and $A \cap U, A \cap V$ are non-empty.

We say that A is **CONNECTED** if it is not disconnected.

In other words, disconnected means that we have an open cover of A with two disjoint sets.

Tip 1.43:

Working with connectedness is usually done by contraposition, since the definition is an existence statement.

Proposition 1.44: Connected subsets of \mathbb{R}

$E \subset \mathbb{R}$ is connected if and only if E is an interval.

Notice that E is an interval if and only if

$$\forall x < y \in E, [x, y] \subset E.$$

Proof. \Rightarrow : By contraposition, assume E is not an interval. Then, $\exists x < y \in E$ such that $[x, y] \not\subset E$. So $\exists z \in [x, y]$ such that $z \notin E$.

Define $U = (-\infty, z)$ and $V = (z, \infty)$. Then, U, V are open, disjoint, and $E \subset U \cup V$. Moreover, $E \cap U$ and $E \cap V$ are non-empty since $x \in E \cap U$ and $y \in E \cap V$. Thus, E is disconnected.

\Leftarrow : By contraposition, assume E is disconnected. Then, $\exists U, V$ open, disjoint, such that $E \subset U \cup V$ and $E \cap U, E \cap V$ are non-empty.

Thus, pick $x \in E \cap U$ and $y \in E \cap V$ such that $x < y$ ³. Consider the supremum of the following set:

$$t^* = \sup\{t \geq x \mid [x, t] \subset U\}.$$

This supremum is well defined since $t^* < y$ because $[x, y] \not\subset U$.

From this we also find that $t^* \in \mathbb{R} \setminus V$ since V is open and disjoint from U . But also $t^* \in \mathbb{R} \setminus U$ since if $t^* \in U$, then by openness of U , $\exists \varepsilon > 0$ such that $B(t^*, \varepsilon) \subset U$. Thus, $[x, t^* + \varepsilon] \subset U$, contradicting the definition of t^* .

So $t^* \in \mathbb{R} \setminus (U \cup V) \subset \mathbb{R} \setminus E$, and thus $[x, y] \not\subset E$. So E is not an interval. \square

Proposition 1.45:

Given $(X, d_X), (Y, d_Y)$ metric spaces, $f : X \rightarrow Y$ continuous and $E \subset X$ connected, then $f(E)$ is connected.

Proof. By contraposition, assume $f(E)$ is disconnected. Then, $\exists V_1, V_2$ open, disjoint, such that $f(E) \subset V_1 \cup V_2$ and $f(E) \cap V_1, f(E) \cap V_2$ are non-empty.

Define now $U_i = f^{-1}(V_i)$, which are open since f is continuous. Then, U_1, U_2 are open, disjoint, and $E \subset U_1 \cup U_2$. Also, they are non-empty since $f(E) \cap V_i$ are non-empty. Thus, E is disconnected. \square

Corollary 1.46: Intermediate Value Theorem

Given a connected metric space (X, d) and $f : X \rightarrow \mathbb{R}$ continuous, such that $f(x) = a \leq f(y) = b$. Then $\exists t \in X$ such that $f(t) = c$ for all $c \in [a, b]$.

Proof. Skipped...

Since X is connected, $f(X)$ is connected, and thus an interval. Since $a, b \in f(X)$, we have $[a, b] \subset f(X)$, and thus $\exists t \in X$ such that $f(t) = c$. \square

Definition 1.47: Curve

Given a metric space (X, d) , a **CURVE** or **PATH** in X is a map $\gamma : [0, 1] \rightarrow X$ which is continuous.

$\gamma(0)$ is called the **STARTING POINT** of γ , and $\gamma(1)$ is called the **ENDPOINT** of γ .

γ is called **CLOSED** or **LOOP** if $\gamma(0) = \gamma(1)$.

²Any sequence in a compact interval has a convergent subsequence

³Can be assumed w.l.o.g.

Definition 1.48: Path Connectedness

Given a metric space (X, d) , we call $E \subset X$ path connected if $\forall x, y \in E, \exists$ a path joining x and y , i.e. $\exists \gamma : [0, 1] \rightarrow E$ continuous such that $\gamma(0) = x$ and $\gamma(1) = y$.

Proposition 1.49:

Given (X, d) , E path connected, then E is connected.

Proof. Assume by contraposition that E is disconnected. Then, $\exists U_1, U_2$ open, disjoint, such that $E \subset U_1 \cup U_2$ and $\exists x_i \in E \cap U_i$.

Assume by contradiction, $\exists \gamma : [0, 1] \rightarrow E$ continuous such that $\gamma(0) = x_1$ and $\gamma(1) = x_2$. A let $V_i = \gamma^{-1}(U_i)$, which are open since γ is continuous and disjoint. So $[0, 1]$ is disconnected, contradicting the fact that $[0, 1]$ is connected. \square

The converse is not true, as the following example shows.

Example 1.50: Topologist's Sine Curve

Consider $(\mathbb{R}^2, d_{\text{Eucl}})$ and the set

$$E = \{0\} \times [-1, 1] \cup \{(t, \sin(\frac{1}{t})) \mid t > 0\}.$$

This set is connected. However it is not path connected since intuitively, if we want to connect $(0, 0)$ to $(1, \sin(1))$, we need to go through infinitely many oscillations of the sine curve, which is not possible with a continuous path.

Theorem 1.51:

In $(\mathbb{R}^n, d_{\text{Eucl}})$, given $U \subset \mathbb{R}^n$ open. U is connected if and only if U is path connected.

We first like to define the composition of paths. Given $\gamma_1 : [0, 1] \rightarrow X$ and $\gamma_2 : [0, 1] \rightarrow X$ such that $\gamma_1(0) = \gamma_2(0)$, we can define $\gamma_1^* = \gamma_1(1 - t)$ as the reverse of γ_1 . Then, the composition is defined as

$$\gamma_3(t) = \begin{cases} \gamma_1^*(2t) & t \in [0, \frac{1}{2}] \\ \gamma_2(2t - 1) & t \in [\frac{1}{2}, 1] \end{cases}.$$

Proof. \Leftarrow : By Proposition 1.49.

\Rightarrow : We want to show that we can pick $x_0 \in U$ and join it with any other point $x \in U$ with a path. This is enough since if we can join x_0 to x and x_0 to y , then we can compose the paths to join x and y .

Define a set $G \subset U$ as

$$G := \{x \in U \mid \exists \text{ path } \gamma : [0, 1] \rightarrow U : \gamma(0) = x_0, \gamma(1) = x\}.$$

We will now proof that G is open and that $U \setminus G$ is open, which since $x_0 \in G$ and U is connected, will imply that $G = U$.

The key observation is that for every $x \in U$, $\exists B_r(x) \subset U$ for some $r > 0$ since U is open. So $y \in B_r(x) \in G$ if and only if $x \in G$. This is because the map $t \in [0, 1] \mapsto \gamma(t) = (1 - t)x + ty$ is a path joining x and y .

This immediately implies that G is open, since if $x \in G$, then $B_r(x) \subset G$.

Moreover, if $x \in U \setminus G$, then $B_r(x) \subset U \setminus G$ since if $y \in B_r(x)$, then $x \in G$ if and only if $y \in G$. Thus, $U \setminus G$ is open. \square

Proposition 1.52: Continuity on compact sets

Suppose $(X, d_X), (Y, d_Y)$ are metric spaces, $f : X \rightarrow Y$ is continuous, and $K \subset X$ is compact. Then, $f|_K : K \rightarrow Y$ is uniformly continuous.

Proof. Let $\varepsilon > 0$. By usual continuity of f , $\forall x \in K, \exists \delta_x > 0$ such that

$$f(B(x, \delta_x)) \subset B\left(f(x), \frac{\varepsilon}{2}\right).$$

Consider the open cover $\mathcal{U} = \{B(x, \frac{\delta_x}{2}) \mid x \in K\}$ of K . Since K is compact, there exists a finite subcover $B(x_1, \frac{\delta_{x_1}}{2}), \dots, B(x_n, \frac{\delta_{x_n}}{2})$ of K . Let $\delta = \min\{\frac{\delta_{x_i}}{2} \mid i = 1, \dots, n\}$. We will show that this δ works for uniform continuity.

If $x \in K$ and $y \in B(x, \delta)$, then

$$x \in B(x_i, \frac{\delta_{x_i}}{2}) \text{ for some } i \in \{1, \dots, n\}.$$

So

$$d(x, y) < \delta \leq \frac{\delta_{x_i}}{2} \Rightarrow d(x_i, y) < \delta(x_i).$$

But then

$$f(B(x, \delta)) \subset f(B(x_i, \delta_{x_i})) \subset B\left(f(x_i), \frac{\varepsilon}{2}\right) \subset B(f(x), \varepsilon).$$

\square

1.2 Normed Vector Spaces

Lec 7

Before starting with differentiating functions of several variables, we need to add some structure to our metric spaces.

Definition 1.53: Normed Vector Space over \mathbb{R}

Let V be a VS over \mathbb{R} . A map $\|\cdot\| : V \rightarrow [0, \infty)$ is called a **NORM** if $\forall v, u \in V$ and $\alpha \in \mathbb{R}$, the following hold:

1. Definite: $\|v\| = 0$ if and only if $v = 0$.
2. Homogeneous: $\|\alpha v\| = |\alpha| \cdot \|v\|$
3. Triangle inequality: $\|v + u\| \leq \|v\| + \|u\|$.

Example 1.54: Examples on \mathbb{R}^n

\mathbb{R}^n with the Euclidean norm $\|x\|_{\text{Eucl}} = \sqrt{\sum_{i=1}^n x_i^2}$ is a normed vector space.

\mathbb{R}^n with the p -norm $\|x\|_p = (\sum |x_i|^p)^{\frac{1}{p}}$ for $p \geq 1$ is a normed vector space.

\mathbb{R}^n with the ∞ -norm $\|x\|_{\infty} = \max |x_i|$ is a normed vector space.

From now on, we will write $|\cdot|$ instead of $\|\cdot\|$ when we use the Euclidean norm.

Definition 1.55: Hilbert-Schmidt norm

Let $M_{m \times n}(\mathbb{R})$ be the set of $m \times n$ matrices with real entries. We define the **HILBERT-SCHMIDT NORM** on $M_{m \times n}(\mathbb{R})$ as

$$\|M\|_2 = \sqrt{\text{tr}(M^T M)} = \sqrt{\sum_{i=1}^n |Me_i|^2}.$$

To see the last equality, we square the Hilbert-Schmidt norm, we get

$$\|M\|_2^2 = \text{tr}(M^T M) = \sum_{i=1}^n \sum_{j=1}^m M_i^j M_i^j.$$

But also

$$[Me_i]^j = M_i^j \Rightarrow |Me_i|^2 = \sum_{j=1}^m M_i^j M_i^j.$$

Lemma 1.56:

For every $x \in \mathbb{R}^n$ and $M : \mathbb{R}^n \rightarrow \mathbb{R}^m$ linear,

$$|Mx| \leq \|M\|_2 \cdot |x|.$$

Proof. We can write $x = \sum_{i=1}^n x_i e_i$. Thus, we compute

$$\begin{aligned} |Mx|^2 &= \sum_{i=1}^m [(Mx)^i]^2 = |M(\sum_{i=1}^n x_i e_i)|^2 \\ &= \left| \sum_{i=1}^n x_i Me_i \right|^2 \\ &\leq \left(\sum_{i=1}^n |x_i Me_i| \right)^2 = \left(\sum_{i=1}^n |x_i| |Me_i| \right)^2 \\ &\leq |x|^2 \cdot \left(\sum_{i=1}^n |Me_i|^2 \right) = |x|^2 \cdot \|M\|_2^2. \end{aligned}$$

Example 1.57: Norms on function spaces

Given $V = \{f : [a, b] \rightarrow \mathbb{R} \mid f \text{ continuous}\}$ we have the L^p -NORM for $p \geq 1$

$$\|f\|_{L^p} := \left(\int_a^b |f(t)|^p dt \right)^{\frac{1}{p}}.$$

And the SUPREMUM NORM

$$\|f\|_{L^\infty} := \sup_{t \in [a, b]} |f(t)|.$$

Proposition 1.58: Norm implies metric

Every normed vector space is a metric space with the metric defined as

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

Proof. d is definite because $\|x - y\| = 0$ if and only if $x - y = 0$ if and only if $x = y$.

d is symmetric because $\|x - y\| = \|(-1)(y - x)\| = \|y - x\|$.

d satisfies the triangle inequality, because

$$\|x - z\| = \|(x - y) + (y - z)\| \leq \|x - y\| + \|y - z\|.$$

□

Definition 1.59: Inner Product Vector Space

Let V be a VS over \mathbb{R} . A **INNER PRODUCT** is a map $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ such that $\forall u, v, w \in V$ and $\alpha \in \mathbb{R}$:

1. Symmetry: $\langle u, v \rangle = \langle v, u \rangle$.
2. Bilinearity: $\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle$.
3. Definite: $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0 \Leftrightarrow v = 0$.

Lemma 1.60:

Let V be an inner product vector space. Then, $\|\cdot\| : V \rightarrow [0, \infty)$ defined as

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

is a norm on V . So every inner product vector space is a normed vector space.

Proof. Exercise Sheet and Linear Algebra: Uses Cauchy-Schwarz inequality,

$$\langle v, w \rangle \leq \sqrt{\langle v, v \rangle} \cdot \sqrt{\langle w, w \rangle}.$$

□

Going back to the beginning of the lecture, we now know that \mathbb{R}^n is an inner product vector space with the inner product defined as $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$, and thus a normed vector space and a metric space.

From now on, we will by default work with the Euclidean norm and metric on \mathbb{R}^n . The most central definitions are the limit of a function on an open set $U \subset \mathbb{R}^n$

□

$$y = \lim_{x \rightarrow x_0} f(x) \Leftrightarrow \forall (x_k) \subset U \text{ with } x_k \rightarrow x_0, f(x_k) \rightarrow y.$$

and the continuity of a function

$$\lim_{x \rightarrow x_0} f(x) = f(x_0).$$

One useful trick to compute a limit of a function $\mathbb{R}^2 \rightarrow \mathbb{R}$ as $r \rightarrow 0$ is to use polar coordinates. We can write $x = (r \cos \theta, r \sin \theta)$, and thus for example

$$\lim_{\rightarrow 0} \frac{xy}{\sqrt{x^2 + y^2}} = \lim_{r \rightarrow 0} \frac{r^2 \cos \theta \sin \theta}{r} = \lim_{r \rightarrow 0} r \cos \theta \sin \theta = 0.$$

2 Multidimensional Differentiation

We start with the most important definition of the course.

Definition 2.1: Differential

Let $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R}^m$. We say that f is **DIFFERENTIABLE** at $x_0 \in U$ if there exists a linear map $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that

$$\lim_{x \rightarrow x_0} \frac{f(x_0 + x) - f(x_0) - L(x)}{|x|} = 0.$$

L is called the **DIFFERENTIAL** of f at x_0 , and is denoted by Df_{x_0} or $Df(x_0)$.

Claim 2.2:

If $n = m = 1$, then $L(x) = f'(x_0)x$ where $f'(x_0)$ is the usual derivative of f at x_0 .

Solution. We know $f'(x_0) = \lim_{x \rightarrow 0} \frac{f(x_0+x) - f(x_0)}{x}$, so we can write

$$\begin{aligned} & \lim_{x \rightarrow 0} \frac{f(x_0 + x) - f(x_0) - f'(x_0)x}{|x|} \\ &= \lim_{x \rightarrow 0} \frac{f(x_0 + x) - f(x_0)}{x} - f'(x_0) \\ &= f'(x_0) - f'(x_0) = 0. \end{aligned}$$

For some intuition consider the following examples.

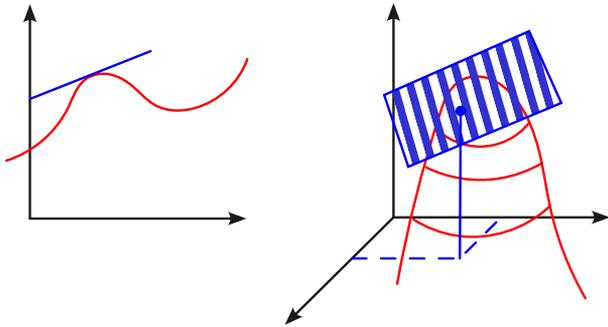


Figure 7: Differential of a function $\mathbb{R} \rightarrow \mathbb{R}$ and of a function $\mathbb{R}^2 \rightarrow \mathbb{R}$.

The equation for the slope in the first case would be

$$y = f(x_0) + f'(x_0)(x - x_0).$$

Whereas the equation for the tangent plane in the second case would be

$$y = f(x_0) + Df_{x_0}(x - x_0).$$

Lemma 2.3: Differential Componentwise

Let $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R}^m$ and $x_0 \in U$. Then f is differentiable at x_0 if and only if $f_i : U \rightarrow \mathbb{R}$ is differentiable at $x_0 \forall i = 1, \dots, m$, where f_i is the i -th component of f . Moreover,

$$Df_{x_0}(x) = \begin{pmatrix} Df_{1,x_0}(x) \\ \vdots \\ Df_{m,x_0}(x) \end{pmatrix}.$$

This lemma allows us to reduce proofs to the case $m = 1$, which is usually easier to work with.

Proof. Recall the definition of the derivative of f at x_0 :

$$\lim_{x \rightarrow 0} \frac{f(x_0 + x) - f(x_0) - L(x)}{|x|} = 0.$$

This can converge if and only if $\forall i = 1, \dots, m$,

$$\lim_{x \rightarrow 0} \frac{f_i(x_0 + x) - f_i(x_0) - L_i(x)}{|x|} = 0.$$

But this is equivalent to f_i being differentiable at x_0 with differential L_i for all $i = 1, \dots, m$. \square

A convenient notation is the following

Definition 2.4: Big and Little O

Given $f, g : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $x_0 \in U$, we say that

$$f(x) = O(g(x)) \text{ as } x \rightarrow x_0 \Leftrightarrow \lim_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} < \infty.$$

Similarly, we say that

$$f(x) = o(g(x)) \text{ as } x \rightarrow x_0 \Leftrightarrow \lim_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} = 0.$$

With this notation, f is differentiable at x_0 if and only if

$$R(x) = f(x_0 + x) - f(x_0) - Df_{x_0}(x) = o(|x|).$$

Definition 2.5: Directional derivatives

Let $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R}^m$ and $x_0 \in U$. Given $v \in \mathbb{R}^n$, we define the **DIRECTIONAL DERIVATIVE** of f at x_0 along v as

$$\partial_v f(x_0) = \lim_{t \rightarrow 0} \frac{f(x_0 + tv) - f(x_0)}{t},$$

provided the limit exists.

Proposition 2.6:

Given $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R}^m$ and $x_0 \in U$. f is differentiable at $x_0 \in U$ implies that $\partial_v f(x_0)$ exists for all $v \in \mathbb{R}^n$, and

$$\partial_v f(x_0) = Df_{x_0}(v).$$

Proof. We can write

$$f(x_0 + x) - f(x_0) - Df_{x_0}(x) = o(|x|) \text{ as } x \rightarrow 0.$$

Thus for any sequence going to 0, we have that

$$\frac{f(x_0 + x) - f(x_0) - Df_{x_0}(x)}{|x|} \rightarrow 0.$$

In particular, for the sequence $x = tv$ with $t \rightarrow 0$, we have

$$\frac{f(x_0 + tv) - f(x_0) - Df_{x_0}(tv)}{|tv|} \rightarrow 0.$$

Since Df_{x_0} is linear, we can write $Df_{x_0}(tv) = tDf_{x_0}(v)$, and since $|tv| = |t||v|$, with $|v|$ finite, we can write

$$\underbrace{\frac{f(x_0 + tv) - f(x_0)}{t}}_{=\partial_v f(x_0)} - Df_{x_0}(v) \rightarrow 0.$$

\square

Definition 2.7: Partial derivatives

Let $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R}^m$ and $x_0 \in U$. Given $i \in \{1, \dots, n\}$, we define the **PARTIAL DERIVATIVE** of f at x_0 along the i -th coordinate as

$$\partial_{e_i} f(x_0) =: \frac{\partial f}{\partial x_i}(x_0) = \partial_i f(x_0).$$

Notice that the partial derivatives are just the directional derivatives along the canonical basis vectors.

Example 2.8:

Let $n = 3, m = 1$ and $f(x_1, x_2, x_3) = \exp(x_2)[1 + x_1 x_3]$. Then, compute the partial derivative $\partial_1 f(0)$ and all partial derivatives at an arbitrary point $x = (x_1, x_2, x_3)$.

Solution. By definition

$$\begin{aligned} \partial_1 f(0, 0, 0) &= \lim_{t \rightarrow 0} \frac{f(0+t, 0, 0) - f(0, 0, 0)}{t} \\ &= \lim_{t \rightarrow 0} \frac{\exp(0)[1+t \cdot 0] - \exp(0)[1+0 \cdot 0]}{t} \\ &= \lim_{t \rightarrow 0} \frac{1-1}{t} = 0. \end{aligned}$$

For the general case, we can define $g(t) = f(x_1 + t, x_2, x_3)$, so that $\partial_1 f(x) = g'(0)$. We can compute

$$\begin{aligned} \partial_1 f(x_1, x_2, x_3) &= \exp(x_2)x_3 \\ \partial_2 f(x_1, x_2, x_3) &= \exp(x_2)(1 + x_1 x_3) \\ \partial_3 f(x_1, x_2, x_3) &= \exp(x_2)x_1. \end{aligned}$$

In principle, the partial derivative can be calculated by fixing all the coordinates except the one we want to differentiate with respect to, and then applying the usual derivative in one variable.

The trick can be extended to a directional derivative along an arbitrary direction v

$$\partial_v f(x_0) = g'(0) \text{ where } g(t) = f(x_0 + tv).$$

Given $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R}^m$. When $\partial_i f(x_0)$ exists for all $x_0 \in U$ then we define the function

$$\partial_i f : U \rightarrow \mathbb{R}^m, x \mapsto \partial_i f(x).$$

Theorem 2.9: Sufficient condition for differentiability

Let $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R}^m$. If $\partial_i f(x)$ exists and is continuous for every $i = 1, \dots, n$, and every $x \in U$, then f is differentiable at every $x \in U$.

Moreover,

$$Df_{x_0}(x) = (\partial_1 f(x_0), \dots, \partial_n f(x_0)) x.$$

More explicitly, if we write $x = (x_1, \dots, x_n)$, then

$$Df_{x_0}(x) = \begin{pmatrix} \partial_1 f_1(x_0) & \dots & \partial_n f_1(x_0) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(x_0) & \dots & \partial_n f_m(x_0) \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Proof. Fix $x_0 \in U$ and take $\varepsilon > 0$ such that $\{x \mid |x_i - x_{0,i}| < \varepsilon\} \subset U$.

Take $x \in \mathbb{R}^n$ and define $x^{(k)} = x_0 + \sum_{i=1}^k x_i e_i$, so that $x^{(0)} = x_0$ and $x^{(n)} = x_0 + x$. Then, we can write

$$\begin{aligned} f(x_0 + x) - f(x_0) &= \sum_{k=1}^n f(x^{(k)}) - f(x^{(k-1)}) \\ &= \sum_{k=1}^n \partial_k f(y^{(k)}) x_k. \end{aligned}$$

We want to show the last equality by using the mean value theorem on a function $g_k(t) = f(x^{(k-1)} + t e_k)$, so that $g_k(0) = f(x^{(k-1)})$ and $g_k(x_k) = f(x^{(k)})$.

By Lemma 2.3, we can assume $m = 1$ and thus $g_k : \mathbb{R} \rightarrow \mathbb{R}$, so that by the mean value theorem, $\exists \xi_k \in [x^{(k-1)}, x^{(k)}]$ such that

$$g_k(x_k) - g_k(0) = g'_k(\xi_k) x_k = \partial_k f(x^{(k-1)} + \xi_k e_k) x_k.$$

Letting $y^{(k)} = x^{(k-1)} + \xi_k e_k$, we get the desired equality.

From this, we can write

$$f(x_0 + x) - f(x_0) = \sum_{k=1}^n \partial_k f(y^{(k)}) x_k$$

But $\sum_{i=1}^{k-1} x_i e_i \rightarrow 0$ as $x \rightarrow 0$, and since $\xi_k \in [0, x_k]$, we have $\xi_k e_k \rightarrow 0$ as $x \rightarrow 0$. Thus, $y^{(k)} \rightarrow x_0$ as $x \rightarrow 0$.

$$\begin{aligned} f(x_0 + x) - f(x_0) &= \sum_{k=1}^n \partial_k f(y^{(k)}) x_k \\ &= \sum_{k=1}^n (\partial_k f(x_0) x_k + o(1)) \cdot |x| \end{aligned}$$

Thus, we can write

$$f(x_0 + x) - f(x_0) = \sum_{k=1}^n \partial_k f(x_0) x_k + o(|x|).$$

Which means that f is differentiable at x_0 with differential

$$Df_{x_0}(x) = (\partial_1 f(x_0), \dots, \partial_n f(x_0)) x.$$

□

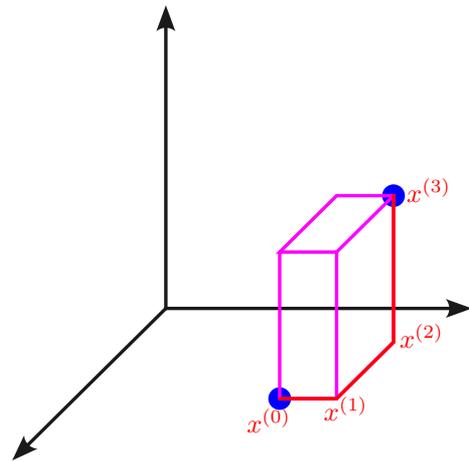


Figure 8: Definition of the $x^{(i)}$ s.

Lec 9

Definition 2.10: C^1 functions

Given $U \subset \mathbb{R}^n$ open, we define $C^1(U, \mathbb{R}^m)$ as

$$C^1(U, \mathbb{R}^m) := \{f : U \rightarrow \mathbb{R}^m : f \text{ cont. diff. in } U\}.$$

When $m = 1$, we write $C^1(U)$.

Definition 2.11: Jacobi Matrix

Given $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R}^m$, differentiable at x_0 . We define the **JACOBI MATRIX** of f at x_0 as

$$Df(x_0) = Jf(x_0) = (\partial_1 f(x_0), \dots, \partial_n f(x_0)).$$

Writing out the Jacobi matrix explicitly, we have

$$Jf(x_0) = \begin{pmatrix} \partial_1 f_1(x_0) & \dots & \partial_n f_1(x_0) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(x_0) & \dots & \partial_n f_m(x_0) \end{pmatrix}.$$

We observe, that if we take a vector $v = (v_1, \dots, v_n)^T = \sum_{i=1}^n v_i e_i$, then

$$Df_{x_0}(v) = \sum_{i=1}^n v_i Df_{x_0}(e_i) = \sum_{i=1}^n v_i \frac{\partial f}{\partial x_i} = Jf(x_0)v.$$

That is, $Jf(x_0)$ is the matrix of Df_{x_0} with respect to the canonical basis.

Theorem 2.12: Chain Rule

Given $U \subset \mathbb{R}^n$ open, $V \subset \mathbb{R}^m$ open and $f : U \rightarrow V$ and $g : V \rightarrow \mathbb{R}^k$. Assume that f is differentiable at $x_0 \in U$ and g is differentiable at $f(x_0) \in V$. Then, $g \circ f : U \rightarrow \mathbb{R}^k$ is differentiable at x_0 , and

$$D(g \circ f)_{x_0} = Dg_{f(x_0)} \circ Df_{x_0}.$$

The idea of the proof is that

$$f(x_0 + x) - f(x_0) \approx L(x) \text{ and } g(y_0 + y) - g(y_0) \approx M(y).$$

Using $y = f(x_0 + x) - f(x_0)$, we can write

$$\begin{aligned} g(y_0 + f(x_0 + x) - f(x_0)) - g(y_0) &\approx M(f(x_0 + x) - f(x_0)) \\ g(f(x_0 + x)) - g(f(x_0)) &\approx M(L(x)). \end{aligned}$$

Proof. We can write

$$\begin{aligned} f(x_0 + x) - f(x_0) &= L(x) + o(|x|) \\ g(y_0 + y) - g(y_0) &= M(y) + o(|y|). \end{aligned}$$

Then substituting $y = f(x_0 + x) - f(x_0)$, we get

$$\begin{aligned} g(f(x_0 + x)) - g(f(x_0)) &= g(y_0 + f(x_0 + x) - f(x_0)) - g(y_0) \\ &= M(f(x_0 + x) - f(x_0)) + o(|f(x_0 + x) - f(x_0)|) \\ &= M(L(x) + o(|x|)) + o(|L(x) + o(|x|)|) \\ &= M(L(x)) + o(|x|). \end{aligned}$$

Where we used the fact, that M and L are linear, so that $M(L(x))$ is linear in x , and thus $o(|L(x)|) = o(|x|)$, and $o(o(|x|)) = o(|x|)$. \square

As a practical consequence of this,

$$J(g \circ f)(x_0) = Jg(f(x_0)) \cdot Jf(x_0).$$

Component-wise this means that

$$\frac{\partial (g \circ f)_k}{\partial x_i}(x) = \sum_{j=1}^m \frac{\partial g_k}{\partial y_j}(f(x)) \cdot \frac{\partial f_j}{\partial x_i}(x).$$

Theorem 2.13: Generalized Mean Value Theorem

Given $U \subset \mathbb{R}^n$ open, $f \in C^1(U)$. Assume $x_0 \in U$ and $h \in \mathbb{R}^n$ such that $x_0 + th \in U \forall t \in [0, 1]$. Then $f(x_0 + h) - f(x_0) = Df_{x_0 + \theta h}(h)$ for some $\theta \in [0, 1]$.

Proof. For $t \in [0, 1]$, define $g(t) = f(x_0 + th)$. Then,

$$f(x_0 + h) - f(x_0) = g(1) - g(0) \stackrel{\text{MVT}}{=} g'(\theta).$$

Since $g(t) = f \circ \gamma(t)$ where $\gamma(t) = x_0 + th$, by the chain rule,

$$g'(t) = Df_{\gamma(t)} \circ D\gamma_t = Df_{x_0 + th}(h).$$

Letting $t = \theta$, we get the desired result. \square

Definition 2.14: Convex Sets

$A \subset \mathbb{R}^n$ is called **CONVEX** if $\forall x, y \in A$ and $\forall t \in [0, 1]$, we have $tx + (1 - t)y \in A$.

Exercise 2.15:

In \mathbb{R}^n , $B_r(x_0)$ and $\overline{B_r(x_0)}$ are convex.

Definition 2.16: Gradient

Given $U \subset \mathbb{R}^n$ open, $f \in C^1(U)$. Define the **GRADIENT** of f at x as the vector

$$\nabla f(x) = Jf(x)^T = \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{pmatrix}.$$

Proposition 2.17:

Given $U \subset \mathbb{R}^n$ open, $f \in C^1(U)$. Assume U convex and that $\sup_{x \in U} |\nabla f(x)| \leq M$ for some $M > 0$. Then, $\forall x, y \in U$,

$$|f(x) - f(y)| \leq M|x - y|.$$

Proof. Let $x = y + h, y = x_0$. Then

$$\begin{aligned} |f(x) - f(y)| &= |f(x_0 + h) - f(x_0)| \\ &\stackrel{\text{MVT}}{=} |Df_{\xi}(h)| \\ &= |Jf(\xi)h| \\ &= |\nabla f(\xi)h| \\ &\stackrel{CS}{\leq} |\nabla f(\xi)| \cdot |h| \\ &\leq M|x - y|. \end{aligned}$$

We can apply the mean value theorem since U is convex, and thus contains the straight line between x and y . \square

From this we can prove a similar result for $m \geq 2$.

Theorem 2.18: Differentiability vs Lipschitz

Given $U \subset \mathbb{R}^n$ open, $f \in C^1(U, \mathbb{R}^m)$. Suppose, that

$$\Lambda := \sup_{x \in U} \|Jf(x)\|_2 < \infty.$$

Then, f is Lipschitz with constant $\sqrt{m}\Lambda$.

Proof. We apply the previous proposition to each component of f . For $i = 1, \dots, m$, we have

$$|f(x) - f(y)|^2 = \sum_{i=1}^m |f_i(x) - f_i(y)|^2 \leq \sum_{i=1}^m \Lambda_i^2 |x - y|^2 = m\Lambda^2 |x - y|^2.$$

Where $\Lambda_i = \sup_{x \in U} |\nabla f_i(x)|$. Since $|\nabla f_i(x)| \leq \|Jf(x)\|_2$, we have $\Lambda_i \leq \Lambda$ for all $i = 1, \dots, m$. \square

As a consequence, if U is any (not necessarily convex) open subset of \mathbb{R}^n , and $f \in C^1(U, \mathbb{R}^m)$ and $B_r(z) \subset U$ is a ball, then f is Lipschitz on $B_r(z)$ with a constant that may depend on z and r . We then call f **LOCALLY LIPSCHITZ**. Indeed every C^1 function is locally Lipschitz.

Proof. $\overline{B_r(z)}$ is closed and bounded, so compact. So the continuous function $\|Jf(x)\|_2$ attains its maximum on $\overline{B_r(z)}$. So we can apply the above theorem to conclude that f is Lipschitz on $B_r(z)$ with constant $\sqrt{m} \max_{x \in \overline{B_r(z)}} \|Jf(x)\|_2$.

Theorem 2.19:

Given $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R}^m$, differentiable with $Df(x) = 0$ for all $x \in U$. Then, f is constant on each connected component of U . In particular, if U is connected, then f is constant.

Proof. It suffices to consider $m = 1$. Assume U is non-empty and choose $x_0 \in U$. Consider the subset

$$G = \{x \in U \mid f(x) = f(x_0)\} \subset U.$$

Since f is continuous, G is closed in U .

Since U is open, given $x \in G$, we find $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U$. Since every point $y \in B_\varepsilon(x)$ can be connected to x by a line segment, we can apply the mean value theorem to conclude that $f(y) = f(x) = f(x_0)$, so $B_\varepsilon(x) \subset G$. Thus, G is open in U .

Since U is connected, and G and $U \setminus G$ are both open in U , we must have $G = U$, so f is constant on U .

Notice that $x_0 \in G$ so it cannot be empty. \square

Lec 10

2.1 Higher Order Derivatives

Similar to Analysis One, we can define higher order derivatives by iterating the definition of the derivative.

Definition 2.20: C^n functions

Given $U \subset \mathbb{R}^n$ open, we define

$$C^0(U, \mathbb{R}^m) = \{f : U \rightarrow \mathbb{R}^m : f \text{ continuous}\}.$$

$$C^k(U, \mathbb{R}^m) = \{f \in C^{k-1} \mid \partial_i f \in C^{k-1}\}.$$

If $f \in C^k(U, \mathbb{R}^m)$, we say that f is k -times continuously differentiable.

If $f \in C^k(U, \mathbb{R}^m)$ and $i_1, \dots, i_k \in \{1, \dots, n\}$, then we define

$$\partial_{i_1} \dots \partial_{i_k} f = \partial_{i_1}(\partial_{i_2} \dots \partial_{i_k} f).$$

Example 2.21:

Let $f(x_1, x_2, x_3) = x_1 x_3^2 \sin(x_2)$. Then, we can compute

$$\begin{aligned} \partial_2 \partial_1 \partial_2 \partial_3 (f) &= \partial_2 \partial_1 \partial_2 (2x_1 x_3 \sin(x_2)) \\ &= \partial_2 \partial_1 (2x_1 x_3 \cos(x_2)) \\ &= \partial_2 (2x_3 \cos(x_2)) \\ &= -2x_3 \sin(x_2). \end{aligned}$$

Proposition 2.22: C^k is a vector space

If $f, g \in C^k(U)$, $U \subset \mathbb{R}^n$ open, $V \subset \mathbb{R}^m$ open. Let $h : V \rightarrow \mathbb{R}^n$ such that $h(V) \subset U$ and $h \in C^k(V, \mathbb{R}^n)$. Then

1. $f + g \in C^k(U)$ and $f \cdot g \in C^k(U)$.
2. $f \circ h \in C^k(V)$.

Proof. 1. We argue by induction over k .

Base Case ($k = 0$): Very Similar to Analysis I \Rightarrow Exercise.

\square Inductive Step: Assume the claim holds for $k - 1$. Then

$$\partial_i(f + g) = \partial_i f + \partial_i g \in C^{k-1}(U),$$

since $\partial_i f, \partial_i g \in C^{k-1}(U)$ and thus by the inductive hypothesis $\partial_i f + \partial_i g \in C^{k-1}(U)$. Also $f + g \in C^{k-1}(U)$, so $f + g \in C^k(U)$.

Similarly, we can write

$$\partial_i(f \cdot g) = \partial_i f \cdot g + f \cdot \partial_i g \in C^{k-1}(U),$$

since $C^k \subset C^{k-1}$.

2. We argue by induction over k .

Base Case ($k = 0$): Since A is open, $f^{-1}(A)$ is open implying $h^{-1}(f^{-1}(A))$ is open, so $f \circ h$ is continuous.

Inductive Step: Assume the claim holds for $k - 1$. Then, by the chain rule,

$$\partial_j(f \circ h) = \sum_{i=1}^n (\partial_i f) \circ h \partial_j h_i.$$

All of these are by assumptions in $C^{k-1}(V)$. By the inductive hypothesis, and by 1. we have $\partial_j(f \circ h) \in C^{k-1}(V)$, and since $f \circ h \in C^{k-1}(V)$, we have $f \circ h \in C^k(V)$. \square

Theorem 2.23: Schwarz's Theorem

Given $U \subset \mathbb{R}^n$ open, $f \in C^2(U, \mathbb{R}^m)$. Then $\forall i, j \in \{1, \dots, n\}$, we have $\partial_i \partial_j f = \partial_j \partial_i f$.

Proof. If $i = j$, there is nothing to prove. By symmetry, wlog $i < j$. It suffices to consider the case $n = 2$ with $i = 1$ and $j = 2$. For a general n fix i_1, i_2 and consider the function

$$\tilde{f}_b(x_1, x_2) = f(y_1, \dots, x_1, \dots, x_2, \dots, y_n),$$

where y_k is fixed for $k \notin \{i_1, i_2\}$. for $b \in 1, \dots, m$.

Thus, WLOG, $n = 2, m = 1$. For $x = (x_1, x_2) \in U$ and $h > 0$ sufficiently small, define

$$F(h) = f(x_1 + h, x_2 + h) - f(x_1 + h, x_2) - f(x_1, x_2 + h) + f(x_1, x_2).$$

Consider the differentiable function

$$\phi : [0, 1] \rightarrow \mathbb{R}, t \mapsto f(x_1 + th, x_2 + h) - f(x_1 + th, x_2).$$

Thus, by the mean value theorem, $\exists \xi_1 \in (0, 1)$ such that

$$F(h) = \phi(1) - \phi(0) = \phi'(\xi_1).$$

By the chain rule, we can write

$$F(h) = (\partial_1 f(x_1 + \xi_1 h, x_2 + h) - \partial_1 f(x_1 + \xi_1 h, x_2)) h.$$

Define now the function

$$\psi : [0, 1] \rightarrow \mathbb{R}, t \mapsto \partial_1 f(x_1 + \xi_1 h, x_2 + th).$$

Then, by the one dimensional mean value theorem, $\exists \xi_2 \in (0, 1)$ such that

$$F(h) = \psi(1) - \psi(0) = \psi'(\xi_2) = \partial_2 \partial_1 f(x_1 + \xi_1 h, x_2 + \xi_2 h) h^2.$$

Similarly, we can also define $\tilde{\phi}, \tilde{\psi}$ by swapping the roles of x_1 and x_2 , and get

$$F(h) = \partial_1 \partial_2 f(x_1 + \tilde{\xi}_1 h, x_2 + \tilde{\xi}_2 h) h^2,$$

for suitable $\tilde{\xi}_1, \tilde{\xi}_2 \in (0, 1)$. Dividing by h^2 and letting $h \rightarrow 0$, we get

$$\begin{aligned} \partial_2 \partial_1 f(x_1 + \xi_1 h, x_2 + \xi_2 h) &= \partial_1 \partial_2 f(x_1 + \tilde{\xi}_1 h, x_2 + \tilde{\xi}_2 h) \\ \partial_2 \partial_1 f(x) &= \partial_1 \partial_2 f(x). \end{aligned}$$

Corollary 2.24:

Given $U \subset \mathbb{R}^n$ open, $f \in C^k(U)$. Then $\forall i_1, \dots, i_k \in \{1, \dots, n\}$ and for every permutation $\sigma \in S_k$, we have

$$\partial_{i_1} \dots \partial_{i_k} f = \partial_{i_{\sigma(1)}} \dots \partial_{i_{\sigma(k)}} f.$$

A useful shorthand is the following.

Definition 2.25: Multi-index

A **MULTI-INDEX** is a tuple $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$. We define $|\alpha| = \alpha_1 + \dots + \alpha_n$ as the length of α . We say that $\beta \leq \alpha$ if $\beta_i \leq \alpha_i$ for all $i = 1, \dots, n$. We define the factorial

$$\alpha! = \alpha_1! \dots \alpha_n!$$

With this notation we can for example write a polynomial of degree k in n variables compactly as

$$P(x) = \sum_{|\alpha| \leq k} c_\alpha x^\alpha.$$

With this notation, we can also neatly write multiple derivatives as

$$\partial^\alpha f = \partial_1^{\alpha_1} \dots \partial_n^{\alpha_n} f,$$

where $f \in C^k$ and $|\alpha| \leq k$.

Theorem 2.26: Taylor's Theorem

Given $U \subset \mathbb{R}^n$ open, $f \in C^{k+1}(U)$, $k \geq 0$. Let $x_0 \in U$ and $h \in \mathbb{R}^n$ such that $x_0 + th \in U \forall t \in [0, 1]$. Then

$$f(x_0 + h) = \sum_{|\alpha| \leq k} \partial^\alpha f(x_0) \frac{h^\alpha}{\alpha!} + R_{k+1} f(x_0, h).$$

Where the remainder term $R_{k+1} f(x_0, h)$ is given by

$$\begin{aligned} R_{k+1} &= \int_0^1 (k+1)(1-t)^k \sum_{|\alpha|=k+1} \partial^\alpha f(x_0 + th) \frac{h^\alpha}{\alpha!} dt \\ &= O(h^{k+1}) \text{ as } h \rightarrow 0. \end{aligned}$$

Proof. Since U is open $\exists \varepsilon > 0$ such that $x + th \in U$ for all $t \in (-\varepsilon, 1 + \varepsilon)$. By Taylor's theorem in one variable, applied to $\varphi(t) : (-\varepsilon, 1 + \varepsilon) \rightarrow \mathbb{R}, t \mapsto f(x_0 + th)$, we have the Taylor approximation for $\varphi(1)$ around 0 with remainder term given by

$$\varphi(1) = \sum_{m=0}^k \frac{\varphi^{(m)}(0)}{m!} 1^m + \int_0^1 \varphi^{(k+1)}(t) \frac{(1-t)^k}{k!} dt.$$

Applying the chain rule to φ , we get for $t \in (-\varepsilon, 1 + \varepsilon)$,

$$\varphi'(t) = \sum_{i=1}^n \partial_i f(x_0 + th) h_i = \sum_{|\alpha|=1} \partial^\alpha f(x_0 + th) h^\alpha.$$

We now want to show that

$$\varphi^{(m)}(t) = m! \sum_{|\alpha|=m} \partial^\alpha f(x_0 + th) \frac{h^\alpha}{\alpha!}.$$

□ Indeed, by chain rule and induction over m , we have

$$\begin{aligned} \varphi^{(m+1)}(t) &= \frac{d}{dt} \left(m! \sum_{|\alpha|=m} \partial^\alpha f(x_0 + th) \frac{h^\alpha}{\alpha!} \right) \\ &= m! \left(\sum_{|\alpha|=m} \sum_{i=1}^n \partial_i \partial^\alpha f(x_0 + th) h_i \frac{h^\alpha}{\alpha!} \right) \end{aligned}$$

Now, if $|\alpha| = m$, then $\partial_i \partial^\alpha f = \partial^{\alpha+e_i} f$, where e_i is the multi-index with 1 in the i -th position and 0 elsewhere. Thus, we can write

$$\varphi^{(m+1)}(t) = m! \left(\sum_{|\beta|=m+1} \partial^\beta f(x_0 + th) h^\beta \sum_{1 \leq i \leq n, \beta_i \geq 1} \frac{\beta_i}{\beta!} \right).$$

Where we used that

$$\frac{1}{\alpha!} = \frac{\alpha_i + 1}{\alpha_1! \dots (\alpha_i + 1)! \dots \alpha_n!} = \frac{\beta_i}{\beta!}.$$

Putting things together, we get our desired result. □

Corollary 2.27:

Let $x_0 \in U$, $f \in C^{k+1}(U)$ and $P(x)$ be a polynomial of degree $k \geq 0$. Assume that $|f(x_0 + h) - P(h)| = o(|h|^k)$ as $h \rightarrow 0$. Then, P is the Taylor polynomial of f at x_0 of degree k .

Mathematically: $\forall |\alpha| \leq k, \partial^\alpha f(x_0) = \partial^\alpha P(0)$.

Proof. By Taylor's theorem, we get

$$\left| P(h) - \sum_{|\alpha| \leq k} \partial^\alpha f(x_0) \frac{h^\alpha}{\alpha!} + R_{k+1} f(x_0, h) \right| = o(|h|^k).$$

But two polynomials of degree k which differ by $o(|h|^k)$ must be equal. (Exercise) □

Example 2.28:

We want to compute the Taylor expansion of degree 2 of

$$f(x, y) = \sqrt{1 + x - y^2}.$$

From analysis I, we know that

$$\sqrt{1+t} = 1 + \frac{t}{2} - \frac{t^2}{8} + O(t^3) \text{ as } t \rightarrow 0.$$

Plugging in $t = x - y^2$, we get

$$f(x, y) = 1 + \frac{x - y^2}{2} - \frac{(x - y^2)^2}{8} + O((x - y^2)^3).$$

Expanding this and neglecting terms of degree higher than 2, we get

$$f(x, y) = 1 + \frac{x}{2} - \frac{y^2}{2} - \frac{x^2}{8} + O(r^3) \text{ as } r \rightarrow 0.$$

For Taylor approximations to be useful, we need to be able to control the remainder term. This is described by analytic functions.

Definition 2.29: Real analytic functions

Given $U \subset \mathbb{R}^n$ open, we say that $f \in C^\infty(U)$ satisfies analytic estimates in U if for every $x_0 \in U$, there exists $\rho > 0$ and $C > 0$ such that

$$\max_{B_\rho(x_0)} |\partial^\alpha f| \leq C |\alpha|! (n\rho)^{-|\alpha|} \text{ for all } \alpha \in \mathbb{N}^n.$$

Thanks to Taylor's theorem, a function is analytic iff it can be written as a power series around any point in its domain.

Theorem 2.30: Analytic Expansion

Given $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R}$, analytic.

Given $x_0 \in U$, let ρ and C be as in definition 2.29. Then, let

$$P_{x_0,k}(h) := \sum_{|\alpha|=1}^k \partial^\alpha f(x_0) \frac{h^\alpha}{\alpha!}.$$

Then for all $r \in (0, \rho)$, the power series converges absolutely, i.e. $\forall k < l$

$$\sup_{h \in B_r(0)} \sum_{|\alpha|=k}^l \left| \partial^\alpha f(x_0) \frac{h^\alpha}{\alpha!} \right| \leq \frac{C \left(\frac{r}{\rho}\right)^k}{1 - \left(\frac{r}{\rho}\right)}.$$

The series converges for $h \in \overline{B_r(0)}$ and we have

$$\sup_{h \in B_r(0)} |P_{x_0,k}(h) - P_{x_0}(h)| \leq \frac{C \left(\frac{r}{\rho}\right)^k}{1 - \left(\frac{r}{\rho}\right)}.$$

Moreover,

$$f(x_0 + h) = P_{x_0,\infty}(h) \quad \forall h \in B_\rho(0).$$

Lec 12

Proof. The key idea is that for $x \in B_r(x_0)$, $0 < r < \rho$, we have

$$\begin{aligned} \sum_{|\alpha|=m} \left| \frac{\partial^\alpha f(x)}{\alpha!} h^\alpha \right| &\leq \sum_{|\alpha|=m} \frac{|\partial^\alpha f(x)|}{\alpha!} r^m \\ &\leq \sum_{|\alpha|=m} C \frac{\rho^{-m} n^{-m} r^m}{\alpha!} |\alpha|! \\ &= C \left(\frac{r}{\rho}\right)^m n^{-m} \sum_{|\alpha|=m} \frac{|\alpha|!}{\alpha!} = C \left(\frac{r}{\rho}\right)^m. \end{aligned}$$

The last step follows since

$$(1 + \dots + 1)^m = \sum_{|\alpha|=m} \frac{|\alpha|!}{\alpha!}.$$

Thus for the tail of the series,

$$\sum_{|\alpha|=k}^\infty \left| \frac{\partial^\alpha f(x_0)}{\alpha!} h^\alpha \right| \leq \sum_{m=k}^\infty C \left(\frac{r}{\rho}\right)^m = \frac{C \left(\frac{r}{\rho}\right)^k}{1 - \left(\frac{r}{\rho}\right)}.$$

So $P_{x_0,\infty}(h) = \lim_{k \rightarrow \infty} P_{x_0,k}(h)$ exists for $h \in B_\rho(x_0)$.

The error is given by Taylor, $f(x_0 + h) = P_{x_0,k}(h) + R_{x_0,k+1}^f$.

$$\begin{aligned} |R_{x_0,k+1}^f(h)| &= \left| \int_0^1 (k+1)(1-t)^k \sum_{|\alpha|=k+1} \partial^\alpha f(x_0 + th) \frac{h^\alpha}{\alpha!} dt \right| \\ &\leq \int_0^1 (k+1)(1-t)^k \sum_{|\alpha|=k+1} \left| \partial^\alpha f(x_0 + th) \frac{h^\alpha}{\alpha!} \right| dt \\ &\leq \int_0^1 (k+1)(1-t)^k C \left(\frac{r}{\rho}\right)^{k+1} dt \leq C \left(\frac{r}{\rho}\right)^{k+1}. \end{aligned}$$

As a consequence, by triangle inequality,

$$\begin{aligned} |f(x_0 + h) - P_{x_0,\infty}(h)| &\leq \underbrace{|f(x_0 + h) - P_{x_0,k}(h)|}_{\text{Rest}} + \underbrace{|P_{x_0,k}(h) - P_{x_0,\infty}(h)|}_{\text{Tail}} \\ &\leq C \left(\frac{r}{\rho}\right)^{k+1} + \frac{C \left(\frac{r}{\rho}\right)^k}{1 - \left(\frac{r}{\rho}\right)} \rightarrow 0 \text{ as } k \rightarrow \infty. \end{aligned}$$

□

Corollary 2.31: Unique Continuation Principle

Given $U \subset \mathbb{R}^n$ open and connected, f, g analytic in U such that $\exists x_0 \in U$ with

$$\partial^\alpha f(x_0) = \partial^\alpha g(x_0) \quad \forall \alpha \in \mathbb{N}^n.$$

Then $f = g$ on U .

In particular, if $V \subset U$ is an open subset such that $f = g$ on V , then $f = g$ on U .

Proof. Let $x_0 \in G := \{x \in U \mid \partial^\alpha f(x) = \partial^\alpha g(x) \forall \alpha \in \mathbb{N}^n\}$. We will show that $U \setminus G$ and G are both open in U . Since U is connected and $x_0 \in G$, we must have $G = U$.

i. $U \setminus G$ is open in U : By definition of G ,

$$\begin{aligned} G &= \bigcap_{\alpha \in \mathbb{N}^n} \{x \in U \mid \partial^\alpha (f - g)(x) = 0\} \\ U \setminus G &\stackrel{\text{De Morgan}}{=} \bigcup_{\alpha \in \mathbb{N}^n} \{x \in U \mid \partial^\alpha (f - g)(x) \neq 0\} \\ &= \bigcup_{\alpha \in \mathbb{N}^n} [\partial^\alpha (f - g)]^{-1}(\mathbb{R} \setminus \{0\}). \end{aligned}$$

Since $\mathbb{R} \setminus \{0\}$ is open and $\partial^\alpha (f - g)$ is continuous, $[\partial^\alpha (f - g)]^{-1}(\mathbb{R} \setminus \{0\})$, the preimage is open. Since arbitrary unions of open sets are open, $U \setminus G$ is open.

ii. G is open in U : Let $x \in G \subset U$. Since f, g are analytic, $\exists \rho > 0$ such that $\forall h \in B_\rho(x)$,

$$f(x + h) = P_{x,\infty}^f(h), g(x + h) = P_{x,\infty}^g(h).$$

But $P^f = P^g$ since Taylor polynomials are defined by the derivatives at $x \in G$.

Hence, $f \equiv g$ on $B_\rho(x)$, so $B_\rho(x) \subset G$, so G is open. □

3 Optimization

In this chapter we want to treat the standard problem of minimizing or maximizing a function $f(x_1, \dots, x_n) \in \bar{U}$ with constraints $g_k(x_1, \dots, x_n) = 0$ for $k = 1, \dots, m$.

3.1 First and Second Order Optimality Conditions

To describe the optimality conditions, we need to introduce some definitions.

Definition 3.1: Local Minimum / Maximum

Given $U \subset \mathbb{R}^n$ open, $f \in C^1(U)$. We say that x_0 is a **LOCAL MINIMUM** of f if $\exists r > 0$ such that

$$f(x_0) \leq f(x) \text{ for all } x \in B_r(x_0).$$

LOCAL MAXIMUM of f if $\exists r > 0$ such that

$$f(x_0) \geq f(x) \text{ for all } x \in B_r(x_0).$$

STRICT LOCAL MINIMUM of f if $\exists r > 0$ such that

$$f(x_0) < f(x) \text{ for all } x \in B_r(x_0) \setminus \{x_0\}.$$

If x_0 is a local minimum or maximum of f , then x_0 is called a **LOCAL EXTREMUM** of f .

Definition 3.2: Critical Point

Given $U \subset \mathbb{R}^n$ open, $f \in C^1(U)$. Then $x_0 \in U$ is a **CRITICAL POINT** of f if $\nabla f(x_0) = 0$.

In other words, $\partial_i f(x_0) = 0$ for all $i = 1, \dots, n$.

Proposition 3.3:

Every local Extremum is a critical point.

Proof. Assume wlog x_0 is a local minimum. Fix $i \in 1, \dots, n$. Then

$$\partial_i f(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + te_i) - f(x_0)}{t}.$$

Letting t approach 0 from both sides, we get

$$\begin{aligned} \lim_{t \rightarrow 0^+} \frac{f(x_0 + te_i) - f(x_0)}{t} &\geq 0 \\ \lim_{t \rightarrow 0^-} \frac{f(x_0 + te_i) - f(x_0)}{t} &\leq 0. \end{aligned}$$

Since the limit exists, we must have $\partial_i f(x_0) = 0$. \square

Theorem 3.4: Lagrange Multipliers

Given $U \subset \mathbb{R}^n$ open, $f, g_j \in C^1(U)$ for $j = 1, \dots, m$. Let $M = \{x \in U \mid g_1(x) = \dots = g_m(x) = 0\}$ be the constraint set. If x_0 is a local minimum of $f|_M$, then $\exists \lambda_*, \lambda_1, \dots, \lambda_m \in \mathbb{R}$, with $\lambda_*^2 + \lambda_1^2 + \dots + \lambda_m^2 = 1$, such that

$$\lambda_* \nabla f(x_0) + \sum_{j=1}^m \lambda_j \nabla g_j(x_0) = 0.$$

$\lambda_*, \lambda_1, \dots, \lambda_m$ are called the **LAGRANGE MULTIPLIERS**.

Proof. First assume that x_0 is a strict local minimum.

For $\varepsilon > 0$, let $f_\varepsilon(x) = f(x) + \frac{1}{2\varepsilon} \sum_{j=1}^m g_j(x)^2$, defined for $x \in \overline{B_{\frac{r}{2}}(x_0)}$, where r is such that $f(x) > f(x_0)$ for all $x \in B_r(x_0) \setminus \{x_0\}$.⁴ Let $\varepsilon_l = \frac{1}{l}$. Choose x_l as a point of minimum of the continuous function f_{ε_l} on the compact set $\overline{B_{\frac{r}{2}}(x_0)}$.

We claim that some subsequence x_{l_k} converges to x_0 . Indeed, by compactness, we can choose a subsequence x_{l_k} converging to some $\bar{x} \in \overline{B_{\frac{r}{2}}(x_0)}$. Since x_l is a minimum we have

$$f_{\varepsilon_l}(x_l) \leq f_{\varepsilon_l}(x_0) = f(x_0).$$

So also

$$\frac{1}{2\varepsilon_l} \sum_{j=1}^m g_j(x_l)^2 \leq f_{\varepsilon_l}(x_l) \leq f(x_0).$$

This implies that $\sum_{j=1}^m g_j(x_l)^2 \leq 2\varepsilon_l f(x_0) \rightarrow 0$ as $l \rightarrow \infty$, so $\sum_{j=1}^m g_j(\bar{x})^2 = 0$, so $\bar{x} \in M$. From this, we get

$$f(x_{l_k}) \leq f_{\varepsilon_{l_k}}(x_{l_k}) \leq f(x_0).$$

Since $x_{l_m} \rightarrow \bar{x}$, by continuity of f we get $f(\bar{x}) \leq f(x_0)$. Since x_0 is a strict local minimum, we must have $\bar{x} = x_0$. Lec 13

Let $y_m = x_{l_m}$ and $\tilde{\varepsilon}_m = \varepsilon_{l_m}$. Since $y_m \in B_r(x_0)$ (for large m , y_m cannot be on the boundary), is an interior minimum of $f_{\tilde{\varepsilon}_m}$, we have

$$0 = \tilde{\varepsilon}_m \nabla f_{\tilde{\varepsilon}_m}(y_m) = \tilde{\varepsilon}_m \nabla f(y_m) + \sum_{j=1}^m g_j(y_m) \nabla g_j(y_m).$$

Let $\mu_m^2 = \tilde{\varepsilon}_m^2 + \dots + [g_m(y_m)]^2 > 0$. Then,

$$\left(\frac{\tilde{\varepsilon}_m}{\mu_m}\right)^2 + \dots + \left(\frac{g_m(y_m)}{\mu_m}\right)^2 = 1.$$

Define

$$\lambda_0^m = \frac{\tilde{\varepsilon}_m}{\mu_m}, \lambda_j^m = \frac{g_j(y_m)}{\mu_m} \text{ for } j = 1, \dots, m.$$

So there exists $\lambda^m = (\lambda_0^m, \lambda_1^m, \dots, \lambda_m^m) \in S^j$ i.e.

$$(\lambda_0^m)^2 + \dots + (\lambda_m^m)^2 = 1.$$

Hence,

$$0 = \lambda_0^m \nabla f(y_m) + \sum_{j=1}^m \lambda_j^m \nabla g_j(y_m).$$

Since S^j is closed and bounded, it is compact. Hence, there exists a subsequence λ^{m_k} converging to some $\lambda_* = (\lambda_*, \lambda_1, \dots, \lambda_m) \in S^j$. Since $y_{m_k} \rightarrow x_0$, by continuity of ∇f and ∇g_j , we get

$$0 = \lambda_* \nabla f(x_0) + \sum_{j=1}^m \lambda_j \nabla g_j(x_0).$$

\square

Theorem 3.5: Spectral Theorem

If A is $n \times n$ symmetric ($A^T = A$), with coefficients in \mathbb{R} , then A diagonalizes in some orthonormal basis and has real eigenvalues.

Also, for $O = (v_1 | \dots | v_n)$, where v_1, \dots, v_n is an orthonormal basis of eigenvectors of A , we have

$$O^T O = I_n, O^{-1} = O^T, O^{-1} A O = \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

⁴This function is called penalizing function as for points not satisfying the constraints, it diverges to $+\infty$ as $\varepsilon \rightarrow 0$.

Lemma 3.6:

Assume v_1, \dots, v_k , with $0 \leq k < n$, are vectors such that $v_i \cdot v_j = \delta_{ij}$. Let $A \in M_{n \times n}(\mathbb{R})$ be symmetric.

If there exist $\mu_i \in \mathbb{R}$ such that $Av_i = \mu_i v_i$ then there exists $w \in \mathbb{R}^n$ such that $|w| = 1$, $w \cdot v_i = 0$ and $Aw = \lambda w$ for some $\lambda \in \mathbb{R}$.

Proof. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $x \mapsto x \cdot Ax$. We want to minimize this function under the constraints

$$\begin{aligned} g_*(x) &= |x|^2 - 1 = 0 \\ g_1(x) &= v_1 \cdot x = 0 \\ &\vdots \\ g_k(x) &= v_k \cdot x = 0. \end{aligned}$$

We apply theorem 3.4 with

$$M = \{x \in \mathbb{R}^n \mid g_*(x) = \dots = g_k(x) = 0\} \subset S^{n-1}.$$

Note that $M \subset \mathbb{R}^n$ is closed and bounded, hence compact.

Since f is continuous, there exists $w \in M$ as a point of minimum of $f|_M$. Hence, $\exists \lambda_0, \lambda_*, \lambda_1, \dots, \lambda_k \in \mathbb{R}$, with $\sum_i \lambda_i^2 = 1$, such that

$$\lambda_0 \nabla f(w) + \lambda_* \nabla g_*(w) + \dots + \lambda_k \nabla g_k(w) = 0.$$

Note that in general, for symmetric matrices

$$v \cdot w = v^T w \Rightarrow v \cdot Aw = Aw \cdot v = (Av) \cdot w = w \cdot Av.$$

Hence, calculating the partial derivatives, we get

$$\begin{aligned} \partial_i f(x) &= \lim_{t \rightarrow 0} \frac{f(x + te_i) - f(x)}{t} \\ &= \lim_{t \rightarrow 0} \frac{(x + te_i) \cdot A(x + te_i) - x \cdot Ax}{t} \\ &= \lim_{t \rightarrow 0} \frac{2te_i \cdot Ax + t^2 e_i \cdot Ae_i}{t} = 2e_i \cdot Ax \\ &\Rightarrow \nabla f(x) = 2Ax. \end{aligned}$$

Replacing A with I , the above calculation gives $\nabla g_*(x) = 2x$. Also

$$\partial_i g_j(x) = \lim_{t \rightarrow 0} \frac{v_j \cdot (x + te_i) - v_j \cdot x}{t} = v_j \cdot e_i \Rightarrow \nabla g_j(x) = v_j.$$

Plugging this into the Lagrange multipliers condition, we get

$$\underbrace{2\lambda_0 Aw + 2\lambda_* w}_{w_1} + \underbrace{\lambda_1 v_1 + \dots + \lambda_k v_k}_{w_2} = 0.$$

Since $w \in M$, we have $w \cdot v_j = 0 \forall j = 1, \dots, k$. This implies that

$$v_j \cdot Aw = w \cdot Av_j = w \cdot \mu_j v_j = \mu_j w \cdot v_j = 0.$$

This means that Aw is perpendicular to v_1, \dots, v_k . Since w_2 is a linear combination of v_1, \dots, v_k , we get $w_1 \cdot w_2 = 0$. Hence, $w_1 + w_2 = 0$ implies $w_1 = w_2 = 0$. If the sum of two perpendicular vectors is 0, then both vectors must be 0. Hence $2\lambda_0 Aw + 2\lambda_* w = 0$, and $\lambda_i = 0$ since $w_2 = 0$ and w_2 is a linear combination of v_1, \dots, v_k which are linearly independent.

By definition of the Lagrange multipliers, $\lambda_0^2 + \lambda_*^2 = 1$. Hence, $Aw = \lambda w$ for $\lambda = -\frac{\lambda_*}{\lambda_0} \in \mathbb{R}$. Note that $\lambda_0 \neq 0$ since otherwise $\lambda_*^2 = 1$, so $\lambda_* = \pm 1$ and $w = 0$, contradicting $|w| = 1$. \square

Proof. [Theorem 3.5] By induction over k , we can construct an orthonormal basis of eigenvectors of A . \square

For second order optimality conditions, we need to introduce the following definition.

Definition 3.7: Hessian

Given $U \subset \mathbb{R}^n$ open, $f : U \rightarrow \mathbb{R} \in C^2$. The **HESIAN MATRIX** of f at $x \in U$ is the $n \times n$ matrix defined by

$$H_{ij} f(x) = \partial_i \partial_j f(x).$$

By Schwarz's theorem, the Hessian is a symmetric matrix. It is also denoted $D^2 f(x)$. The **LAPLACIAN** of f is the trace of the Hessian:

$$\Delta f(x) = \text{tr}(Hf(x)) = \sum_{i=1}^n \partial_{ii} f(x).$$

Definition 3.8:

$A \in M_{n \times n}(\mathbb{R})$ is **NON-NEGATIVE DEFINITE** if $\forall x \in \mathbb{R}^n, x \cdot Ax \geq 0$.

Proposition 3.9:

Given $U \subset \mathbb{R}^n$ open, $f \in C^3(U)$. If f has a local minimum at $x_0 \in U$, then $Hf(x_0)$ is non-negative definite.

Proof. By Taylor, $f(x_0 + x) = f(x_0) + \nabla f(x_0) \cdot x + \frac{1}{2} x \cdot Hf(x_0) x + O(|x|^3)$ as $x \rightarrow 0$.

By the spectral theorem, $O^T H O = \Lambda$.

Let $y = O x \in \mathbb{R}^n$. Then

$$f(x_0 + x) = f(x_0) + \underbrace{\nabla f(x_0) \cdot x}_{=0} + \frac{1}{2} y \cdot \Lambda y + O(|y|^3) \text{ as } y \rightarrow 0.$$

Since x_0 is a local minimum, we must have $y \cdot \Lambda y \geq 0$ for all $y \in \mathbb{R}^n$. This implies that Λ is non-negative definite, so $Hf(x_0)$ is non-negative definite. \square

Lec 14

Definition 3.10:

A matrix A is called

- **POSITIVE DEFINITE** if $\min\{\lambda_i\} > 0$
- **NONNEGATIVE DEFINITE** if $\min\{\lambda_i\} \geq 0$
- **NEGATIVE DEFINITE** if $\max\{\lambda_i\} < 0$
- **NON-POSITIVE DEFINITE** if $\max\{\lambda_i\} \leq 0$
- **INDEFINITE** if $\min\{\lambda_i\} < 0 < \max\{\lambda_i\}$

Proposition 3.11: Hessian test

Let $x_0 \in U$ be a critical point of $f \in C^3(U)$.

- 1) x_0 local minimum $\Rightarrow Hf(x_0)$ non-negative definite.
- 2) $Hf(x_0)$ positive definite $\Rightarrow x_0$ local minimum.
- 3) x_0 local maximum $\Rightarrow Hf(x_0)$ non-positive definite.
- 4) $Hf(x_0)$ negative definite $\Rightarrow x_0$ local maximum.
- 5) $Hf(x_0)$ indefinite $\Rightarrow x_0$ is neither local minimum nor local maximum.

Proof. We will only show 2) and 3).

2) By Taylor, $f(x_0 + \underbrace{tOy}_{=x}) = f(x_0) + \frac{1}{2} \sum_{i=1}^n \lambda_i y_i^2 + O(|y|^3)$ as $y \rightarrow 0$.

Let $\lambda = \min\{\lambda_i\} > 0$. Then, for $|y| \leq r_0$ we have

$$\begin{aligned} f(x_0 + tOy) &\geq f(x_0) + \frac{1}{2} \sum_{i=1}^n \lambda y_i^2 - C|y|^3 \\ &\geq f(x_0) + \frac{1}{2} \lambda |y|^2 - C|y|^3 \end{aligned}$$

We can choose r small enough such that $\frac{1}{2} \lambda r^2 - Cr^3 > 0$. Then $\forall y \in B_r(x_0)$, $f(x_0 + x) > f(x_0)$, so x_0 is a local minimum.

3) Assume by contradiction $\exists \lambda_j > 0$. Then by Taylor,

$$f(x_0 + tOe_i) \geq f(x_0) + \lambda_j t^2 + O(t^3) > f(x_0) \text{ for } t > 0 \text{ small enough.}$$

This contradicts the fact that x_0 is a local maximum. \square

Theorem 3.12: Fundamental Theorem of Algebra

A polynomial of degree n with complex coefficients has n complex roots, up to repetitions.

Proof. We show that there exists one root. Dividing and inductively applying the fundamental theorem of algebra to the quotient, we get the desired result.

Let $P(t) = a_0 + a_1 t + \dots + a_n t^n$ be a polynomial of degree n with $a_i \in \mathbb{C}$ and $a_n = 1$. We want to find t_0 such that $|P(t_0)| = 0$. We thus try to minimize $|P(z)|$ for $z \in \mathbb{C}$. The issue here is, that \mathbb{C} is not compact, so we cannot apply Weierstrass theorem.

Step 1: Let $\beta = \inf_{z \in \mathbb{C}} |P(z)|$. We show that $|P(z)| \rightarrow \infty$ as $|z| \rightarrow \infty$. Indeed,

$$\begin{aligned} P(z) &= a_0 + a_1 z + \dots + a_{n-1} z^{n-1} + z^n \\ &\stackrel{\Delta}{\geq} |z|^n - (|a_0| + |a_1||z| + \dots + |a_{n-1}||z|^{n-1}) \\ &\geq R^n - CR^{n-1} \text{ for } |z| \geq R \\ &\geq 1 + 10\beta > \beta \text{ for } R \text{ large enough.} \end{aligned}$$

Here we defined $C = |a_0| + |a_1| + \dots + |a_{n-1}|$.

Hence, the Infimum is attained in $\overline{B_R(0)}$ so $\exists z_0 \in \overline{B_R(0)}$ such that $|P(z_0)| = \beta$. This implies that

$$\inf_{\mathbb{C}} |P(z)| = \min_{\overline{B_R(0)}} |P(z)|,$$

since P is continuous and $\overline{B_R(0)}$ is compact.

Since we had a strict inequality, $z_0 \in B_R(0)$, i.e. z_0 does not lie on the boundary. If $\beta = 0$, then $P(z_0) = 0$ so we have a root.

Assume thus, $\beta > 0$. Then $P(z_0 + z)$ has a minimum at $z = 0$. By Taylor, we can write

$$P(z_0 + z) = \sum_{k=1}^n a_k (z_0 + z)^k = \sum_{k=1}^n \underbrace{b_k z^k}_{:=Q}, b_k \in \mathbb{C}.$$

Notice, that $|b_0| = |Q(0)| = |P(z_0)| = \beta > 0$. Let $l = \min\{k \geq 1 \mid b_k \neq 0\}$, i.e. l is the order of the first non-zero term in the Taylor expansion of $P(z_0 + z)$ around $z = 0$. Then

$$\begin{aligned} |P(z_0 + z)| &= |b_0 + b_l z^l + \dots| \\ &= |b_0| \left| 1 + \frac{b_l}{b_0} z^l + \dots \right| \\ &= |b_0| \left| c_l z^l + \sum_{k=1+l}^n c_k z^k \right| \text{ for } c_k = \frac{b_k}{b_0}. \end{aligned}$$

We have a contradiction if $c_l z^l < 0$. Note that we must have

$$c_l z^l = \rho e^{i\alpha} |z|^l e^{il \arg(z)} \in \mathbb{R}.$$

This is exactly the case if $\alpha + l \arg(z) = \pi$. Thus, we choose γ accordingly (i.e. $\alpha + l\gamma = \pi$) and let $z = r e^{i\gamma}$ for $r > 0$ small enough. Then

$$\begin{aligned} |P(z_0 + z)| &= |b_0| \left| 1 + \rho r^l e^{i\pi} + O(r^{l+1}) \right| \\ &\leq |b_0| \left| 1 - \rho r^l + Cr^{l+1} \right| \\ &< |b_0| = \beta \text{ for } r \text{ small enough.} \end{aligned}$$

This contradicts the fact that $z = 0$ is a minimum of $P(z_0 + z)$, so we must have $\beta = 0$, so $P(z_0) = 0$ and z_0 is a root of P . \square

Definition 3.13: Convex Functions

Let $A \subset \mathbb{R}^n$ be a convex set. $f : A \rightarrow \mathbb{R}$ is **CONVEX** if $\forall x, y \in A, t \in [0, 1]$,

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$

Index

- L^p -norm, 8
- accumulation point, 2
- boundary, 3
- bounded, 5
- Cauchy sequence, 2
- closed, 3, 6
- closure, 3
- complete, 2
- connected, 6
- connectedness, 6
- continuous, 4
- contraction, 4
- convex, 11, 17
- cover, 5
- critical point, 15
- curve, 6
- differentiable, 9
- differential, 9
- dimension, 1
- directional derivative, 9
- disconnected, 6
- endpoint, 6
- euclidean distance, 1
- euclidean norm, 1
- euclidean structure, 1
- fixed point, 4
- gradient, 11
- Hessian Matrix, 16
- Hilbert-Schmidt norm, 7
- indefinite, 16
- inner product, 8
- interior, 3
- Jacobi matrix, 11
- Lagrange multipliers, 15
- Laplacian, 16
- limit, 2
- Lipschitz constant, 4
- Lipschitz continuous, 4
- local Extremum, 15
- local maximum, 15
- local minimum, 15
- locally Lipschitz, 12
- loop, 6
- Manhattan metric, 1
- metric space, 1
- multi-index, 13
- negative definite, 16
- non-negative definite, 16
- non-positive definite, 16
- nonnegative definite, 16
- norm, 7
- open, 3
- open ball, 3
- open cover, 5
- partial derivative, 10
- path, 6
- positive definite, 16
- scalar product, 1
- sequence, 2
- sequentially compact, 5
- starting point, 6
- strict local minimum, 15
- subcover, 5
- subsequence, 2
- supremum norm, 8
- topologically compact, 5
- uniformly continuous, 4