Analysis 2

Lecture notes, University (of Technology) Graz
based on the lecture by Wolfgang Ring

Lukas Prokop

June 7, 2018

Contents

This lecture took place on 2018/03/06.

1 Mathematical Redux and topological fundamentals

1.1 Metric

Definition 1.1. *Let $X \neq \emptyset$ be a set. We define a map $d : X \times X \to [0, \infty)$. d should behave like a geometrical distance. We require $\forall x, y, z \in X$:*

- $d(x, y) = d(y, x)$ *[called* symmetry*]*
- $d(x, y) = 0 \iff x = y$ *[called* positive definiteness*]*
- $\forall x, y, z \in X : d(x, z) \leq d(x, y) + d(y, z)$ *[called* triangle inequality*]*

Then d is called metric *or* distance function *on X. (X, d) is called* metric space.

Example 1.1.

- $X \subseteq \mathbb{C}$, $d(x, y) = \left|x - y\right|$. *It satisfies* $|x - z| \leq \left|x - y\right| + \left|y - z\right|$
- $X \subseteq \mathbb{R}^n$, $\left\|x - y\right\| = \langle x - y, x - y\rangle^{\frac{1}{2}}$

Claim.

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

$$\|x\| = \langle x, x\rangle^{\frac{1}{2}} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

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

It holds that $\left\|x + y\right\| \leq \|x\| + \left\|y\right\|$ *[triangle inequality].*

Proof.

$$
\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y\rangle \\
&= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle \\
&= \|x\|^2 + 2\langle x,y\rangle + \|y\|^2 \\
&\leq \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 \qquad \text{[see Cauchy-Schwarz inequality]} \\
&= (\|x\| + \|y\|)^2 \\
\|x-y\|^2 &= \langle x-y, x-y\rangle \\
&= \|x\|^2 - 2\langle x,y\rangle + \|y\|^2 \\
\|x+y\|^2 + \|x-y\|^2 &= 2\left(\|x\|^2 + \|y\|^2\right)
\end{aligned}
$$

□

1.2 Cauchy-Schwarz inequality

Theorem 1.1 (Cauchy-Schwarz inequality)**.**

$$\left|\langle x,y\rangle\right| \leq \|x\|\,\|y\|$$

Proof.

$$0 \leq \langle x-\lambda y, x-\lambda y\rangle = \|x\|^2 - 2\lambda\langle x,y\rangle + \lambda^2\|y\|^2 \qquad \forall\lambda\in\mathbb{R}$$

Let $\lambda = \frac{\langle x,y\rangle}{\|y\|^2}$. Then,

$$
\begin{aligned}
0 &\leq \|x\|^2 - 2\frac{\left|\langle x,y\rangle\right|^2}{\|y\|^2} + \frac{\left|\langle x,y\rangle\right|^2}{\|y\|^4}\cdot\|y\|^2 \\
&\implies 0 \leq \|x\|^2 - \frac{\left|\langle x,y\rangle\right|^2}{\|y\|^2} \\
&\implies \left|\langle x,y\rangle\right|^2 \leq \|x\|^2\cdot\|y\|^2
\end{aligned}
$$

□

1.3 Euclidean norm

Definition 1.2. $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$ *is called* Euclidean norm *(length) of vector* $x\in\mathbb{R}^n$.
$\|x\| = \langle x,x\rangle^{\frac{1}{2}}$ *It holds that*

1. $\|\lambda x\| = |\lambda| \, \|x\| \, \forall x \in \mathbb{R}^n, \lambda \in \mathbb{R}$
2. $\|x\| = 0 \iff x = 0$ *in* $\mathbb{R}^n$
3. $\left\|x + y\right\| \leq \|x\| + \left\|y\right\|$

In general: Let V be a vector space over $\mathbb{R}$. A map $\|\cdot\|$, which assigns every vector x a non-negative real number satisfying the properties above, is called norm on V*. Then* $(V, \|\cdot\|)$ *is called a* normed vector space*.*

Let $X \subseteq \mathbb{R}^n$ (V is a normed vector space), then $d(x, y) = \left\|x - y\right\|$ is a metric on X.

$$\left\|y - x\right\| = \left\|(-1)(x - y)\right\| = |-1| \cdot \left\|x - y\right\| = \left\|x - y\right\|$$

$$d(x, y) = 0 \iff \left\|x - y\right\| = 0 \iff x - y = 0 \iff x = y$$

$$d(x, z) = \|z - x\| = \left\|z - y + y - x\right\| \leq \left\|z - y\right\| + \left\|y - x\right\| = d(z, y) + d(y, x)$$

1.4 Metric space

Example 1.2 (metric space)**.** *Metric space, distance is not a norm. Consider an area in* $\mathbb{R}^3$*.*

$d(x, y)$ is the shortest path, connecting x and y in X. See Figure 1

Example 1.3 (French railway)**.** *All connections between two cities pass through Paris except one city is Paris.*

Example 1.4. *$X = \mathbb{R}^2$. Let $p \in \mathbb{R}^2$ be fixed.*

$$d(x, y) = \begin{cases} \left|x - y\right| & \text{if } x, y, p \text{ are on one line} \\ \left|x - p\right| + \left|p - y\right| & \text{if } x, y, p \text{ are not on one line} \end{cases}$$

1.5 Open sets, convergence and accumulation points

Now we put some terminology into the context of a metric space. (X, d) is a metric space.

Definition 1.3. *Let $x \in X$, $r \geq 0$.*

$$K_r(x) = \{z \in X \,|\, d(x, z) < r\}$$

Is an open sphere *with radius r and center x.*

Definition 1.4.

$$\overline{K_r(x)} = \{z \in X \,|\, d(x, z) \leq r\}$$

Closed sphere with center x and radius r.

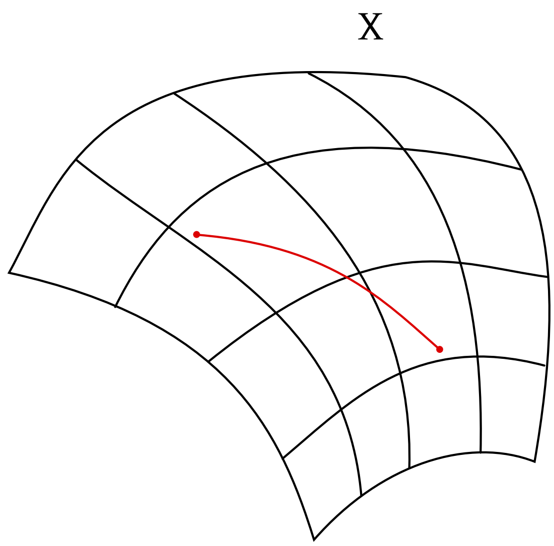

Figure 1: Example in $\mathbb{R}^3$. The red line illustrates the shortest path

Definition 1.5 (Sequences in X)**.** *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in X (hence, $x_n \in X \forall n \in \mathbb{N}$)*

1. *$(x_n)_{n\in\mathbb{N}}$ is called* convergent *and limit $x \in X$ if*

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} : n \geq N \implies d(x_n, x) < \varepsilon$$

Denoted as $\lim_{n\to\infty} x_n = x$.

2. *$(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence if*

$$\forall \varepsilon > 0 \exists N \in \mathbb{N} : n, m \geq N \implies d(x_n, x_m) < \varepsilon$$

Every convergent sequence is also a Cauchy sequence.

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be convergent with limit x. Let $\varepsilon > 0$ be arbitrary. Because $(x_n)_{n\in\mathbb{N}}$ is convergent, there exists $N \in \mathbb{N}$ such that $n \geq N \implies d(x_n, x) < \frac{\varepsilon}{2}$. Now let $n, m \geq N$. Then it holds that

$$d(x_n, x_m) \leq \underbrace{d(x_n, x)}_{<\frac{\varepsilon}{2}} + \underbrace{d(x, x_m)}_{<\frac{\varepsilon}{2}} < \varepsilon$$

$\square$

Definition 1.6. (X, d) *is called* complete metric space *if every Cauchy sequence in X is also convergent (has a limit).*

$\mathbb{R}$ is complete. $\mathbb{R}^n$ is also complete. $\mathbb{Q} \subseteq \mathbb{R}$ is incomplete.

Definition 1.7. *Let $(x_n)_{n \in \mathbb{N}}$ be a sequence of X is called "accumulation point" (dt. Häufungspunkt) of the sequence. $\forall \varepsilon > 0$, it holds that $K_\varepsilon(x)$ contains infinitely many sequence elements.*

This lecture took place on 2018/03/08.

(X, d) is called *metric space*.

$$d(x, y) = 0 \iff x = y$$

$$\forall x, y \in X : d(x, y) = d(y, x)$$

$$d(x, z) \leq d(x, y) + d(y, z) \forall x, y, z \in X$$

1.6 Norm

Let V be a vector space. $\|\cdot\|$ is called *norm on* V.

$$\|x\| = 0 \iff x = 0$$

$$\forall \lambda \in \mathbb{R}, \mathbb{C} : \forall x \in V : \|\lambda x\| = |\lambda| \, \|x\|$$

$$\forall x, y, z \in V : \left\|x + y\right\| \leq \|x\| + \left\|y\right\|$$

Let $X \subseteq V$ be a subset of normed vector space V. Then X is a metric space with $d(x, y) = \left\|x - y\right\|$.

For $V = \mathbb{R}^n$. Then

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}}$$

is a norm on $\mathbb{R}^n$. $\|x\|_2$ is called *Euclidean norm on* $\mathbb{R}^n$.

Other norms in $\mathbb{R}^n$:

$$\|x\|_\infty = \max \{ |x_i| \mid i = 1, \ldots, n \}$$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

for $1 \leq p < \infty$.

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

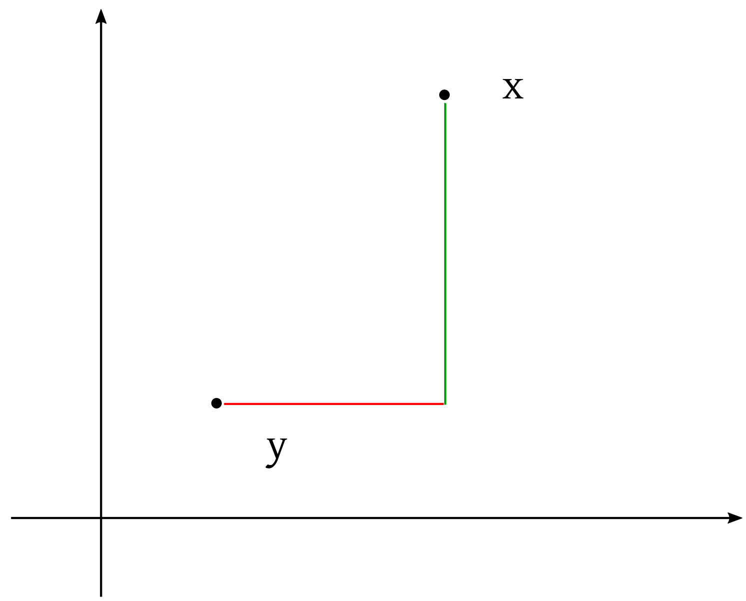

Figure 2: Visualizing $\|x\|_1$

e.g. $\|x\|_1$ in $\mathbb{R}^2$

$$\left\|x-y\right\| = \left|x_1 - y_1\right| + \left|x_1 - y_2\right|$$

is the so-called *Manhattan metric*.

The concepts “subsequence”, “final element of a sequence”, “reordering of a sequence” correspond one-by-one to metric spaces.

Definition 1.8 (Accumulation point)**.** *Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence in* X. $x \in X$ *is called* accumulation point of sequence X *if* $\forall\varepsilon > 0$ *the sphere* $K_\varepsilon(x)$ *contains infinitely many elements.*

Lemma 1.1. $x \in X$ *is accumulation point of sequence* $(x_n)_{n\in\mathbb{N}}$ *if and only iff there exists a subsequence* $(x_{n_k})_{k\in\mathbb{N}}$ *such that* $x = \lim_{k\to\infty} x_{n_k}$.

Proof. See Analysis 1 course □

1.7 Contact point

Let $B \subseteq X$, X is a metric space. Then B with d is a metric space itself.

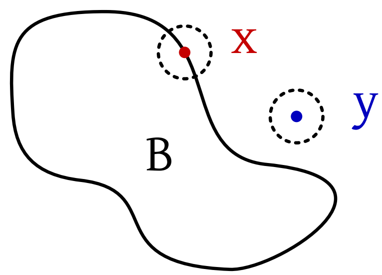

Figure 3: Contact points in set B

Definition 1.9. *Let $B \subseteq X$ and $x \in X$. We say, x is a* contact point of B *if* $\forall \varepsilon > 0 : K_\varepsilon(x) \cap B \neq \emptyset$.

[$y \in X$ is not a contact point of $B \iff \exists \varepsilon > 0 : K_\varepsilon(y) \cap B = \emptyset$]

See Figure 3.

We let $\overline{B} = \left\{ x \in X \,\middle|\, x \text{ is contact point of } B \right\}$.

$\overline{B}$ is called closed hull of B.

B is called closed if $B = \overline{B}$, hence, every contact point is also element of B.

Remark 1.1. *Because $\forall x \in B$ holds $K_r(x) \cap B \supseteq \{x\}\ \forall r > 0$ is x always contact point of B. Also $B \subseteq \overline{B}$ (always)*

Lemma 1.2. *x is contact point of $B \iff \exists (x_n)_{n\in\mathbb{N}}$ with $x_n \in B$ and $\lim_{n\to\infty} x_n = x$.*

Proof. Let x be a contact point of B.

Direction $\Rightarrow$: Because $K_{\frac{1}{n}}(x) \cap B \neq \emptyset$, choose $X_n \in K_{\frac{1}{n}}(x) \cap B$. The sequence $(x_n)_{n\in\mathbb{N}}$ has property $d(x_n, x) < \frac{1}{n}$. Let $\varepsilon > 0$ be arbitrary. Choose $N \in \mathbb{N}$ sch that $N > \frac{1}{\varepsilon}$ (consider the Archimedean axiom). Then for $n \geq N$, $d(x_n, x) < \frac{1}{n} \leq \frac{1}{N} < \varepsilon$, hence $\lim_{n\to\infty} x_n = x$.

Direction $\Leftarrow$: Let $x = \lim_{n\to\infty} x_n$ and $x_n \in B$. Let $\varepsilon > 0$ be arbitrary and $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon \forall n \geq N$. Then $d(x_N, x) < \varepsilon$, hence

$$x_N \in \underbrace{K_\varepsilon(x) \cap B}_{\neq \emptyset}$$

So x is contact point of B. $\square$

Lemma 1.3. *It holds that $\forall B \subseteq X : \overline{B} = \overline{\overline{B}}$, hence $\overline{B}$ itself is closed.*

Proof. Show that $x \in \overline{B}$. Let $x \in \overline{\overline{B}}$.

$$\iff \forall \varepsilon > 0 : K_\varepsilon(x) \cap \overline{B} \neq \emptyset$$

Therefore let $\varepsilon > 0$ be arbitrary and $x \in \overline{\overline{B}}$.

Show that $K_\varepsilon(x) \cap B \neq \emptyset$.

Because $x \in \overline{\overline{B}} : \exists y \in \overline{B} : y \in K_{\frac{\varepsilon}{2}}(x)$. Because $y \in \overline{B} : \exists z \in B : z \in K_{\frac{\varepsilon}{2}}(y)$. Hence,

$$d(z, x) \leq \underbrace{d(z, y)}_{<\frac{\varepsilon}{2}} + \underbrace{d(y, x)}_{<\frac{\varepsilon}{2}} < \varepsilon$$

so $z \in K(x, \varepsilon) \cap B$. So x is contact point of $B \implies x \in \overline{B}$. □

Lemma 1.4. *Let X be a metric space.*

- $A_i \subseteq X$ *be closed* $\forall i \in I$. *Then* $A = \bigcap_{i \in I} A_i = \{x \in X \,|\, x \in A_i \forall i \in I\}$ *is closed itself.*
- $A_1, \ldots, A_n \subseteq X$ *are closed. Then* $\bigcup_{k=1}^{n} A_k$ *is closed in X.*
- φ *is closed, X is closed.*

Proof. See Analysis 1 course. □

Definition 1.10. *Let* $x \in X$ *is called accumulation point of set* $B \subseteq X$ *if* $\forall \varepsilon > 0 : (K_\varepsilon(x) \setminus \{x\}) \cap B \neq \emptyset$.

Remark 1.2. *Accumulation* points *only exist in the context of* sets. *Accumulation* values *only exist in the context of* sequences.

For example $(+1, -1, +1, -1, +1, \ldots)$ *has accumulation* values $+1$ *and* -1.

Lemma 1.5. *Let* $x \in X$ *is accumulation point on B* $\iff$ *every sphere* $K_\varepsilon(x)$ *contains infinitely many points of B.*

Proof. Direction $\Leftarrow$ is trivial.

Direction $\Rightarrow$: Choose $x_1 \in (K_1(x) \setminus \{x\}) \cap B$, hence $x_1 \neq x$, $x_1 \in B$ and $d(x_1, x) < 1$. Let $r_1 = 1$.

Inductive: choose $r_n = \min(\frac{1}{n}, d(x_{n-1}, x))$ and $x_n \in (K_{r_n}(x) \setminus \{x\}) \cap B$. Then $d(x_n, x) > 0$ (because $x_n \neq x$) where $d(x_n, x) < r_n < \frac{1}{n}$.

$$0 < d(x_n, x) < \frac{1}{n}$$

Furthermore, $d(x_n, x) < r_n \leq d(x_{n-1}, x)$. So $x_n \neq x_{n-1}$.

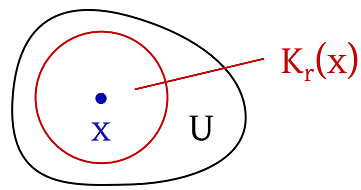

Figure 4: x is an inner point of U if $\exists r > 0 : K_r(x) \subseteq U$

Inductive: $x_n \neq x_{n-1} \neq x_{n-2} \neq \cdots \neq x_1$. Now consider arbitrary $\varepsilon > 0$ and N large enough such that $\frac{1}{N} < \varepsilon$.

Then it holds that $\forall n \geq N : 0 < d(x_n, x) < \frac{1}{n} \leq \frac{1}{N} < \varepsilon$. So $K_\varepsilon(x) \cap B$ contains infinitely many points $x_N, x_{N+1}, x_{N+2}, \ldots$. □

Definition 1.11. *Let $U \subseteq X$ and $x \in U$. We say x is an* inner point of U *if $\exists r > 0 : K_r(x) \subseteq U$. We let $\mathring{U} = \left\{ x \in U \,\middle|\, x \text{ is inner point of } U \right\}$ and call it* interior of *U (offenen Kern von U or das Innere von U). $O \subseteq X$ is called* open *(open set), if every point $x \in O$ is also an inner point of O. Hence $\mathring{O} = O$. Compare with image 4.*

Example 1.5. *Let $K_r(x)$ with $r > 0$ be an open sphere in X. Then $K_r(x)$ is an open set in X. Compare with image 5.*

Proof. Why? Let $y \in K_r(x)$. Show that y is an inner point of the sphere. $d(y, x) = s < r$. Define $r' = r - s > 0$.

Claim: $K_r'(y) \subseteq K_r(x)$.

Let $z \in K_{r'}(y)$, hence $d(z, y) < r'$. Then,

$$d(z, x) \leq \underbrace{d(x, y)}_{<r'} + \underbrace{d(y, z)}_{=s} < r' + s = r$$

So it holds that $z \in K_r(x)$ and therefore $K_{r'}(y) \subseteq K_r(x)$. □

Lemma 1.6. *Let $U \subseteq X$ be arbitrary. Then $\mathring{U} \subseteq X$ be an open set in X.*

Proof. Let $x \in \mathring{U}$, hence x is an inner point of U. Show that x is an inner point of $\mathring{U}$, also $\exists r > 0 : K_r(x) \subseteq \mathring{U}$.

Because $x \in \mathring{U}$, $r > 0$*exists*: $K_r(x) \subseteq U$. Claim: Every point $y \in K_r(x)$ is also an inner point of U. Obvious (previous example), because $r' > 0$ exists such that $K_{r'}(y) \subseteq K_r(x) \subseteq U$ so $y \in \mathring{U}$ and $K_r(x) \subseteq \mathring{U}$. □

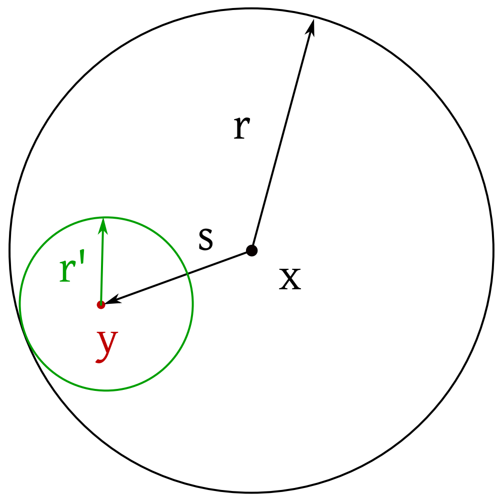

Figure 5: Let $K_r(x)$ with $r > 0$ be an open sphere in X. Then $K_r(x)$ is an open set in X.

Theorem 1.2. *Let X be a metric space.*

$$A \subseteq X \text{ is closed in } X \iff O = X \setminus A = A^C \text{ is open}$$

Proof. Let A be closed and $O + A^C$. We choose $x \in O$ and show that x is in the interior of O.

Assume the opoosite.

$$\forall \varepsilon > 0 : \underbrace{\neg\left(K_\varepsilon(x) \subseteq O\right)}_{\iff K_\varepsilon(x) \cap O^C \neq \emptyset}$$

where $O^C = A$.

Direction $\Leftarrow$. So x is contact point of A. Because A is closed, it holds that $x \in A$. This contradicts with $x \in O = A^C$. Thus O is open.

Direction $\Rightarrow$. Let $O = A^C$ be open and let x be a contact point of A. Show that $x \in A$.

Assume the opposite, hence $x \in A^C = O$ and O is open. So $\exists r > 0 : K_r(x) \subseteq O$, so $K_r(x) \cap A = \varnothing$ where $A = O^C$. Hence x is not a contact point of A.

So every contact point of A is also an element of A and A is closed. $\square$

Theorem 1.3. *Let X be a metric space. Then it holds that*

- *If $O_i \subseteq X$ is open in X $\forall i \in I$. Then also $O = \bigcup_{i \in I} O_i$ is open in X.*

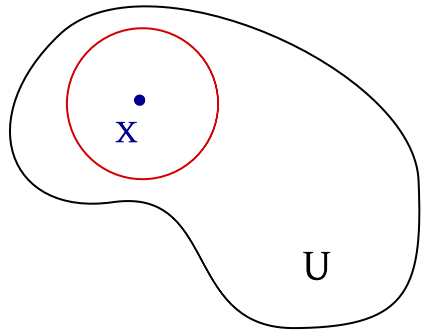

Figure 6: Neighborhood of x

- *If $O_1, O_2, \ldots, O_n$ is open in X, then $\bigcap_{k=1}^{n} O_k$ is open in X.*
- *X is open, $\emptyset$ is open.*

Proof. By Lemma 1.4, Theorem 1.2 and De Morgan's Laws:

$$\left(\bigcup_{i\in I} A_i\right)^C = \bigcap_{iinI} A_i^C$$

□

1.8 Topology

Definition 1.12. *Given a set X. If a subset $T \subseteq \mathcal{P}(X)$ is defined such that the elements $O \in T$ (hence $O \subseteq X$) satisfy the conditions of Theorem 1.3, then T is called* topology *on X. (X, T) is called topological space.*

The sets $O \in T$ are called open sets *in terms of T. The complements $A = O^C$ for $O \in T$ are called* closed sets.

Definition 1.13. *Let $x \in U \subseteq X$. We claim that U is a neighborhood of x, if $r > 0$ exists such that $x \in K_r(X) \subseteq U$*

See Figure 6

Remark 1.3. *$O \subseteq X$ is open iff O is neighborhood of every point $x \in O$.*

Definition 1.14. *Let X and Y be metric spaces and $x_0 \in X$. Let $f : X \to Y$ be given. We say f is continuous in x_0 if*

$$\forall \varepsilon > 0 \exists \delta > 0 : \forall x \in X : d_X(x, x_0) < \delta \implies d_Y(f(x), f(x_0)) < \varepsilon$$

Here, d_X is a metric on X and d_Y is a metric on Y.

This lecture took place on 2018/03/13.

Theorem 1.4. *Let X and Y be metric spaces. $f : X \to Y$. Let $x_0 \in X$ be given. Then the following statements are equivalent:*

1. *f is continuous in x_0*
2. *For every neighborhood V of $y_0 = f(x_0)$ it holds that $f^{-1}(V)$ is a neighborhood of x_0*
3. *For every sequence $(x_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} f(x_n) = f(x_0)$*

Proof. See Analysis 1. □

Definition 1.15. *Let $f : X \to Y$ is called continuous on X, if f is continuous in every point $x_0 \in X$.*

Theorem 1.5. *Let $f : X \to Y$ be given. Then f is continuous on $X \iff \forall$ open $O \subseteq Y : U = f^{-1}(O)$ open in X.*

Remark 1.4. *This characterization of continuity also works in topological spaces.*

Proof. Direction $\Rightarrow$.

Let f be continuous in X and let $O \subseteq Y$ be open. Let $U = f^{-1}(O)$ and choose $x_0 \in U$. Then $f(x_0) \in O$, hence O is a neighborhood of $f(x_0)$. By Theorem 1.4 (b), it follows that $U = f^{-1}(O)$ is a neighborhood of x_0.

Hence, U is neighborhood of every of its points, hence open in X.

Direction $\Leftarrow$.

Let the preimages of open sets be open and $x_0 \in X$ and $y_0 = f(x_0)$. Let V be a neighborhood of $y_0 = f(x_0)$, hence $\exists \varepsilon > 0 : K_\varepsilon(f(x_0)) \subseteq V$. Because $K_\varepsilon(f(x_0))$ is an open set, it holds that $f^{-1}(K_\varepsilon(f(x_0))) \in x_0$ is open in X.

Therefore, there exists $\delta > 0$ such that $K_\delta(x_0) \subseteq f^{-1}(K_\varepsilon(f(x_0))) \subseteq f^{-1}(V)$. Hence, $f^{-1}(V)$ is a neighborhood of x_0. Then by Theorem 1.4 (b), it follows that f is continuous in x_0 (chosen arbitrarily). Hence f is continuous on X. □

1.9 Variations of continuity notions

1.9.1 Uniform continuity

Definition 1.16. *Let $f : X \to Y$ be given. We call "f uniformly continuous on X" if*

$$\forall \varepsilon > 0 \exists \delta > 0 : \forall x, y \in X \land d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \varepsilon$$

Remark 1.5. *Compare it with the definition of "continuous in X":*

$$\forall x \in X \forall \varepsilon > 0 \exists \delta > 0 : \forall y \in X : d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \varepsilon$$

The difference is the location of the $\forall x \in X$ quantifier.

Every uniformly continuous map is continuous.

Example: $f : (0, \infty) \to (0, \infty)$ with $f(x) = \frac{1}{x}$ is continuous, but not continuously continuous.

1.9.2 Lipschitz continuity

Definition 1.17. *$f : X \to Y$ is called* Lipschitz continuous *with* Lipschitz constant $L \geq 0$ *if $\forall x, y \in X : d_Y(f(x), f(y)) \leq L \cdot d_X(x, y)$.*

Rudolf Lipschitz [1832–1903], University of Bonn

Theorem 1.6. *Every Lipschitz continuous function is uniformly continuous.*

Proof. For $\varepsilon > 0$, choose $\delta = \frac{\varepsilon}{L+1}$. Then it holds that $d_X(x, y) < \delta = \frac{\varepsilon}{L+1} \implies d_Y(f(x), f(y)) \leq L \cdot d_X(x, y) < \frac{L}{L+1} \cdot \varepsilon < \varepsilon$. □

- Most often $X \subseteq V$, $Y \subseteq W$. V and W are normed vector spaces and $d(x, y) = \left\| x - y \right\|$

1.10 Banach Fixed Point Theorem

Definition 1.18. *A Lipschitz continuous map $f : X \to X$ with Lipschitz constant $L < 1$ is called* contraction on *X. Compare with Figure 7*

Theorem 1.7 (Banach fixed-point theorem)**.** *Let $f : X \to X$ be a contraction and X be complete. Then there exists a uniquely defined $\hat{x} \in X$ such that $\hat{x} = f(\hat{x})$. $\hat{x}$ is called fixed point on f. Furthermore it holds that $x_0 \in X$ is arbitrary and $x_n = f(x_{n-1})$ for all $n \geq 1$. Compare with Figure 8.*

$$\lim_{n \to \infty} x_n = \hat{x}$$

Remark 1.6. *The following proof is a very common exam question.*

Proof. Let $x_0 \in X$ be arbitrary. x_n is constructed inductively by $x_n = f(x_{n-1})$ for all $n \geq 1$.

Claim. *$(x_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in X.*

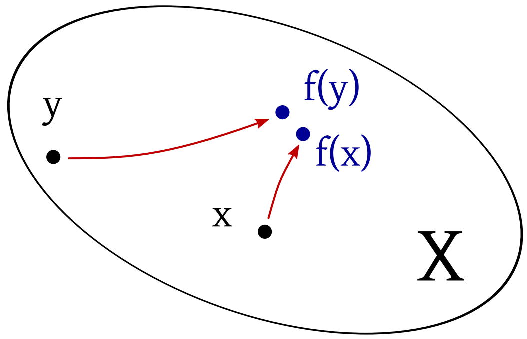

Figure 7: A contraction maps to points closer to each other

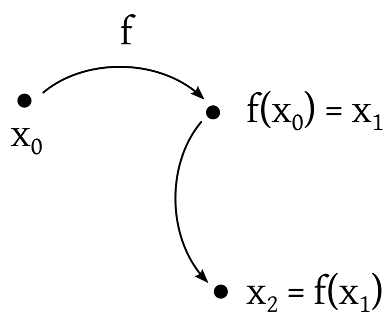

Figure 8: Banach's Fixed Point Theorem states that applying f iteratively gives a point coming closer and closer to the previous one

$d(x_n, x_{n+k}) \leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots + d(x_{n+k-1}, x_{n+k})$

by triangle inequality

$$
\begin{aligned}
&= d(x_n, x_{n+1}) + d(f(x_n), f(x_{n+1})) + d(f(x_{n+1}), f(x_n + 2)) + \cdots + d(f(x_{n+k-2}), f(x_{n+k-1})) \\
&\leq d(x_n, x_{n+1}) + L\left(d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots + d(x_{n+k-2}, x_{n+k-1})\right)
\end{aligned}
$$

this inequality is given by contraction

$$
\begin{aligned}
&= d(x_n, x_{n+1})(1+L) + L\left(d(f(x_n), f(x_{n+1})) + \cdots + d(f(x_{n+k-3}), f(x_{n+k-2}))\right) \\
&\leq d(x_n, x_{n+1})(1+L) + L^2\left[d(x_n, x_{n+1} + \cdots + d(x_{n+k-3}, x_{n+k-2})\right] \\
&\leq \cdots \leq d(x_n, x_{n+1})(1 + L + L^2 + \cdots + L^{k-1}) \\
&= d(f(x_{n-1}, f(x_n))\left(\sum_{j=0}^{k-1} L^j\right) \leq L d(x_{n-1}, x_n) \cdot \left(\sum_{j=0}^{k-1} L^j\right) \\
&\leq L^n d(x_0, x_1) \cdot \underbrace{\left(\sum_{j=1}^{k-1} L^j\right)}_{\leq \sum_{j=0}^{\infty} L^j = \frac{1}{1-L}} \\
&\leq \frac{L^n}{1-L} d(x_0, x_1)
\end{aligned}
$$

$$
d(x_n, x_{n+k}) \leq \frac{L^n}{1-L} d(x_0, x_1) \forall n \in \mathbb{N} \forall k \in \mathbb{N}_0
$$

with $0 \leq L < 1$.

$$
\begin{aligned}
\underbrace{\frac{L^n}{1-L}}_{>0} d(x_0, x_1) < \varepsilon \iff \\
L^n < \frac{\varepsilon}{d(x_0, x_1) + 1}(1 - L) \qquad (L > 0) \\
\iff n \underbrace{\ln L}_{<0} < \ln \frac{\varepsilon}{d(x_0, x_1) + 1}(1 - L) \\
\iff n > \frac{1}{\ln L} \ln \frac{\varepsilon}{d(x_0, x_1) + 1}(1 - L)
\end{aligned}
$$

Hence $(x_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in X. X is complete, hence $\exists \hat{x} \in X$: $\hat{x} = \lim_{n \to \infty} x_n$. Because $\hat{x} = \lim_{n \to \infty} x_{n+1} = \lim_{n \to \infty} f(x_n) = f(\hat{x})$ where the last equality is given by continuity of f. Therefore $\hat{x} = f(\hat{x})$ is a fixed point on f.

It remains to prove uniqueness:

Let $\tilde{x} = f(\tilde{x})$. Then it holds that $d(\hat{x}, \tilde{x}) = d(f(\hat{x}), f(\tilde{x})) \leq L d(\hat{x}, \tilde{x})$ with $L < 1$. If $d(\hat{x}, \tilde{x}) > 0$, then $1 \leq L$. This is a contradiction. Hence $d(\hat{x}, \tilde{x}) = 0$ must hold, hence $\hat{x} = \tilde{x}$. □

Remark 1.7.

- *The Fixed Point Theorem provides an algorithm for numeric computation of $\hat{x}$.*
- *It can reformulate problems $f(x) = 0$ (in $\mathbb{R}^n$) to*

$$
f(x) + x = g(x) = x
$$

- *Attention: The conditions of the Fixed Point Theorem cannot be changed to the structure*

$$d(f(x), f(y)) < L \cdot d(x, y) \wedge L \leq 1$$

or

$$d(f(x), f(y)) \leq L \cdot d(x, y) \wedge L < 1$$

This will be discussed in the practicals.

Lemma 1.7. *Let X be a complete metric space. Let $A \subseteq X$ be closed. Then (A, d) is itself a complete, metric space.*

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in A ($x_n \in A$). Then $(x_n)_{n\in\mathbb{N}}$ is also a Cauchy sequence in X. Because X is complete, there exists $\hat{x} = \lim_{n\to\infty} x_n$. Therefore $\hat{x}$ is a contact point of A. Because A is closed, it holds that $\hat{x} \in A$.

Therefore every Cauchy sequence in A has a limit point in A, hence A is complete. □

2 Compactness

2.1 Definition

Definition 2.1. *A metric space (X, d) is called* compact *if every sequence $(x_n)_{n\in\mathbb{N}}$ has a convergent subsequence.*

Specifically, this definition is called sequence compactness. The other definition defines compactness as closed and bounded subset of an Euclidean space. The latter definition only works for a subset of branches in mathematics. Therefore the generalization is recommended to be remembered.

Lemma 2.1. *Let X be a compact, metric space. Then X is complete.*

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in X. By compactness, it follows that $\exists (x_{n_k})_{k\in\mathbb{N}}$ with $\lim_{k\to\infty} x_{n_k} = \hat{x}$. Choose $\varepsilon > 0$ arbitrary and L large enough such that $k \geq L \implies d(x_{n_k}, \hat{x}) < \frac{\varepsilon}{2}$. Furthermore choose $N \in \mathbb{N}$ large enough such that $n, m \geq N \implies d(x_n, x_m) < \frac{\varepsilon}{2}$ (satisfied, because $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence). Choose $K \geq L$ and $n_k \geq N$. Let n_k be fixed this way. Then it holds $\forall n \geq N : d(x_n, \hat{x}) \leq d(x_n, x_{n_k}) + d(x_{n_k}, \hat{x}) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$. The first summand $\frac{\varepsilon}{2}$ results from the Cauchy sequence property, the second summand $\frac{\varepsilon}{2}$ results by convergence of (x_{n_k}). Hence $(x_n)_{n\in\mathbb{N}}$ is convergent with limit $\hat{x}$. □

2.2 Boundedness

Definition 2.2. *A metric space X is called* bounded *if there exists $M \geq 0$, such that $d(x, y) \leq M \forall x, y \in X$.*

It holds for arbitrary $x \in X$ that $\forall y \in X : y \in K_M(x)$. So, $X \subseteq K_M(x)$. On the contrary, let $X \subseteq \overline{K_M(x)}$ and let $y \in X$ and $z \in X$ be arbitrary. Then it holds that $d(y,z) \leq d(y,x) + d(x,z) \leq M + M = 2M$. Hence, X is bounded.

So, X is bounded $\iff \exists x \in X \wedge M \geq 0 : X \subseteq \overline{K_M(x)}$.

Lemma 2.2. *Every compact, metric space is also bounded.*

Proof. Assume X is unbounded.

We construct a sequence of points $(x_n)_{n\in\mathbb{N}}$ with $d(x_n, x_m) \geq 1 \forall n, m \in \mathbb{N}$ with $n \neq m$.

We use the following auxiliary result: Let $B = \bigcup_{j=1}^{n} K_1(z_j)$ for arbitrary $n \in \mathbb{N}$ and arbitrary $z_j \in X$. Then B is bounded. This result will be part of the practicals.

We construct $(x_n)_{n\in\mathbb{N}}$ inductively. Choose arbitrary $x_0 \in X$. Assume $(x_1, \ldots, x_{n-1})$ are already found. Then it holds that

$$\underbrace{X}_{\text{unbounded}} \not\subseteq \underbrace{\bigcup_{j=1}^{n-1} K_1(x_j)}_{\text{bounded}}$$

hence $\exists x_n \in X \setminus \bigcup_{j=1}^{n-1} K_1(x_j)$. Because $x_n \notin K_1(x_j)$ for $j = 0, \ldots, n-1$ it holds that $d(x_n, x_j) \geq 1 \forall j < n$. We get $(x_n)_{n\in\mathbb{N}}$ with $d(x_n, x_m) \geq 1 \forall n \in \mathbb{N} \forall m < n$, hence $m \neq n$. Because $d(x_n, x_m) \geq 1$, i.e. $(x_n)_{n\in\mathbb{N}}$ does not contain any Cauchy sequence as subsequence, $(x_n)_{n\in\mathbb{N}}$ does not have a convergent subsequence. Therefore X is not compact. □

This lecture took place on 2018/03/15.

Every compact metric space is bounded. Every compact metric space is complete. In $\mathbb{C}$ ($\mathbb{R}^n$) it holds that $A \subseteq \mathbb{C}$ is closed. Then A with metric $d(x,y) = \left|x - y\right|$ is complete as metric space.

If A is additionally bounded, then A is compact (see course Analysis 1, Bolzano-Weierstrass).

Attention! Let V be an infinite-dimensional, complete, normed vector space. For example, $V = C([a,b], \mathbb{R}) = \left\{ f : [a,b] \to \mathbb{R} \,\middle|\, f \text{ is continuous in } [a,b] \right\}$ with norm $\left\|f\right\|_\infty = \max\left\{\left|f(x)\right| : x \in [a,b]\right\}$ and metric $\left\|f - g\right\|_\infty = \max\left\{\left|f(x) - g(x)\right| : x \in [a,b]\right\}$. $C([a,b], \mathbb{R})$ is a complete, normed vector space. It holds that $\overline{K_1(0)}$ is not compact in $C([a,b], \mathbb{R})$ (i.e. V, for every infinite-dimensional vector space).

Again: do not remember "compactness" not as closed and bounded, as this only holds in the finite-dimensional case.

In the last proof, we have shown: If a sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in X$ and $d(x_n, x_m) \geq 1$ (or $\geq \varepsilon$) $\forall n \neq m \implies X$ is not compact.

2.3 Total boundedness

Definition 2.3. *X is called totally bounded, if for every $\varepsilon > 0$, finitely many points $X_1^\varepsilon, X_2^\varepsilon, \ldots, X_{N(\varepsilon)}^\varepsilon$ such that $X \subseteq \bigcup_{i=1}^{N(\varepsilon)} K_\varepsilon(X_i^\varepsilon)$.*

Hence, for every $x \in X$, there exists some X_j^ε such that $d(X, X_j^\varepsilon) < \varepsilon$.

Remark 2.1 (For the practicals)**.** *Let X be totally bounded, then there does not exist some sequence $(x_n)_{n\in\mathbb{N}}$ with $d(x_n, x_m) \geq \varepsilon \forall n \neq m$. It holds, that X is compact if and only if X is totally bounded and complete.*

2.4 Compactness, continuity and openness

Theorem 2.1. *Let $f : X \to Y$ be continuous. Let X be compact. Then image $f(X) \subseteq Y$ is also compact.*

Be aware, that this proof is a common exam question and students often begin with the wrong order.

Proof. Let $(y_n)_{n\in\mathbb{N}}$ be an arbitrary sequence in $f(X)$. Show that $(y_n)_{n\in\mathbb{N}}$ has a convergent subsequence. Because $y_n \in f(X)$, there exists at least one x_n with $y_n = f(x_n)$. Then $(x_n)_{n\in\mathbb{N}}$ is a sequence in X, X is compact, hence there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ with $\lim_{k\to\infty} x_{n_k} = \hat{x} \in X$. Because f is continuous, it holds that $\lim_{k\to\infty} f(x_{n_k}) = \lim_{k\to\infty} y_{n_k} = f(\hat{x}) =: \hat{y}$. So $(y_n)_{n\in\mathbb{N}}$ has a convergent subsequence. Hence $f(X) \subseteq Y$ is compact. □

Theorem 2.2 (Conclusion)**.** *Let X be compact, $f : X \to \mathbb{R}$ continuous on X. Then there exists $\underline{x}$ and $\overline{x} \in X$, such that*

$$f(\underline{x}) \leq f(x) \leq f(\overline{x}) \qquad \forall x \in X$$

Hence, f has a maximum and a minimum.

Proof. $f(X) \subseteq \mathbb{R}$ is compact (Theorem 2.1), hence $f(X)$ is bounded and complete, hence closed in $\mathbb{R}$. There exists $\xi \in \mathbb{R}$ with $\xi = \sup f(X)$, because $f(X)$ is complete and ξ is a contact point of $f(X)$, it holds that $\xi \in f(X)$, hence $\exists \overline{x} \in X$: $\xi = f(\overline{x})$. Furthermore, ξ is an upper bound of $f(X) \to f(x) \leq \xi = f(\overline{x}) \forall x \in X$.

For $\underline{x}$, it works the same way. □

Theorem 2.3. *Let $f : X \to Y$ is continuous on X and X is compact. Then f is uniformly continuous on X.*

Indirect proof. Assume X is compact, $f : X \to Y$ is continuous, but not uniformly continuous. Uniform continuity:

$$\forall \varepsilon > 0 \exists \delta > 0 : \forall x, y \in X : d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \varepsilon$$

Not uniformly continuous:

$$\exists \varepsilon > 0 \forall \delta_n = \frac{1}{n}(n \in \mathbb{N}) \exists x_n, y_n \in X : d_X(x_n, y_n) < \frac{1}{n} \wedge d_Y(f(x_n), f(y_n)) \geq \varepsilon$$

Now choose some (x_n) and (y_n). We will use a specific ε later. Because X is compact, there exists a convergent subsequence of $(x_n)_{n\in\mathbb{N}}$, hence $\lim_{k\to\infty} x_{n_k} = \hat{x}$. The sequence $(y_{n_k})_{k\in\mathbb{N}}$ has a convergent subsequence itself:

$$\lim_{l\to\infty} y_{(n_k)_l} = \hat{y}$$

Because $(x_{n_k})_{n\in\mathbb{N}}$ is convergent, the subsequence $(x_{(n_k)_l})_{l\in\mathbb{N}}$ converges towards the same limit $\hat{x}$.

$$\tilde{x}_l := x_{n_{k_l}} \qquad \tilde{y}_l := y_{n_{k_l}}$$

because $l \leq x_{n_l}$ and

$$d_X(\tilde{x}_l, \tilde{y}_l) = d_X(x_{n_{k_l}}, y_{n_{k_l}}) \underbrace{<}_{\text{by assumption}} \frac{1}{n_{k_l}} \leq \frac{1}{l}$$

Claim. *For $\hat{x} = \lim_{l\to\infty} \tilde{x}_l$ and $\hat{y} = \lim_{l\to\infty} \tilde{y}_l$, it holds that $\hat{x} = \hat{y}$.*

Proof. Let $\varepsilon' > 0$ be arbitrary, l large enough such that

- $\frac{1}{l} < \frac{\varepsilon'}{3}$
- $d_X(\tilde{x}_l, \hat{x}) < \frac{\varepsilon'}{3}$
- $d_X(\tilde{y}_l, \hat{y}) < \frac{\varepsilon'}{3}$

Therefore it holds that

$$d_X(\hat{x}, \hat{y}) \leq d_X(\hat{x}, \tilde{x}_l) + d_X(\tilde{x}_l, \tilde{y}_l) + d_X(\tilde{y}_l, \hat{y}) < \frac{\varepsilon'}{3} + \frac{1}{l} + \frac{\varepsilon'}{3} < \varepsilon'$$

Therefore it holds that $d_X(\hat{x}, \hat{y}) = 0$, hence $\hat{x} = \hat{y}$. □

Because f is continuous and $\tilde{x}_l \to \hat{x}$ and $\tilde{y}_l \to \hat{x}$, there exists $l \in \mathbb{N}$ such that

$$d_Y(f(\tilde{x}_l), f(\hat{x})) < \frac{\varepsilon}{2}$$

and also

$$d_Y(f(\tilde{y}_l), f(\hat{x})) < \frac{\varepsilon}{2}$$

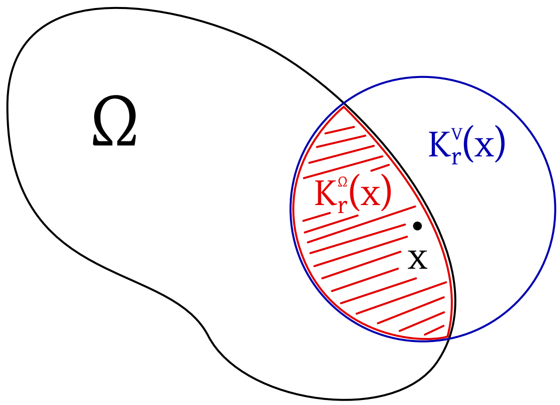

Figure 9: Subsets of $(\mathbb{R}^n, \|\cdot\|)$ as metric spaces

where ε is the epsilon from the very beginning of the proof.

$$\implies d_Y(f(\tilde{x}_l), f(\hat{x})) + d_Y(f(\tilde{y}_l), f(\hat{x})) < \varepsilon$$

This contradicts to

$$d_Y(f(\tilde{x}_l), f(\tilde{y}_l)) = d_Y(f(x_{n_{k_l}}), f(y_{n_{k_l}})) \geq \varepsilon$$

Hence, f is uniformly continuous. $\square$

Subsets of $(\mathbb{R}^n, \|\cdot\|)$ (or $(V, \|\cdot\|)$) as metric spaces.

We consider $\Omega \subseteq V$ where V is a normed vector space. (Ω, d) is $d(x, y) = \left\|x - y\right\|$ is a metric space.

$$K_r^\Omega(x) = \left\{ y \in \Omega \,\middle|\, \left\|y - x\right\| < r \right\}$$

is a sphere with center x and radius r in Ω.

$$K_r^V(x) = \left\{ y \in V \,\middle|\, \left\|y - x\right\| < r \right\}$$

obvious: $K_r^\Omega(x) = \Omega \cap K_r^V(x)$.

Lemma 2.3. *Let $O' \subseteq \Omega \subseteq V$.*

Then it holds that O' is open in Ω $\iff$ there exists $O \subseteq V$ is open in V such that $O' = O \cap \Omega$.

Proof. $\Rightarrow$ Let $O' \subseteq \Omega$ be open in Ω and $x \in O'$ be arbitrary. Then there exists $r(x) > 0 : x \in K^{\Omega}_{r(x)}(x) = K^{V}_{r(x)}(x) \cap \Omega \subseteq O'$. Then it holds that

$$O' = \bigcup_{x\in O'} = \{x\} \subseteq \bigcup_{x\in O'} K^{\Omega}_{r(x)}(x) = \left(\bigcup_{x\in O'} (K^{V}_{r(x)}(x)\right) \cap \Omega) = \underbrace{\left(\bigcup_{x\in O'} K^{V}_{r(x)}(x)\right)}_{=O\subseteq V \text{ is open in } V} \cap \Omega \subseteq O'$$

So every $\subseteq$ in this inclusion chain is actually an equality. So $O' = O \cap \Omega$.

$\Leftarrow$ Let $O' = O \cap \Omega$ and $x \in O'$ be chosen arbitrarily. Because $x \in O$ and O is open in V.

$$\exists r > 0 : K^{V}_{r}(x) \subseteq O \implies \underbrace{K^{V}_{r}(x) \cap \Omega}_{=K^{\Omega}_{r}(x)} \subseteq O \cap \Omega = O'$$

So O' is open in Ω.

$\square$

Remark 2.2. *$A' \subseteq \Omega$ is closed in $\Omega \iff \exists A \subseteq V$ closed in V with $A' = A \cap \Omega$.*

Remark 2.3. *Let T be an arbitrary topological space with topology τ on T (a system of open sets). Furthermore let $\Omega \subseteq T$.*

Then Ω itself is a topological space with $O' \subseteq \Omega$ is open $\iff \exists O \subset T$ open in T with $O' = O \cap \Omega$.

Also called "subspace topology", "trace topology" or "relative topology".

Attention!

$$O' \subseteq \Omega \text{ open in } \Omega \implies O' \text{ open in } V$$

does *not* hold in general.

Example 2.1.

$$\Omega = [0,1] \cap [0,1)$$

$K_{\frac{1}{2}}(p) \cap \Omega$ is open in Ω but not open in $\mathbb{R}^2$.

Analogously,

$$A' \subseteq \Omega \text{ is closed} \implies A' \text{ closed in } V$$

does *not* hold in general.

Remark 2.4. *K is compact in $\Omega \implies K$ is compact in V*

Let $(x_n)_{n\in\mathbb{N}}$ is a sequence in K. Compactness $\implies \exists (x_{n_k})_{k\in\mathbb{N}} : x_{n_k} \to \hat{x}$ for $k \to \infty$ and $K \subseteq \Omega \subseteq V$.

Then $(x_n)_{n\in\mathbb{N}}$ also has a convergent subsequence in V.

2.5 Normed vector spaces

Definition 2.4. *Let V be a vector space and $\|\cdot\|_1$ and $\|\cdot\|_2$ are normed on V. We say, $\|\cdot\|_1$ is equivalent to norm $\|\cdot\|_2$, if $0 < m \leq M$ exist such that*

$$m\,\|v\|_1 \leq \|v\|_2 \leq M\,\|v\|_1 \;\forall v \in V$$

Remark 2.5. *Equivalence of norms is an equivalence relation.*

reflexivity $\|\cdot\|_1$ *is equivalent to* $\|\cdot\|_1$ *with* $m = M = 1$.

symmetry

$$m\,\|v\|_1 \leq \|v\|_2 \implies \|v\|_1 \leq \frac{1}{m}\,\|v\|_2 \wedge \|v\|_2 \leq M \cdot \|v\|_1 \implies \frac{1}{M}\,\|v\|_2 \leq \|v\|_1$$

$$\implies \underbrace{\frac{1}{M}}_{m'}\,\|v\|_2 \leq \|v_1\| \leq \underbrace{\frac{1}{m}}_{M'}\,\|v\|_2$$

hence the equivalence relations of norms are symmetrical.

transitivity *Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be equivalent. Let $\|\cdot\|_2$ and $\|\cdot\|_3$ be equivalent.*

$$m \cdot \|v\|_1 \leq \|v\|_2 \leq M\,\|v\|_1 \;\forall v \in V$$
$$m' \cdot \|v\|_2 \leq \|v\|_3 \leq M'\,\|v\|_2 \;\forall v \in V$$
$$\implies m \cdot m'\,\|v\|_1 \leq m'\,\|v\|_2 \leq \|v\|_3 \leq M'\,\|v\|_2 \leq M \cdot M'\,\|v\|_1$$

This lecture took place on 2018/03/20.

Addendum:

- Let $(x_n)_{n\in\mathbb{N}}$ be in (X, d), then it holds that

$$\underbrace{x = \lim_{n\to\infty} x_n}_{\text{in } X} \iff \underbrace{\lim_{n\to\infty} d(x_n, x) = 0}_{\text{in } \mathbb{R}}$$

$$(\iff \lim_{n\to\infty} \|x_n - x\| = 0 \text{ in normed vector spaces } V)$$

- Reversed triangle inequality: Let V be a normed vector space. Let $x, y \in V$.

$$\|x\| = \left\|x - y + y\right\| \leq \left\|x - y\right\| + \left\|y\right\|$$

  Hence,

$$\|x\| - \left\|y\right\| \leq \left\|x - y\right\|$$

By exchanging x and y,

$$\|y\| - \|x\| \leq \|x - y\|$$

Hence, it holds that

$$\big|\|x\| - \|y\|\big| \leq \|x - y\|$$

- Define the map $n : V \to [0, \infty)$ on $(V, \|\cdot\|)$ with $n(x) = \|x\|$. Then n is continuous on V because

$$|n(x_1) - n(x_2)| = |\|x_1\| - \|x_2\|| \leq \|x_1 - x_2\|$$

  Hence, n is Lipschitz continuous with constant 1.

Regarding the equivalence of norms:

Lemma 2.4. *Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be equivalent norms on V. Then it holds that*

1. $\lim_{n\to\infty} \|x_n - x\|_1 = 0 \iff \lim_{n\to\infty} \|x_n - x\|_2 = 0$, *hence $(x_n)_{n\in\mathbb{N}}$ is convergent with limit x in regards of $\|\cdot\|_1 \iff (x_n)_{n\in\mathbb{N}}$ is convergent with limit x in regards of $\|\cdot\|_2$.*
2. *$O \subseteq V$ is open in regards of $\|\cdot\|_1 \iff O$ is open in regards of $\|\cdot\|_2$, hence $\tau_1 = \tau_2$ (topologies are equivalent).*
3. *$K \subseteq V$ is compact in regards of $\|\cdot\|_1 \iff K$ is compact in regards of $\|\cdot\|_2$.*

Proof. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent, hence $\exists m, M > 0 : m\|x\|_1 \leq \|x\|_2 \leq M\|x\|_1 \; \forall x \in V$.

1. Let $\varepsilon > 0$ and $\lim_{n\to\infty} \|x_n - x\|_1 = 0$. Choose $N \in \mathbb{N}$ such that $n \geq N \implies \|x_n - x\|_1 < \frac{\varepsilon}{M}$. For those n it holds that

$$\|x_n - x\|_2 \leq M\|x_n - x\|_1 < \frac{\varepsilon}{M} \cdot M = \varepsilon$$

   Hence, $\lim_{n\to\infty} \|x_n - x\|_2 = 0$.
2. $K_r^2(x) = \left\{ y \in V \,\middle|\, \|y - x\|_2 < r \right\}$. For $y \in K_r^2(x)$ it holds that

$$m\|y - x\|_1 \leq \|y - x\|_2 < r$$

   hence,

$$\|y - x\|_1 < \frac{r}{m} \implies y \in K^1_{\frac{r}{m}}(x)$$

   hence $K_r^2(x) \subseteq K^1_{\frac{r}{m}}(x)$. Let $y \in K^1_{\frac{r}{M}}(x)$. Then it holds that,

$$\|y - x\|_2 \leq M\|y - x\|_1 < M \cdot \frac{r}{M} = r$$

hence $y \in K_r^2(x)$. $\implies K^1_{\frac{r}{M}}(x) \subseteq K_r^2(x)$. Now let O be open in regards of $\|\cdot\|_2$, hence

$$\forall x \in O \exists r > 0 : K_r^2(x) \subseteq O \implies K^1_{\frac{r}{M}}(x) \subseteq K_r^2(x) \subseteq O$$

so O is open in regards of $\|\cdot\|_1 \implies O$ is open in regards of $\|\cdot\|_2$ analogously.

3. Let K be compact in regards of $\|\cdot\|_1$ and $(x_n)_{n\in\mathbb{N}}$ be a sequence in K. Then there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ with $\left\|x_{n_k} - x\right\|_1 \to 0$ for $k \to \infty$ $\xRightarrow{\text{by the first property}} \left\|x_{n_k} - x\right\|_2 \to 0$. Hence $(x_{n_k})_{k\in\mathbb{N}}$ is also a convergent subsequence in regards of $\|\cdot\|_2$.

$\square$

Remark 2.6 (Proven in the practicals). *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^k$*

$$\|x\|_\infty = \max\left\{\left|x^i\right| \,\middle|\, i = 1, \ldots, n\right\}$$

$$x = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^k \end{bmatrix}$$

It holds that $\lim_{n\to\infty} \|x_n - x\|_\infty = 0 \iff \lim_{n\to\infty} \left|x_n^i - x^i\right| = 0$ *for all* $i \in \{1, \ldots, k\}$.

Theorem 2.4 (Bolzano-Weierstrass theorem in $\mathbb{R}^k$). *Let $K \subseteq \mathbb{R}^k$ be closed and bounded. Then K is compact in* $(\mathbb{R}^k, \|\cdot\|_\infty)$.

Proof. Let $\|x\|_\infty \leq M \forall x \in K \iff \left|x^i\right| \leq M \forall x \in K$ and $i \in \{1, \ldots, k\}$. Choose $(x_n)_{n\in\mathbb{N}}$ an arbitrary sequence in K $(x_n^i)_{n\in\mathbb{N}}$ is a bounded sequence in $\mathbb{R}$. Because $(x_n^1)_{n\in\mathbb{N}}$ is bounded, there exists a convergent subsequence $\left(x^1_{n_{l_1}}\right)_{l_1\in\mathbb{N}}$

$$\lim_{l_1\to\infty} x^1_{n_{l_1}} = x^1$$

Consider $(x^2_{n_{l_1}})_{l_1\in\mathbb{N}}$, a subsequence of a bounded sequence, hence bounded itself. By the Bolzano-Weierstrass theorem in $\mathbb{R}$, there exists a convergent subsequence $(x^2_{n_{l_{1 l_2}}})_{l_2\in\mathbb{N}}$ with $\lim_{l_2\to\infty} x^2_{n_{l_{1 l_2}}} = x^2$. Consider $x^1_{n_{l_{1 l_2}}}$ as subsequence of $x^1_{n_{l_1}}$ is already convergent, hence $\lim_{l_2\to\infty} x^1_{n_{l_{1 l_2}}} = x^1$. Furthermore, up to index i, it holds that:

$$\lim_{l_k\to\infty} x_{n_{l_{1 l_2 \ldots l_k}}} = x^i \qquad \text{for } i = 1, \ldots, k$$

Hence, with $\tilde{x_{l_k}} = x_{n_{l_{1 l_2 \ldots l_k}}}$ gives a subsequence of x_n, converging by each coordinate. Thus,

$$\lim_{l_k\to\infty} \left\|\tilde{x}_{l_k} - x\right\|_\infty = 0$$

Because $\tilde{x}_{l_n} \in K$ and K be closed, it holds that $x \in K$. Hence K is compact. $\square$

Theorem 2.5 (Norm equivalence in $\mathbb{R}^k$)**.** *In $\mathbb{R}^k$, all norms are equivalent.*

Proof. We show: Let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^n$. Then $\|\cdot\|$ is equivalent to $\|\cdot\|_\infty$. By transitivity of norm equivalence, two arbitrary norms are equivalent to each other.

1. Let $(e_1, e_2, \dots, e_k)$ be the canonical basis in $\mathbb{R}^k$.

$$x = \begin{bmatrix} x^1 \\ \vdots \\ x^k \end{bmatrix} = \sum_{j=1}^{k} x^j e_j$$

Furthermore let $M' = \max\left\{\left\|e_j\right\| : j = 1, \dots, k\right\}$ with $\left\|e_j\right\| \neq 0$ and $M' > 0$. Then it holds that

$$\|x\| = \left\|\sum_{j=1}^{k} x^j e_j\right\| \leq \sum_{j=1}^{k} \left\|x^j e_j\right\| = \sum_{j=1}^{k} \left|x^j\right| \left\|e_j\right\| \leq M' \sum_{j=1}^{k} \underbrace{\left|x_j\right|}_{\leq \|x\|_\infty} \leq \underbrace{M' \cdot k}_{M} \|x\|_\infty = M \|x\|_\infty$$

2. We consider $\nu : \mathbb{R}^k \to [0, \infty)$. $\nu(x) = \|x\|$ as map on $(\mathbb{R}^k, \|\cdot\|_\infty)$.

Claim. *ν is continuous on* $(\mathbb{R}^k, \|\cdot\|_\infty)$.

Proof. Show that,

$$\left|\nu(x) - \nu(y)\right| = \left|\|x\| - \left\|y\right\|\right| \underbrace{\leq}_{\text{reversed triangle ineq.}} \left\|x - y\right\| \underbrace{\leq}_{\text{because of (1)}} M\left\|x - y\right\|$$

Hence ν is Lipschitz continuous. □

We consider $S_\infty^{k-1} = \left\{x \in \mathbb{R}^k\right\} \|x\|_\infty = 1 = \text{boundary}(K_1^\infty(0)$. S_∞^{k-1} is bounded. Let $(x_n)_{n\in\mathbb{N}}$ is a sequence in S_∞^{k-1} with $x = \lim_{n\to\infty} x_n$. Because $n(x) = \|x\|_\infty$ is continuous, it holds that

$$\lim_{n\to\infty} \underbrace{\|x_n\|_\infty}_{=1} = \underbrace{\|x\|}_{=1}$$

Hence $x \in S_\infty^{k-1}$. Hence, S_∞^{k-1} is closed in $(\mathbb{R}^k, \|\cdot\|_\infty)$. Hence S_∞^{k-1} is compact in $(\mathbb{R}^k, \|\cdot\|_\infty)$, $\nu : S_\infty^{k-1} \to [0, \infty)$, with S_∞^{k-1} compact, is continuous. Has a minimum n on S_∞^{k-1}. Thus there exists $\overline{x} \in S_\infty^{k-1}$: $\underbrace{m}_{>0} = \left\|\underbrace{\overline{x}}_{\neq 0}\right\| \leq$

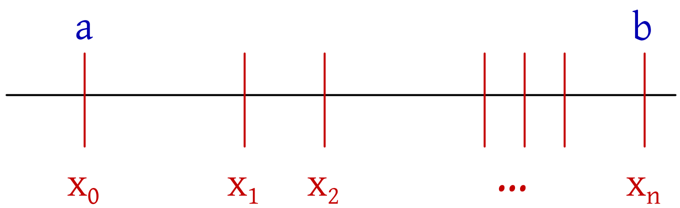

Figure 10: Illustration of a partition

$\|x\| \, \forall x \in S_\infty^{-1}$. Let $x \in \mathbb{R}^k$ be arbitrary with $x \neq 0$. Then it holds that $\frac{x}{\|x\|_\infty} \in S_\infty^{k-1}$ and it holds that

$$m \leq \left\| \frac{x}{\|x\|_\infty} \right\| = \frac{1}{\|x_\infty\|} \|x\| \implies m \|x\|_\infty \leq \|x\|$$

Inequality also holds true for $x = 0$.

$\square$

3 Integration calculus

3.1 Partitions and refinements

Definition 3.1. *Let $a < b$ with $a, b \in \mathbb{R}$. We consider functions of $[a, b]$. We call $(x_j)_{j=0}^n$ a* partition *of $[a, b]$ if $a = x_0 < x_1 < x_2 < \cdots < x_n = b$. x_j decomposes $[a, b]$ in subintervals (x_{j-1}, x_j). $\varphi : [a, b] \to \mathbb{R}$ is called* step function *in $[a, b]$ in regards of partition $(x_j)_{j=0}^n$ if $\varphi|_{(x_{j-1}, x_j)} = c_j$, so constant for $j = 1, \ldots, n$.*

φ is called step function *in $[a, b]$ if there exists a partition such that φ is a subsequence.*

$$\tau[a, b] = \{\varphi : [a, b] \to \mathbb{R} : \varphi \text{ is subsequence}\}$$

- *Let $(\xi_i)_{i=0}^m$ be a partition of $[a, b]$ and $(x_j)_{j=0}^n$ is a partition as well. Then we call $(\xi_i)_{i=0}^m$ a* refinement *of $[a, b]$ and $(x_j)_{j=1}^n$ as well. Then $(\xi_i)_{i=0}^n$ is a refinement of $(x_j)_{j=0}^k$ if $\{x_0, x_1, \ldots, x_n\} \subseteq \{\xi_0, \xi_1, \ldots, \xi_m\}$*

  Compare with Figure 11. Functions values in boundaries x_{j-1} and x_j do not have any constraints and will be relevant for an integral. A φ can be a step function in terms of many, various partitions.

Lemma 3.1. *Let $\varphi \in \tau[a, b]$ be a step function in terms of partition $(x_j)_{j=0}^n$ and let $(x_i)_{i=0}^n$ be a refinement of $(x_j)_{j=0}^n$ in terms of $(x_i)_{i=0}^m$.*

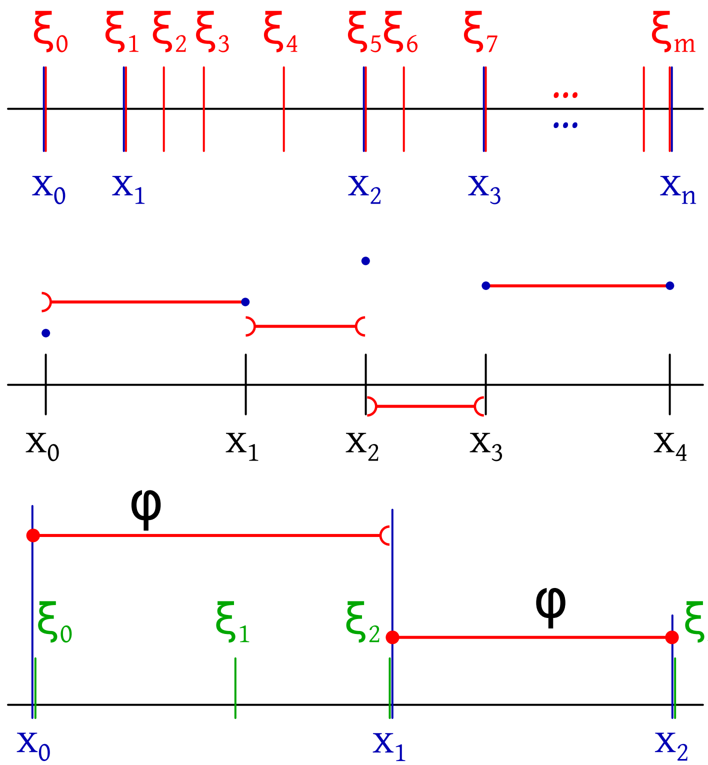

Figure 11: (top) Refinement (middle) function values in points x_j are unrestricted (bottom) step functions on a refinement

Proof. Refinement: For every $j \in \{0, \dots, n\}$ there exists $i_j \in \{0, \dots, m\}$ such that $X_j = \xi_{i_j}$. $i_0 = 0, i_n = m$. $i_{j-1} < i_j$.

Let $i \in \{1, \dots, m\}$. Then there exists a uniquely determined $j \in \{1, \dots, n\}$ such that $\xi_{i_{j-1}} < \xi_i \leq \xi_j$ Compare with Figure 12.

Then it holds that $(\xi_{i-1}, \xi_i) \subseteq \underbrace{(\xi_{i_{j-1}})}_{=(x_{j-1}, x_j)}, \xi_{i_j})$ and $\varphi|_{(\xi_{i-1}, \xi_j)} = c_j =$ const. So φ is a subsequence in regards of $(\xi_i)_{i=0}^m$. $\square$

Definition 3.2. *Let $\varphi \in \tau[a, b]$ in terms of partition $(X_j)_{j=0}^n$ with $\varphi|_{(X_{j-1}, X_j)} = c_j$ and $\triangle X_j = X_j - X_{j-1} > 0$ for $g = 1, \dots, n$. Then we define ...*

$$\int_a^b \varphi \, dx = \sum_{j=1}^n c_j \triangle x_j$$

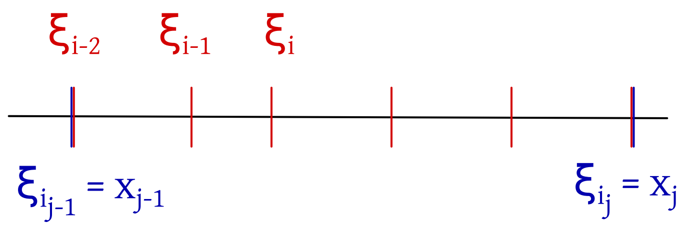

Figure 12: ξ on a refinement x_{i_j}

is called integral of φ *in terms of partition* $(x_j)_{j=0}^n$

This lecture took place on 2018/03/22.

Step function φ. $\varphi|_{x_{j-1},x_j} = c_j$

$$\delta x_j = x_j - x_{j-1}$$

$$\int_a^b \varphi \, dx = \sum_{j=1}^n c_j \cdot \delta x_j$$

Lemma 3.2. *Let $(x_i)_{j=0}^n$ be a partition of $[a,b]$ and $(\xi_i)_{i=0}^m$ be a refinement of $(x_j)_{j=0}^n$. Furthermore let φ be a subsequence with respect to $(x_j)_{j=0}^n$ (so also with respect to $(\xi_j)_{i=0}^m$). Then the integrals of φ with respect to $(x_j)_{j=0}^n$ and $(\xi_i)_{i=0}^m$ are equal.*

Proof. There exist indices i_j for $j = 0, n$ such that $x_j = \xi_{ij}$.

$$i_0 = 0 \qquad i_n = m \qquad i_{j-1} < i_j$$

$$\delta x_j = x_j - x_{j-1} = \xi_{i_j} - \xi_{i_{j-1}} = \xi_{i_j} - \xi_{i_{j-1}} = \underbrace{\sum_{i=i_{j-1}+1}^{i_j}}_{\text{telescoping sum}} (\xi_i - \xi_{i-1}) = \sum_{i=i_{j-1}+1}^{i_j} \delta\xi_i$$

$$\varphi|_{x_{j-1},x_j} = c_j \implies \varphi|_{(\xi_{i-1},\xi_i)} = c_j \text{ for } i = i_{j-1}+1, \dots, i_j$$

$$\tilde{c}_i = \varphi|_{(\xi_{i-1},\xi_i)}$$

$$\underbrace{\sum_{i=1}^m \tilde{c}_i \delta\xi_i}_{\text{integral of } \varphi \text{ w.r.t } (\xi_i)_{i=0}^m} = \sum_{j=1}^n \sum_{i=i_{j-1}+1}^{i_j} \tilde{c}_i \delta\xi_i = \sum_{j=1}^n c_j \underbrace{\sum_{i=i_{j-1}+1}^{i_j} \delta\xi_i}_{=x_j} = \sum_{j=1}^n c_j \delta x_j$$

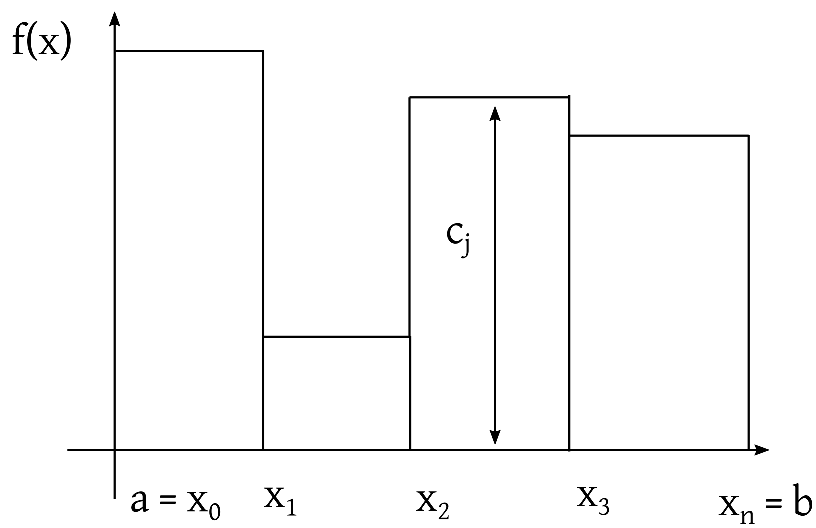

Figure 13: Integral of a step function as sum of areas of rectangles

This is the integral of φ with respect to $(x_j)_{j=0}^n$. □

Lemma 3.3. *Let φ be a step function with respect to $(x_j)_{j=0}^n$ and $(w_i)_{i=0}^L$. Then the integrals of φ with respect to $(x_j)_{j=0}^n$ and with respect to $(w_l)_{l=0}^L$ equal.*

Proof. Let $\{\xi_i \,|\, i = 1, \ldots, m\} = \left\{x_j \,\middle|\, j = 0, \ldots, n\right\} \cup \{w_l \,|\, l = 0, \ldots, L\}$ with $\xi_0 = a$, $\xi_m = x_n = w_L = b$ and $\xi_{i-1} < \xi_i$ for $i = 1, \ldots, m$. Then $(\xi_i)_{i=0}^m$ is a refinement of $(x_j)_{j=0}^n$ as well as $(w_l)_{l=0}^L$. By Lemma 3.2, the integral of φ with respect to $(x_j)_{j=0}^n$ = integral of φ with respect to $(\xi_i)_{i=1}^m$ = integral of φ with respect to $(w_l)_{l=0}^L$. Here we discard the statement "with respect to $(x_j)_{j=0}^n$". □

Lemma 3.4. *Let f, g be step functions on $[a, b]$. $f, g \in \tau[a, b]$.*

- *for $\alpha, \beta \in \mathbb{R}$, let $\alpha f + \beta g \in \tau[a, b]$ and*

$$\int_a^b (\alpha f + \beta g)\, dx = \alpha \int_a^b f\, dx + \beta \int_a^b g\, dx$$

  Hence, the integral is linear on $[a, b]$. $\tau[a, b]$ is a vector space.

- *$f \leq g$ in $[a, b]$, then $\int_a^b f\, dx \leq \int_a^b g\, dx$ (monotonicity).*

- $\left|\int_a^b f\,dx\right| \leq \int_a^b \left|f\right|\,dx$ *(*$\left|f(x)\right|$ *is also a step function)*

Proof. 1. Let $f, g \in \tau[a,b]$. Let $(\xi_i)_{i=0}^m$ be a partition such that $f|_{(\xi_{i-1},\xi_i)} = c_i$ and $g|_{(\xi_{i-1},\xi_i)} = d_i$. Then

$$\int_a^b (\alpha f+\beta g)\,dx = \sum_{i=1}^m (\alpha c_i+\beta d_i)\delta\xi_i = \alpha \sum_{i=1}^m c_i\delta\xi_i+\beta\sum_{i=1}^m d_i\delta\xi_i = \alpha\int_a^b f\,dx+\beta\int_a^b g\,dx$$

Furthermore,

$$(\alpha f + \beta g)|_{(\xi_{i-1},\xi_i)} = \alpha c_i + \beta d_i = \text{ const.}$$

Thus,

$$\alpha f + \beta g \in \tau[a,b]$$

2. Let $h \in \tau[a,b]$ and $h(x) \geq 0 \forall x \in [a,b]$, then it holds that $v_i = h|_{(\xi_{i-1},\xi_i)} \geq 0$ and

$$\int_a^b h\,dx = \sum_{i=1}^m \underbrace{h_i}_{\geq 0}\underbrace{\triangle\xi_i}_{\geq 0} \geq 0$$

Now let $f, g \in \tau[a,b]$; $f \leq g$. Then $h = g - f \in \tau[a,b]$ (membership because of (1.)) and $h \geq 0$. Therefore,

$$0 \leq \int_a^b h\,dx = \int_a^b (g-f)\,dx = \int_a^b g\,dx - \int_a^b f\,dx$$

3. $f \leq \left|f\right|$, hence $\int_a^b f\,dx \leq \int_a^b \left|f\right|\,dx$ and also $-f \leq \left|f\right|$, so

$$\int_a^b (-f)\,dx = -\int_a^b f\,dx \leq \int_a^b \left|f\right|\,dx$$

$$\implies \left|\int_a^b f\,dx\right| \leq \int_a^b \left|f\right|\,dx$$

It is left to prove: $\left|f\right| \in \tau[a,b]$ (i.e. $\left|f\right|$ is a step function)

Let $f|_{(\xi_{i-1},\xi_i)} = c_i \implies \left|f\right||_{(\xi_{i-1},\xi_i)} = |c_i| = $ constant. Hence $\left|f\right| \in \tau[a,b]$.

□

3.2 Characteristic functions

Definition 3.3. *Let $a \subseteq \mathbb{R}^k$. We call $\chi_A : \mathbb{R}^n \to \mathbb{R}$ with*

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

a characteristic function *(indicator function) of set A. Often denoted as $\chi_A = \mathbb{1}$. Compare with Figure 14.*

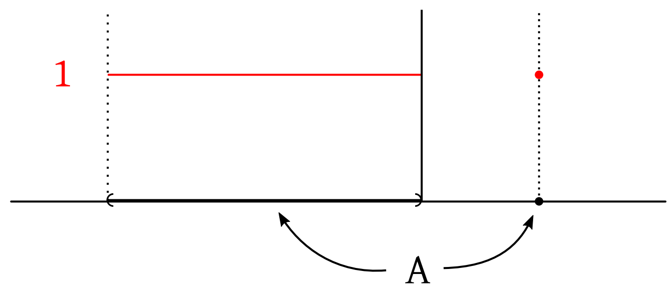

Figure 14: A characteristic function takes value 1 inside a set A which can be an interval (left) or a single point (right)

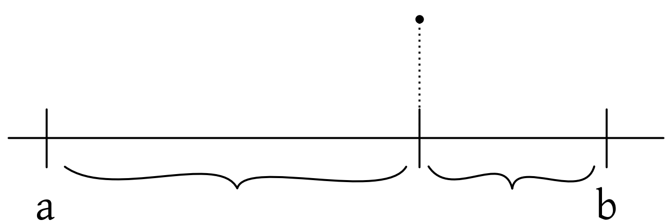

Figure 15: Linear combinations of characteristic functions are step functions

Remark 3.1. *Let $A = (a', b')$ with $a \leq a' < b' \leq b$. Then $\chi_{(a',b')} \in \tau[a, b]$. Also for $x \in [a, b]$, it holds that $\chi_{\{x\}} = \tau[a, b]$. Therefore*

- *every linear combination of characteristic functions of open subintervals (a', b') of $[a, b]$ or*
- *characteristic functions of one-point sets $\chi_{\{x\}}, x \in [a, b]$*

are step functions on $[a, b]$. Compare with Figure 15.

$$\sum_{j=1}^{n} \alpha_j \chi_{(a_j,b_j)} + \sum_{k=1}^{m} \beta_k \chi_{\{x_k\}} \in \tau[a, b]$$

On the opposite, $f \in \tau[a, b]$, hence

$$f|_{(x_{j-1},x_j)} \underbrace{=}_{j=1,\ldots,n} c_j \text{ and } f(x_j) \underbrace{=}_{j=0,\ldots,n} d_j$$

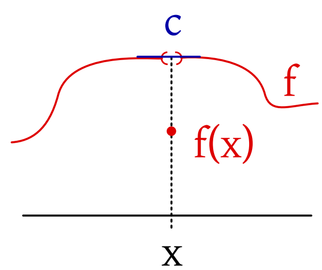

Figure 16: The limit value must not be necessarily the function value

$$f = \sum_{j=1}^{n} c_j \chi_{(x_{j-1},x_j)} + \sum_{j=0}^{n} d_j \chi_{\{x_j\}} = (*)$$

for $x \in (x_{j-1}, x_j)$ *it holds that* $\xi_{(x_{j-1},x_j)}(x) = 1$.

$$\chi_{(x_{l-1},x_l)}(x) = 0 \text{ for } l \neq j$$

$$\chi_{\{x_l\}}(x) = 0 \text{ for } l = 0, \dots, n$$

i.e. $\sum_{j=1}^{n} c_l \chi_{(x_{l-1},x_l)}(x) + \sum_{l=0}^{n} d_j \chi_{\{x_l\}}(x) = c_j \cdot 1 + 0 = c_j$ *hence* $(*) = c_j$ *on* (x_{j-1}, x_j). *Therefore* $f \in \tau[a,b] \iff f$ *is linear combination of characteristic functions of open intervals or one-pointed sets.*

3.3 Limit points

Definition 3.4. *Let* X *be a metric space* $A \subseteq X$ *and* $x \in X$ *is an accumulating point*[1] *of* A. *Let* $f : A \to \mathbb{R}$. *We say,* f *has limit* $c \in \mathbb{R}$ *in* x *(*$\lim_{\xi \to x} f(\xi) = c$*) if*

$$\forall \varepsilon > 0 \exists \delta > 0 : \forall \xi \in A, \xi \neq x \text{ and } d(\xi, x) < \delta : \left| f(\xi) - c \right| < \varepsilon$$

Remark 3.2. $x \in A$ *and* $c = f(x) \implies f$ *is continuous in* x.

We usually consider $A = [a,b] \subseteq \mathbb{R}$, $x \in [a,b]$.

It is possible, that f *in* x *has a limit,* $x \in A$ *and* $c = \lim_{\xi \to x} f(\xi) \neq f(x)$. *Compare with Figure 16.*

Definition 3.5. *Now let* $A \subseteq \mathbb{R}$ *and* x *is a accumulation point of* A. *Let* $f : A \to \mathbb{R}$ *be given. We say* f *has a right-sided limit* c *in* x *with* $c = \lim_{\xi \to x^+} f(\xi) = c$ *if*

$$\forall \varepsilon > 0 \exists \delta > 0 : \forall \xi \in A, \xi > x : |\xi - x| = \xi - x < \delta \implies \left| f(\xi) - c \right| < \varepsilon$$

[1] An accumulation point has 3 equivalent definitions (sequence, intersection, infinitely many elements in sphere).

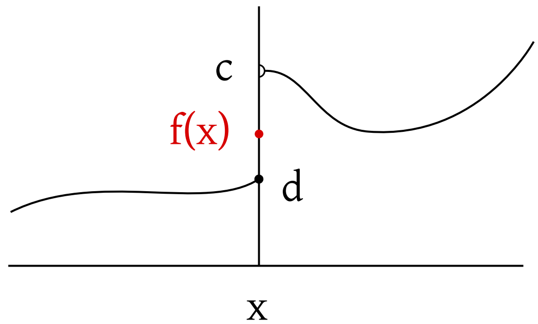

Figure 17: Left-sided and right-sided limit

Compare with Figure 17. The left-sided limit follows analogously (with $\xi < x$).

$$c = \lim_{\xi \to x^+} f(\xi) \qquad d = \lim_{\xi \to x^-} f(\xi)$$

Lemma 3.5 (Sequence criterion for limits of functions)**.** *Let $f : A \subseteq X \to \mathbb{R}$ be given. x is an accumulation point of A. Then it holds that*

$$\lim_{\xi \to x} f(\xi) = c \iff \forall (\xi_n)_{n \in \mathbb{N}} : \xi_n \in A, \xi_n \neq x \text{ and } \lim_{n \to \infty} \xi_n = x \text{ it holds that } \lim_{n \to \infty} f(\xi_n) = c$$

For one-sided limits $A \subseteq \mathbb{R}$ it holds that

$$c = \lim_{\xi \to x^+} f(\xi) \iff \forall (\xi_n)_{n \in \mathbb{N}} : \xi \in A \qquad \xi_n > x \text{ with } \lim_{n \to \infty} \xi_n = x \text{ it holds that } \lim_{n \to \infty} f(\xi_n) = c$$

Proof. See Analysis 1 lecture notes. □

Remark 3.3. *Attention! We, therefore, use two different definitions of limits.*

Lemma 3.6 (Cauchy criterion of limits of functions)**.** *Let $f : A \subseteq X \to \mathbb{R}$. Let x be an accumulation point of A. Let X be a metric space. Then it holds that f has a limit in x if and only if*

$$\forall \varepsilon > 0 \exists \delta > 0 : \forall \xi, \eta \in A : \xi \neq x_i : \eta \neq x$$

with $d(\xi, x) < \delta$ and $d(\eta, x) < \delta$ it holds that $\left| f(\xi) - f(\eta) \right| < \varepsilon$. Analogously for one-sided limits with $A \subseteq \mathbb{R}$. Additionally, we need the constraint that $\xi > x$ and $\eta > x$ for $\lim_{\xi \to x^+} f(\xi)$. And accordingly, $\xi < x$ and $\eta < x$ for $\lim_{\xi \to x^-} f(\xi)$.

Proof. $\Leftarrow$ Let $c = \lim_{\xi \to x} f(\xi)$ and let $\varepsilon > 0$ be chosen arbitrarily. Then there exists $\delta > 0$ such that $d(\xi, x) < \delta$ and $\xi \neq x$

$$\implies \left| f(\xi) - c \right| < \frac{\varepsilon}{2}$$

For ξ, η: $d(\xi, x) < \delta$ and $d(\eta, x) < \delta$ with $\xi, \eta \neq x$ is therefore

$$\left|f(\xi) - f(\eta)\right| = \left|f(\xi) - c + c - f(\eta)\right| \leq \left|f(\xi) - c\right| + \left|f(\eta) - c\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{=}\varepsilon$$

$\Rightarrow$ Assume the Cauchy criterion holds. We show that

1. for every sequence $(\xi_n)_{n\in\mathbb{N}}$, $\xi_n \in A \setminus \{x\}$ with $\lim_{n\to\infty} \xi_n = x$ it holds that $(f(\xi_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$ and therefore convergent in $\mathbb{R}$.
2. all Cauchy sequences have the *same* limit c.

We prove (1.)

Let $(\xi_n)_{n\in\mathbb{N}}$ be as above. Let $\varepsilon > 0$ be arbitrary. and N_ε large enough such that $\forall n \in N_\varepsilon$ it holds that $d(\xi_n, x) < \delta$ (δ chosen appropriately to ε according to the Cauchy criterion).

By the Cauchy criterion, $\left|f(\xi_n) - f(\xi_m)\right| < \varepsilon$ for all $m, n \geq N_\varepsilon$. Therefore $(f(\xi_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. If $\mathbb{R}$ is complete, then there exists $c = \lim_{n\to\infty} f(\xi_n)$. QED.

We prove (2.)

Let $\xi_n \to x$ as above and $\xi'_n \to x$ as above and $c = \lim_{n\to\infty} f(\xi_n)$ as well as $c' = \lim_{n\to\infty} f(\xi'_n)$. Let $\varepsilon > 0$ be arbitrary, N_ε such that $n \geq N_\varepsilon \implies \left|f(\xi_n) - c\right| < \frac{\varepsilon}{3}$ and $N'_\varepsilon \in \mathbb{N}$ such that $n \geq N'_\varepsilon \implies \left|f(\xi'_n) - c'\right| < \frac{\varepsilon}{3}$.

Furthermore choose $\delta > 0$ such that

$$d(\xi, x) < \delta \land d(\eta, x) < \delta \implies \left|f(\xi) - f(\eta)\right| < \frac{\varepsilon}{3}$$

(because of the Cauchy criterion). M_ε such that

$$n \geq M_\varepsilon \implies d(\xi_n, x) < \delta \land M'_\varepsilon : n \geq M'_\varepsilon \implies d(\xi'_n, x) < \delta$$

Let $n \geq \max\{N_\varepsilon, N'_\varepsilon, M_\varepsilon, M'_\varepsilon\}$.

This lecture took place on 2018/04/10.

Then it holds that

$$|c - c'| \leq \underbrace{\left|c - f(\xi_n)\right|}_{<\frac{\varepsilon}{3}} + \underbrace{\left|f(\xi_n) - f(\xi'_n)\right|}_{<\frac{\varepsilon}{3}} + \underbrace{\left|f(\xi'_n) - c'\right|}_{<\frac{\varepsilon}{3}} \qquad \forall\varepsilon > 0$$

Hence, $c = c'$. We have shown that $\exists c \in \mathbb{R} : \forall(\xi_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} \xi_n = x$ it holds that $\lim_{n\to\infty} f(\xi_n) = c$. So $\lim_{\xi\to\infty} f(\xi) = c$ because of Lemma 3.5. QED.

□

3.4 Regulated functions

Definition 3.6 (Regulated function)**.** *Let $a < b$, $f : [a,b] \to \mathbb{R}$. We call f a* regulated function on $[a,b]$ *if*

1. *$\forall x \in (a,b)$, f in x has a right-sided and a left-sided limit.*
2. *in $x = a$, f has a right-sided limit.*
3. *in $x = b$, f has a left-sided limit.*

$$\mathcal{R}[a,b] = \left\{ f : [a,b] \to \mathbb{R} \,\middle|\, f \text{ is a regulated function} \right\}$$

Definition 3.7 (Equivalent definition)**.**

1. *$\forall x \in [a,b)$, f has a right-sided limit in x*
2. *$\forall x \in (a,b]$, f has a left-sided limit in x*

Example 3.1. *Let f be continuous in $[a,b]$. Let $\varphi \in \tau[a,b]$ be a regulated function. Then $\varphi \in \mathcal{R}[a,b]$.*

Rationale:

Let $x_0 = a < x_1 < \cdots < x_n = b$ and $\varphi|_{(x_{j-1},x_j)} = c_j$.

Let $x \in [a,b]$ be chosen arbitrarily.

Case 1 *Let $x \in (x_{j-1}, x_j)$ for some $j \in \{1,\ldots,n\}$*

$$\implies \lim_{\xi \to x} \varphi(\xi) = c_j$$

Choose δ small enough such that $(x-\delta, x+\delta) \subseteq (x_{j-1}, x_j)$. $\forall \xi$ with $\xi \in (x-\delta, x+\delta)$ it holds that

$$\left|\varphi(\xi) - c_j\right| = 0$$

Case 2 *Let $x = x_j$ for $j = 1,\ldots,n-1$.*

$$\implies \lim_{\xi \to x_j^+} \varphi(\xi) = c_{j+1}$$

$$\lim_{\xi \to x_j^-} \varphi(\xi) = c_j$$

Compare with Figure 18.

Case 3 *Let $x = x_0 = a \implies \lim_{\xi \to a^+} \varphi(\xi) = c_1$.*

$$x = x_n = b \implies \lim_{\xi \to b^-} \varphi(\xi) = c_n$$

Let $f : [a,b] \to \mathbb{R}$ be monotonically increasing oder monotonically decreasing. Then $f \in \mathcal{R}[a,b]$. The proof will be done in the practicals.

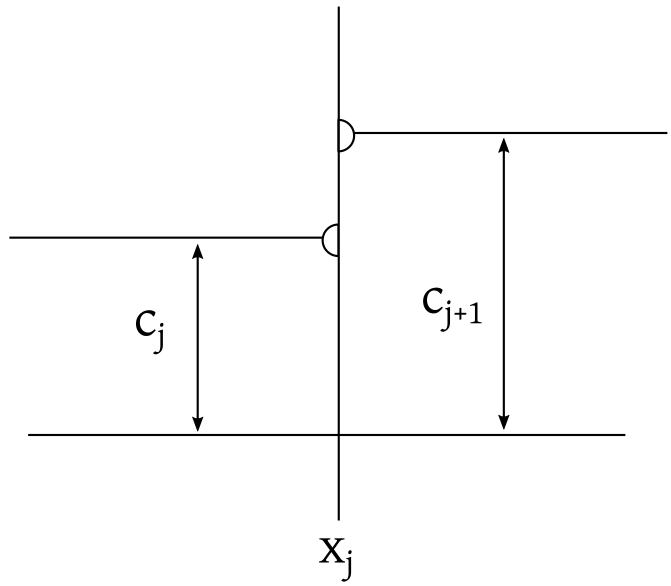

Figure 18: Regulated function

3.5 Bounded functions on bounded sets

Definition 3.8 (Boundedness)**.** *Let $X \neq \emptyset$ be a set. $f : X \to \mathbb{K}$ with $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. We say: f is bounded on X, if $f(X) \subseteq \mathbb{K}$ is a bounded set in $\mathbb{K}$. Hence, $\exists m \geq 0 : |f(x)| \leq m \forall x \in X$. We let,*

$$\mathcal{B}(X) = \left\{ f : X \to \mathbb{K} \,\middle|\, f \text{ is bounded} \right\}$$

$\mathcal{B}(X)$ has vector space structure. $f, g \in \mathcal{B}(X), \lambda \in \mathbb{K}$.

$$(f + g)(x) = f(x) + g(x)$$

$$(\lambda \cdot f)(x) = \lambda \cdot f(x)$$

$f + g \in \mathcal{B}(X)$ and $\lambda f \in \mathcal{B}(X)$. Let $|f(x)| \leq m \forall x \in X$ and $|g(x)| \leq m' \forall x \in X$. Then it holds that

$$|(f + g)(x)| = |f(x) + g(x)| \leq |f(x)| + |g(x)| \leq m + m'$$

Remark 3.4. *It is very interesting, that X does not require any kind of algebraic structure.*

We let

$$\|f\|_\infty = \sup \underbrace{\left\{ |f(x)| \,\middle|\, x \in X \right\}}_{\text{bounded in } \mathbb{R}} = \min \left\{ m \geq 0 \,\middle|\, |f(x)| \leq m \forall x \in X \right\}$$

Some work is required to show that $\|\cdot\|_\infty$ is a norm on $\mathcal{B}(X)$.

Hence, $(\mathcal{B}(X), \|\cdot\|_\infty)$ is a normed vector space. Convergence in $\mathcal{B}(X)$: It holds that $f_n \to f$ in $(\mathcal{B}(X), \|\cdot\|_\infty)$ if and only if $\forall \varepsilon > 0 \exists N \in \mathbb{N} : n \geq N \implies \left\|f_n - f\right\|_\infty < \varepsilon$.

$$\begin{aligned} \left\|f_n - f\right\|_\infty < \varepsilon &\iff \sup\left\{\left|f_n(x) - f(x)\right| : x \in X\right\} \\ &\iff \left|f_n(x) - f(x)\right| \leq \varepsilon \forall x \in X \end{aligned}$$

Hence, $f_n \to f$ in $(\mathcal{B}(X), \|\cdot\|_\infty) \iff \forall \varepsilon > 0 \exists N \in \mathbb{N} : n \geq N \implies \left|f_n(x) - f(x)\right| \leq \varepsilon \forall x \in X$. We say "$f_n$ converges *uniformly* to f on X".

Theorem 3.1 (Approximation theorem for regulated function)**.** *Let $f : [a,b] \to \mathbb{R}$. Then it holds that $f \in \mathcal{R}[a,b] \iff \forall \varepsilon > 0$ there exists some step function $\varphi \in \tau[a,b]$ such that $\left|\varphi(x) - f(x)\right| < \varepsilon \forall x \in [a,b]$ ($\left\|\varphi - f\right\|_\infty < \varepsilon$).*

Especially $\varepsilon_n = \frac{1}{n}$ and φ_n as above. Then it holds that $\left\|\varphi_n - f\right\|_\infty < \frac{1}{n}$, hence $f = \lim_{n\to\infty} \varphi_n$ uniformly on $[a,b]$.

Proof.

Direction $\implies$. Let $f \in \mathcal{R}[a,b]$.

Proof by contradiction. We negate our hypothesis:

$$\exists \varepsilon > 0 : \forall \varphi \in \tau[a,b] \exists x \in [a,b] : \left|\varphi(x) - f(x)\right| \geq \varepsilon \tag{1}$$

Assume (1) holds for $f \in [a,b]$. We construct nested intervals $[a_n, b_n]$ with $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ and $b_{n+1} - a_{n+1} = \frac{1}{2}(b_n - a_n)$ and (1) holds on $[a_n, b_n] \forall n \in \mathbb{N}$. Hence $\forall \varphi \in \tau[a_n, b_n] \exists x \in [a_n, b_n]$ such that $\left|\varphi(x) - f(x)\right| \geq \varepsilon$. This is what we want to show.

Let $a_0 = a$ and $b_0 = b$. Then (1) holds on $[a_0, b_0]$ by assumption. $n \to n+1$: Construction of $[a_{n+1}, b_{n+1}]$. Let $m_n = \frac{1}{2}(a_n + b_n)$.

Claim. (1) *holds either on $[a_n, m_n]$ or on $[m_n, b_n]$.*

Proof. Because if the opposite of (1) holds on $[a_n, m_n]$ as well as $[m_n, b_n]$, then there exists $\varphi_1^n \in \tau[a_n, m_n]$ with $\left|\varphi_n^1(x) - f(x)\right| < \varepsilon \forall x \in [a_n, m_n]$ and if the opposite of (1) holds on $[m_n, b_n]$:

$$\exists \varphi_n^2 \in \tau[m_n, b_n] : \left|\varphi_n^2(x) - f(x)\right| < \varepsilon \forall x \in [m_n, b_n]$$

$$\varphi^n(x) := \begin{cases} \varphi_n^1(x) & \text{if } x \in [a_n, m_n] \\ \varphi_n^2(x) & \text{if } x \in (m_n, b_n] \end{cases}$$

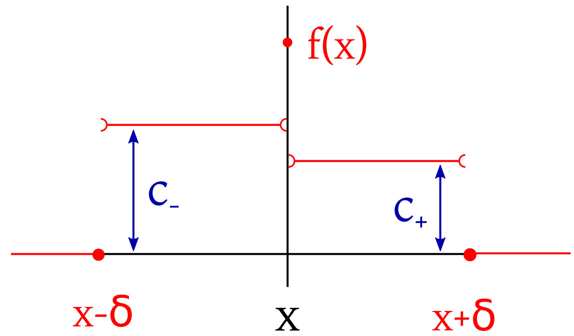

Figure 19: Construction of $\hat{\varphi}(\xi)$

Then φ^n is piecewise constant, hence $\varphi^n \in \tau[a_n, b_n]$ and it holds that

$$\left|\varphi^n(x) - f(x)\right| = \begin{cases} \underbrace{\left|\varphi_1^n(x) - f(x)\right|}_{<\varepsilon} & \text{for } x \in [a_n, m_n] \\ \underbrace{\left|\varphi_2^n(x) - f(x)\right|}_{<\varepsilon} & \text{for } x \in [m_n, b_n] \end{cases}$$

Therefore, $\left|\varphi^n(x) - f(x)\right| < \varepsilon$, which contradicts with (1) on $[a_n, b_n]$. We conclude: (1) holds on $[a_n, m_n]$ or on $[m_n, b_n]$. □

Now, choose $[a_{n+1}, b_{n+1}]$ as one of the subintervals in which (1) holds. Let $X \in \bigcap_{n\in\mathbb{N}}[a_n, b_n]$ (by completeness of $\mathbb{R}$).

Case $\mathbf{x} \in (\mathbf{a}, \mathbf{b})$ Let $x \in (a, b)$. Let ε as above such that (1) holds on every interval $[a_n, b_n]$. Let $c_+ = \lim_{\xi\to x^+} f(\xi)$ and $c_- = \lim_{\xi\to x^-} f(\xi)$ (feasible, because $f \in \mathcal{R}[a, b]$).

By the limit property, $\exists\delta > 0 : |\xi - x| < \delta$ and $\xi < x$, then $\left|f(\xi) - c_-\right| < \varepsilon$ and $|\xi - x| < \delta$ and $x < \delta$ then $\left|f(\xi) - c_+\right| < \varepsilon$.

Additionally, choose δ sufficiently small enough such that $(x - \delta, x + \delta) \subseteq [a, b]$.

$$\hat{\varphi}(\xi) := \begin{cases} 0 & \text{for } \xi \in [a, b] \setminus (x - \delta, x + \delta) \\ c_- & \text{for } \xi \in (x - \delta, x) \\ c_+ & \text{for } \xi \in (x, x + \delta) \\ f(x) & \text{for } \xi = x \end{cases}$$

Compare with Figure 19. $\hat{\varphi} \in \tau[a,b]$ and it holds that

$$\forall \xi \in (x-\delta, x+\delta) : \left|\hat{\varphi}(\xi) - f(\xi)\right| = \begin{cases} \underbrace{\left|c_- - f(\xi)\right|}_{<\varepsilon} & \text{for } \delta \in (x-\delta, x) \\ \underbrace{\left|f(x) - f(x)\right|}_{=0} & \text{for } \xi = x \\ \underbrace{\left|c_+ - f(\xi)\right|}_{<\varepsilon} & \text{for } \xi \in (x, x+\delta) \end{cases}$$

Hence, $\left|\hat{\varphi}(\xi) - f(\xi)\right| < \varepsilon$. Now let N be sufficiently large enough such that $[a_N, b_N] \subseteq (x-\delta, x+\delta)$ (possible because $([a_n, b_n])_{n\in\mathbb{N}}$ gives nested intervals tightening on x). Then for $[a_N, b_N]$ it holds that:

$$\hat{\varphi}|_{[a_N,b_N]} \in \tau[a_N, b_N]$$

and $\forall \xi \in [a_N, b_N] \subseteq (x-\delta, x+\delta)$ it holds that $\left|\hat{\varphi}(\xi) - f(\xi)\right| < \varepsilon$. This contradicts with (1) on $[a_N, b_N]$.

Case x = a and x = b Is analogous to one-sided limits.

Direction $\Longleftarrow$. Let $f = \lim_{n\to\infty} \varphi_n$ be uniform on $[a,b]$.

Claim. *$\forall x \in [a,b)$ there exists a right-sided limit of f in x.*

Proof. Let $\varepsilon > 0$ be arbitrary. $N \in \mathbb{N}$ sufficiently large such that $\left|f(\xi) - \varphi_N(\xi)\right| < \frac{\varepsilon}{2} \forall \xi \in [a,b]$. φ_N is piecewise constant. Choose $\delta > 0$ such that $\varphi_N|_{(x,x+\delta)} = c$. Now let $\xi, \eta \in (x, x+\delta)$ be chosen arbitrarily. Then it holds that

$$\left|f(\xi) - f(\eta)\right| \leq |f(\xi) - \underbrace{c}_{=\varphi_N(\xi)}| + |\underbrace{c}_{=\varphi_N(\eta)} - f(\eta)| = \underbrace{|f(\xi) - \varphi_N(\xi)|}_{<\frac{\varepsilon}{2}} + \underbrace{|\varphi_N(\eta) - f(\eta)|}_{<\frac{\varepsilon}{2}} < \varepsilon$$

Therefore f has a right-sided limit in x by the Cauchy criterion. f has left-sided limit in every point. □

$x \in (a,b]$ follows analogously.

□

Corollary. *Every regulated function $f \in \mathcal{R}[a,b]$ is bounded. Let $\varphi \in \tau[a,b]$ with $\left\|f - \varphi\right\|_\infty < 1$. φ is bounded, hence $\exists m \in [0,\infty)$: $\left|\varphi(x)\right| \leq m \forall x \in [a,b]$. Then it holds that $\left|f(x)\right| \leq \left|f(x) - \varphi(x)\right| + \left|\varphi(x)\right| < 1 + m \forall x \in [a,b]$, hence $f \in \mathcal{B}[a,b]$.*

$$\mathcal{R}[a,b] \subseteq \mathcal{B}[a,b]$$

Corollary. *Let $f \in \mathcal{R}[a,b] \iff f = \sum_{j=0}^{\infty} \psi_j$ with $\psi_j \in \tau[a,b]$ and the series converges uniformly on $[a,b]$.*

Proof.

Direction $\Longleftarrow$. Let $f = \sum_{j=0}^{\infty} \psi_j$ with uniform convergence. Let $\varphi_n = \sum_{j=0}^{\infty} \psi_j \in \tau[a,b]$ and $f = \lim_{n\to\infty} \phi_n$ uniform on $[a,b] \xRightarrow{\textit{Theorem } 3.1} f \in \mathcal{R}[a,b]$.

Direction $\Longrightarrow$. Let $f \in \mathcal{R}[a,b]$ and $f = \lim_{n\to\infty} \varphi_n$ with $\varphi_n \in \tau[a,b]$ (by Theorem 3.1).

$$\psi_0 = \varphi_0$$
$$\psi_j = \varphi_j - \varphi_{j-1} \qquad \text{for } j \geq 1$$
$$\sum_{j=0}^{n} \psi_j = \varphi_0 + \sum_{j=1}^{n} (\varphi_j - \varphi_{j-1}) = \varphi_0 + \sum_{j=1}^{n} \varphi_j - \sum_{j=0}^{n-1} \varphi_j = \varphi_n$$

converges uniformly to f.

□

4 Integration of regulated functions

Definition 4.1 (Definition with a theorem)**.** *Let $f \in \mathcal{R}[a,b]$ and $\varphi_n \in \tau[a,b]$ with $f = \lim_{n\to\infty} \varphi_n$ is uniform on $[a,b]$. We let*

$$\int_a^b f \, dx = \lim_{n\to\infty} \int_a^b \varphi_n \, dx$$

for the integral of f on $[a,b]$.

Theorem: This limit (on the right-hand side) always exists and is independent of the particular choice of the approximating sequence.

Proof. φ_n is chosen as above.

$$i_n = \int_a^b \varphi_n \, dx$$

Show: i_n is cauchy sequence in $\mathbb{R}$.

This lecture took place on 2018/04/12.

Let $\varepsilon > 0$ be chosen arbitrary. Choose $N \in \mathbb{N}$ such that

$$n \geq N \implies \left\| f - \varphi_n \right\|_\infty < \frac{\varepsilon}{2(b-a)}$$

For $n, m \geq N$ it holds for $x \in [a, b]$ that

$$\left|\varphi_n(x) - \varphi_m(x)\right| \leq \left|\varphi_n(x) - f(x)\right| + \left|f(x) - \varphi_m(x)\right|$$

$$\leq \left\|\varphi_n - f\right\|_\infty + \left\|f - \varphi_m\right\|_\infty < \frac{\varepsilon}{2(b-a)} + \frac{\varepsilon}{2(b-a)} = \frac{\varepsilon}{b-a}$$

$\left|\varphi_n - \varphi_m\right|$ is a step function.

$$\left|\varphi_n - \varphi_m\right| \leq \frac{\varepsilon}{b-a} \cdot \underbrace{\chi_{[a,b]}}_{\in \tau[a,b]}$$

Integral for subsequence is monotonous:

$$|i_n - i_m| = \left|\int_a^b \varphi_n \, dx - \int_a^b \varphi_m \, dx\right| = \left|\int_a^b (\varphi_n - \varphi_m) \, dx\right| \leq \int_a^b \left|\varphi_n - \varphi_m\right| \, dx$$

$$\underbrace{<}_{\text{by monotonicity}} \int_a^b \frac{\varepsilon}{b-a} \cdot \chi_{[a,b]} \, dx = \frac{\varepsilon}{b-a} \underbrace{\int_a^b \chi_{[a,b]}}_{1 \cdot (b-a)} \, dx = \varepsilon$$

So $(i_n)_{n \in \mathbb{N}}$ is a Cauchy sequence. $\mathbb{R}$ is complete, hence $i = \lim_{n \to \infty} i_n$ exists.

Uniqueness: (dt. mithilfe des Reissverschlussprinzips)

Let $(\varphi_n)_{n \in \mathbb{N}}$, $(\Phi_n)_{n \in \mathbb{N}}$ be two sequences of step functions, converging uniformly towards f.

$$i_n = \int_a^b \varphi_n \, dx \quad \text{and} \quad j_n = \int_a^b \Phi_n \, dx$$

$$i = \lim_{n \to \infty} i_n \qquad j = \lim_{n \to \infty} j_n$$

Show that $i = j$.

Now we construct a sequence $(\mu_n)_{n \in \mathbb{N}}$ of step functions.

$$\underbrace{(\varphi_1, \Phi_1, \varphi_2, \Phi_2, \dots)}_{(\mu_n)_{n \in \mathbb{N}}}$$

μ_n is a sequence of step functions converging uniformly towards f (the proof is left as an exercise to the reader).

Because of part 1 of the proof:

$$m_n = \int_a^b \mu_n \, dx \text{ converges with limit } m$$

$(i_n)_{n \in \mathbb{N}}$ as well as $(j_n)_{n \in \mathbb{N}}$ are subsequences of $(m_n)_{n \in \mathbb{N}}$. Hence it holds that $i = \lim_{n \to \infty} i_n = m = \lim_{n \to \infty} j_n = j$. $\square$

Theorem 4.1 (Elementary properties of an integral)**.** *Let $f, g \in \mathcal{R}[a, b]$, $\lambda, \mu \in \mathbb{R}$. Then it holds that*

Linearity

$$\lambda f + \mu g \in \mathcal{R}[a, b] \text{ and } \int_a^b (\lambda f + \mu g)\, dx = \lambda \int_a^b f\, dx + \mu \int_a^b g\, dx$$

Monotonicity *If $f(x) \leq g(x) \forall x \in [a, b]$ ($f \leq g$) it holds that*

$$\int_a^b f\, dx \leq \int_a^b g\, dx$$

Boundedness $\left|f\right| \in \mathcal{R}[a, b]$ *and*

$$\left|\int_a^b f\, dx\right| \leq \int_a^b \left|f\right|\, dx$$

Proof.

Linearity. Let $x \in [a, b)$ and $c_+ = \lim_{\xi \to x_+} f(\xi)$ as well as $d_+ = \lim_{\xi \to x_+} g(\xi)$ ($f, g \in \mathcal{R}[a, b]$). Then it holds that

$$\lim_{\xi \to x^+} (\lambda f(\xi) + \mu g(\xi)) = \lambda \lim_{\xi \to x^+} f(\xi) + \mu \lim_{\xi \to x^+} g(\xi) = \lambda c_+ + \mu d_+$$

exists. Analogously for the left side, hence $\lambda f + \mu g \in \mathcal{R}[a, b]$.

Claim. *Let $\varphi_n, \Phi_n \in \tau[a, b]$ with $\varphi_n \to f$ and $\Phi_n \to g$ is uniform on $[a, b]$. Hence $\lambda\varphi_n + \mu\Phi_n \to \lambda f + \mu g$ is continuous on $[a, b]$.*

Proof. Let $\varepsilon > 0$ be arbitrary, N such that $n \geq N \implies \left\|\varphi_n - f\right\|_\infty < \frac{\varepsilon}{2(|\lambda|+1)}$ and M such that $n \geq M \implies \left\|\Phi_n - g\right\|_\infty < \frac{\varepsilon}{2(\left|\mu\right|+1)}$.

Then it holds that

$$\begin{aligned} \left\|\lambda\varphi_n + \mu\Phi_n - \lambda f - \mu g\right\|_\infty &\leq |\lambda| \left\|\varphi_n - f\right\|_\infty + \left|\mu\right| \left\|\Phi_n - g\right\|_\infty \\ &< \frac{|\lambda|}{2(|\lambda| + 1)} \cdot \varepsilon + \frac{\left|\mu\right|}{2(\left|\mu\right| + 1)} \cdot \varepsilon < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

□

We continue:

$$\int_a^b (\lambda f + \mu g)\, dx = \lim_{n \to \infty} \int_a^b (\lambda\varphi_n + \mu\Phi_n)\, dx = \lim_{n \to \infty} \left(\lambda \int_a^b \varphi_n\, dx + \mu \int_a^b \Phi_n\, dx\right)$$

$$= \lambda \underbrace{\lim_{n \to \infty} \int_a^b \varphi_n\, dx}_{\text{exists}} + \mu \underbrace{\lim_{n \to \infty} \int_a^b \Phi_n\, dx}_{\text{exists}} = \lambda \int_a^b f\, dx + \mu \int_a^b g\, dx$$

Monotonicity. Show: Let $h \in \mathcal{R}[a,b]$ with $h \geq 0$ in $[a,b]$. Then it holds that $\int_a^b h\,dx \geq 0$.

Claim. *There exists $(\tilde{\varphi}_n)_{n\in\mathbb{N}}$ with $\tilde{\varphi}_n \to h$ uniform on $[a,b]$ and $\tilde{\varphi}_n \geq 0$.*

Proof. Let $(\varphi_n)_{n\in\mathbb{N}}$, $\varphi_n \in \tau[a,b]$ with $\varphi_n \to h$ uniform on $[a,b]$. We define $\tilde{\varphi}_n$ such that

$$\varphi_n = \sum_{j=1}^{m_n} c_j \chi_{(x_{j-1},x_j)} + \sum_{j=0}^{m_n} d_j \chi_{\{x_j\}}$$

$$\tilde{\varphi}_n := \sum_{j=1}^{m_n} \underbrace{\tilde{c}_j}_{\geq 0} \chi_{(x_{j-1},x_j)} + \sum_{j=0}^{m_n} \underbrace{h(x_j)}_{\geq 0} \chi_{\{x_j\}}$$

and $\tilde{c}_j := \max c_j, 0 \geq 0$. So it holds that $\tilde{\varphi}_n \geq 0$.

For $x = x_l$ ($l \in \{0,\ldots,m_n\}$) it holds that

$$\left|\tilde{\varphi}_n(x_l) - h(x_l)\right| = \left|\sum_{j=1}^{m_n} \tilde{c}_j \underbrace{\chi_{(x_{j-1},x_j)}(x_l)}_{=0 \text{ bc. } x_l \notin (x_{j-1},x_j)} + \sum_{j=0}^{m_n} h(x_j) \underbrace{\chi_{\{x_j\}}(x_l)}_{=\delta_{j,l}} - h(x_l)\right|$$

$$= |h(x_l) - h(x_l)| = 0 \leq \left|\varphi_n(x_l) - h(x_l)\right|$$

For $x \in (x_{j-1}, x_j)$ it holds that

$$\left|\tilde{\varphi}_n(x) - h(x)\right| = \left|\sum_{j=1}^{m_n} \tilde{c}_j \underbrace{\chi_{(x_{j-1},x_j)}(x)}_{\delta_{l,j}} + \sum_{j=0}^{m_n} h(x) \cdot \underbrace{\chi_{\{x_j\}}(x)}_{=0 \text{ bc. } x \neq x_j} - h(x)\right|$$

$$= |\tilde{c}_l - h(x)| = \begin{cases} |c_l - h(x)| & \text{if } c_l \geq 0 \\ |h(x)| = h(x) & \text{if } c_l < 0 \end{cases}$$

$$\leq \begin{cases} |c_l - h(x)| & \text{if } c_l \geq 0 \\ h(x) - c_l & \text{if } c_l < 0 \end{cases}$$

$$= \begin{cases} \left|\varphi_n(x) - h(x)\right| & \text{if } c_l = \varphi_n(x) \geq 0 \\ \left|h(x) - \varphi_n(x)\right| & \text{if } c_l = \varphi_n(x) < 0 \end{cases}$$

$$= \left|\varphi_n(x) - h(x)\right|$$

hence, $\left|\tilde{\varphi}_n(x) - h(x)\right| \leq \left|\varphi_n(x) - h(x)\right|$ for $x \in (x_{l-1}, x_l)$ as well as $x = x_i$, hence

$$\left\|\tilde{\varphi}_n - h\right\|_\infty \leq \underbrace{\left\|\varphi_n - h\right\|_\infty}_{\to 0 \text{ for } n \to \infty}$$

Hence $\left\|\tilde{\varphi}_n - h\right\|_\infty \to 0$ for $n \to \infty$, hence $\tilde{\varphi}_n$ converges uniformly to h. There exists

$$\int_a^b h\,dx = \underbrace{\lim_{n\to\infty}\int_a^b \underbrace{\tilde{\varphi}_n}_{\geq 0}\,dx}_{\geq 0} \geq 0$$

□

To show monotonicity, we let $f \leq g$ in $[a,b]$, hence $h = g - f \geq 0$ in $[a,b]$

$$\implies 0 \leq \int_a^b h\,dx = \int_a^b g\,dx - \int_a^b f\,dx$$

$$\implies \int_a^b f\,dx \leq \int_a^b g\,dx$$

Boundedness. Consider $\left|f\right|$. Proving $\left|f\right| \in \mathbb{R}[a,b]$ is left as an exercise to the reader.

$$f \leq \left|f\right| \text{ on } [a,b] \xRightarrow{\text{monotonicity}} \int_a^b f\,dx \leq \int_a^b \left|f\right|\,dx$$

$$-f \leq \left|f\right| \text{ on } [a,b] \xRightarrow{\text{monotonicity}} \int_a^b (-f)\,dx = -\int_a^b f\,dx \leq \int_a^b \left|f\right|\,dx$$

$$\implies \left|\int_a^b f\,dx\right| \leq \int_a^b \left|f\right|\,dx$$

□

Remark 4.1. *$\mathcal{R}[a,b]$ is a vector space.*

1. *$f,g \in \mathbb{R}[a,b] \implies \lambda f + \mu g \in \mathcal{R}[a,b]$. $\|\cdot\|_\infty$ is a norm on $\mathcal{R}[a,b]$. ($\mathcal{R}[a,b], \|\cdot\|_\infty$) is a normed vector space. Subspace of ($\mathcal{B}[a,b], \|\cdot\|_\infty$). We will show in the practicals that ($\mathcal{R}[a,b], \|\cdot\|_\infty$) is complete.*

Theorem 4.2 (Mean value theorem of integration calculus)**.** *Let f be continuous on $[a,b]$ and $p \in \mathcal{R}[a,b]$ and $p \geq 0$ in $[a,b]$. Then $f \cdot p \in \mathcal{R}[a,b]$ and there exists $\xi \in [a,b]$ such that*

$$\int_a^b f \cdot p\,dx = f(\xi) \cdot \int_a^b p\,dx$$

Proof. Let $m = \min\{f(z) : z \in [a,b]\}$ (exists because f is continuous and $[a,b]$ is compact).

$$M = \max\{f(z) : z \in [a,b]\}$$

$$f([a,b]) = [m,M] \text{ (by the mean value theorem)}$$

It holds that

$$m \cdot \underbrace{p(x)}_{\geq 0} \leq f(x) \cdot p(x) \leq M \cdot p(x)$$

By monotonicity,

$$m \int_a^b p(x)\,dx \leq \int_a^b fp\,dx \leq M \int_a^b p\,dx$$

Therefore, there exists $\eta \in [m, M]$.

$$\eta \cdot \int_a^b p(x)\,dx = \int_a^b fp\,dx$$

Mean value theorem: For $\eta \in [m, M]$ there exists $\xi \in [a, b]$ such that

$$\eta = f(\xi) \text{ (f is continuous!)}$$

Hence,

$$f(\xi) \int_a^b p\,dx = \int_a^b f \cdot p\,dx$$

$f \cdot p$ is regulated function (over one-sided limits). □

Lemma 4.1. *Let $f \in \mathcal{R}[a, b]$ and $a \leq \alpha < \beta < \gamma \leq b$. Then*

$$f|_{[\alpha,\beta]} \in \mathcal{R}[\alpha, \beta], f|_{\beta,\gamma} \in \mathcal{R}[\beta, \gamma]$$

$$f|_{[\alpha,\gamma]} \in \mathcal{R}[\alpha, \gamma] \text{ (immediate over onesided limit)}$$

and it holds that

$$\int_\alpha^\gamma f\,dx = \int_\alpha^\beta f\,dx + \int_\beta^\gamma f\,dx$$

Compare with Figure 20.

Proof. Show that this statement holds for $\varphi \in \tau[a, b]$. Without loss of generality, $\alpha = a, \gamma = b$.

$$\gamma = \sum_{j=1}^m c_j \chi_{(x_{j-1}, x_j)} + \sum_{j=0}^m 0 \cdot \chi_{x_j}$$

The zero represents that this term is not relevant for the integral in any way.

Case 1 $\beta = x_l$ for some $l \in \{1, \ldots, m-1\}$

$$\int_\alpha^\gamma \varphi\,dx = \sum_{j=1}^m c_j (x_j - x_{j-1})$$

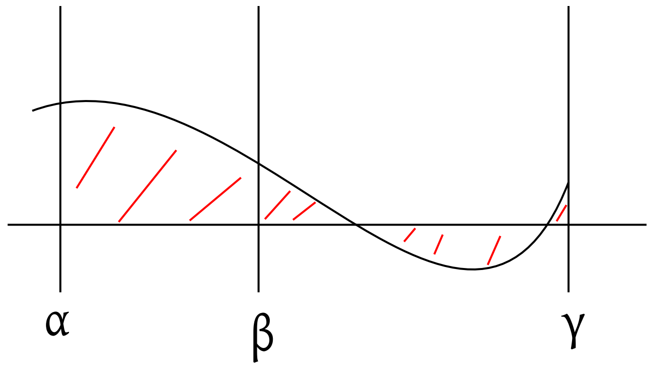

Figure 20: Positive and negative area covered by the integral

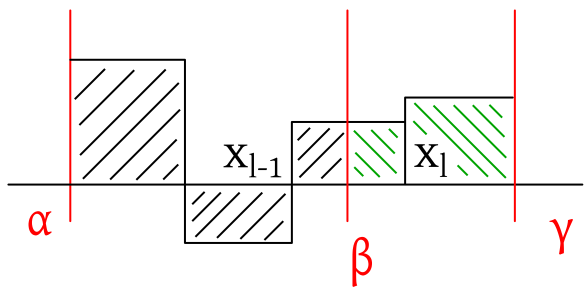

Figure 21: Setting for $\beta \in (x_{l-1}, x_l)$

$$\int_\alpha^\beta \varphi\, dx = \int_\alpha^{x_l} \varphi\, dx = \sum_{j=1}^{l} c_j(x_j - x_{j-1})$$

$$\int_\beta^\gamma \varphi\, dx = \int_{x_l}^{\gamma} \varphi\, dx = \sum_{j=l+1}^{m} c_j(x_j - x_{j-1})$$

And now,

$$\sum_{j=1}^{l} c_j(x_j - x_{j-1}) + \sum_{j=l+1}^{m} c_j(x_j - x_{j-1}) = \sum_{j=1}^{m} c_j(x_j - x_{j-1})$$

Case 2 $\beta \in (x_{l-1}, x_l)$ for some $l \in \{1, \ldots, m\}$. Compare with Figure 21.

$$\int_\alpha^\beta \varphi\, dx = \sum_{j=1}^{l-1} c_j(x_j - x_{j-1}) + c_l \cdot (\beta - x_{l-1})$$

$$\int_\beta^\gamma \varphi\, dx = c_l(x_l - \beta) + \sum_{j=l+1}^m c_j(x_j - x_{j-1})$$

Now we consider the addition of these two expressions:

$$\int_\alpha^\beta \varphi\, dx + \int_\beta^\gamma \varphi\, dx$$

$$= \sum_{j=1}^{l-1} c_j(x_j - x_{j-1}) + \underbrace{c_l(\beta - x_{l-1}) + c_l(x_l - \beta)}_{=c_l(x_l - x_{l-1})} + \sum_{j=l+1}^m c_j(x_j - x_{j-1})$$

$$= \sum_{j=1}^m c_j(x_j - x_{j-1}) = \int_\alpha^\gamma \varphi\, dx$$

Let $\varphi_n \in \tau[\alpha, \beta]$ with $\varphi_n \to f$ uniform on $[\alpha, \beta] \implies \varphi_n|_{[\alpha,\beta]} \to f|_{[\alpha,\beta]}$ uniform on $[\alpha, \beta]$ and also $\varphi_n|_{[\beta,\gamma]} \to f|_{[\beta,\gamma]}$ uniform on $[\beta, \gamma]$.

$$\int_\alpha^\gamma f\, dx = \lim_{n\to\infty} \int_\alpha^\gamma \varphi_n\, dx = \lim_{n\to\infty} \left(\int_\alpha^\beta \varphi_n\, dx + \int_\beta^\gamma \varphi_n\, dx \right)$$

$$= \underbrace{\lim_{n\to\infty} \int_\alpha^\beta \varphi_n\, dx}_{\substack{\text{exists because} \\ \varphi_n|_{[\alpha,\beta]} \to f|_{[\alpha,\beta]} \text{uniform}}} + \lim_{n\to\infty} \int_\beta^\gamma \varphi_n\, dx = \int_\alpha^\beta f\, dx + \int_\beta^\gamma f\, dx$$

□

Remark 4.2 (Notation). *Let $\alpha < \beta$, $\alpha, \beta \in [a, b]$ and $f \in \mathcal{R}[a, b]$. We let*

$$\int_\beta^\alpha f\, dx := -\int_\alpha^\beta f\, dx$$

By this convention, it holds that

$$\int_\alpha^\alpha f\, dx = -\int_\alpha^\alpha f\, dx \implies \int_\alpha^\alpha f\, dx = 0$$

Lemma 4.2. *Let $f \in \mathcal{R}[a, b]$ and $\alpha, \beta, \gamma \in [a, b]$ (without particular order). Then it holds that*

$$\int_\alpha^\gamma f\, dx = \int_\alpha^\beta f\, dx + \int_\beta^\gamma f\, dx$$

Proof. Special case: 2 points are equal

$$\alpha = \gamma \implies \int_a^\alpha f\, dx = 0$$

$$\int_\alpha^\beta f\,dx + \int_\beta^\alpha f\,dx = \int_\alpha^\beta f\,dx - \int_\alpha^\beta f\,dx = 0$$

$$\beta = \gamma \qquad \beta = \alpha$$

Case: $\alpha < \beta < \gamma$ follows immediately

And just as a representative other case: $\alpha < \gamma < \beta$

$$\int_\alpha^\beta f\,dx \underbrace{=}_{\text{by Lemma 1.7}} \int_\alpha^\gamma f\,dx + \underbrace{\int_\gamma^\beta f\,dx}_{-\int_\beta^\gamma f\,dx}$$

$$\int_\alpha^\beta f\,dx + \int_\beta^\gamma f\,dx = \int_\alpha^\gamma f\,dx$$

□

This lecture took place on 2018/04/17.

Lemma 4.3. *Let $f \in \mathcal{R}[a,b]$. Then there exists an at most countable set $A \subseteq [a,b]$ such that f is continuous in every point $x \in [a,b] \setminus A$.*

Proof. Let $f \in \mathcal{R}[a,b]$ and $(\varphi_n)_{n\in\mathbb{N}}$ with $\varphi_n \in \tau[a,b]$ and $\varphi \to f$ converging uniformly on $[a,b]$.

$$\varphi_n = \sum_{j=1}^{m_n} c_j^n \chi_{(X_{j-1}^n, X_j^n)} + \sum_{j=0}^{m_n} d_j^n \chi_{\left\{x_j^n\right\}}$$

$$x_0^n = a < x_1^n < \ldots < x_{m_n}^n = b$$

are separating points for φ_n

$$A = \left\{X_j^n : n \in \mathbb{N}, j \in \{0, \ldots, m_n\}\right\}$$

A is a countable union of finite sets $A_n = \left\{x_0^n, x_{m_n}^n\right\}$. A is countable (as unions of finite sets are).

Now we show: f is continuous in every point $x \in [a,b] : x \notin A$. Let $\varepsilon > 0$ be arbitrary. Choose $N \in \mathbb{N}$ sufficiently large such that $\left\|\varphi_N - f\right\|_\infty < \frac{\varepsilon}{2}$. Because $x \in A$, there exists $j \in \{1, \ldots, m_N\}$ such that $x \in (x_{j-1}^N, X_j^N)$ is open. Choose $\delta > 0$ such that $(x-\delta, x+\delta) \subset (x_{j-1}^N, x_j^n)$, hence $\forall \xi \in (x-\delta, x+\delta)$ it holds that $\varphi_N(\xi) = c_j^N$. Now consider $\xi \in (x-\delta, x+\delta)$, hence $|\xi - x| < \delta$. Then it holds that

$$\left|f(\xi) - f(x)\right| = \left|f(\xi) - \underbrace{\varphi_N(x)}_{c_j^N = \varphi_N(\xi)} + \varphi_N(x) - f(x)\right|$$

$$\leq \underbrace{\left|f(\xi) - \varphi_N(\xi)\right|}_{\leq \|f - \varphi_N\|_\infty} + \underbrace{\left|\varphi_N(x) - f(x)\right|}_{\leq \|\varphi_N - f\|_\infty} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Hence f is continuous in x. □

Remark 4.3 (Notation)**.** *Let $f \in \mathcal{R}[a,b]$. For $x \in [a,b)$, there exists $f_+(x) := \lim_{\xi \to x_+} f(\xi)$. For $x \in (a,b]$, there exists $f_-(x) := \lim_{\xi \to x_-} f(\xi)$. Because of Lemma 4.3, it holds that $f_+(x) = f_-(x) = f(x)$ for all $x \in [a,b] \setminus A$ and A is at most countable.*

Definition 4.2 (One-sided derivatives)**.** *Let $g : [a,b] \to \mathbb{R}$ and $x \in [a,b)$. We say g has the* right-sided derivative *$g'_+(x)$ if*

$$\lim_{\xi \to x_+} \frac{g(\xi) - g(x)}{\xi - x} =: g'_+(x)$$

exists. Analogously we define the left-sided derivative

$$g'_-(x) = \lim_{\xi \to x_-} \frac{g(\xi) - g(x)}{\xi - x}$$

for $x \in (a,b]$. Compare with Figure 22.

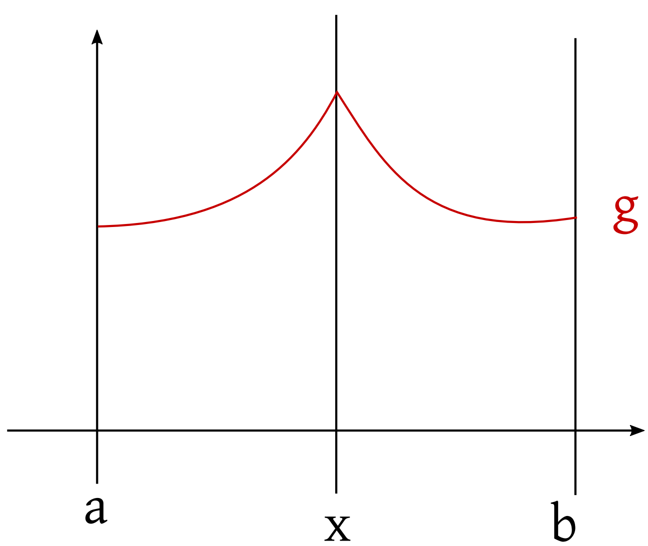

Figure 22: In this example, the left- and right-sided derivatives are not equal. $f'_+(x) \neq f'_-(x)$

Remark 4.4. *If g in x has a one-sided derivative, then it holds that*

$$\lim_{\xi \to x_\pm} (g(\xi) - g(x)) = 0$$

Hence g is continuous in x.

Remark 4.5. *$g : [a,b] \to \mathbb{R}$ is differentiable in point $x \in (a,b)$ with derivative $g'(x)$ $\iff$ g has a left- and right-sided derivative in x and it holds that $g'_-(x) = g'_+(x)$ $(= g'(x))$.*

Theorem 4.3 (Fundamental theorem of differential/integration calculus, variation 1)**.** *Isaac Barrow (1630–1677), Isaac Newton (1642–1726), Gottfried Wilhelm von Leibniz (1646–1716).*

Let $f \in \mathcal{R}[a,b]$, $\alpha \in [a,b]$ and we define

$$F(x) = \int_\alpha^x f \, d\xi$$

Then F is right-sided differentiable in every point $x \in [a,b)$ and in every $x \in (a,b]$ left-sided differentiable. Furthermore it holds that

$$F'_+(x) = f_+(x) \forall x \in [a,b) \tag{2}$$
$$F'_-(x) = f_-(x) \forall x \in (a,b] \tag{3}$$

Remark 4.6.

$$\frac{d}{dx}\left(\int_\alpha^x f \, d\xi\right) = f(x)$$

for all x such that f is continuous in x. For those x, $F'(x)$ is differentiable in x with $F'(x) = f(x)$.

Definition 4.3. *Let $f \in \mathcal{R}[a,b]$ and $\varphi : [a,b] \to \mathbb{R}$ such that φ is one-sided differentiable on $[a,b]$. If $\Phi'_+(x) = f_+(x) \forall x \in [a,b)$ and $\Phi'_-(x) = f_-(x) \forall x \in (a,b]$ then we call Φ an antiderivative of regulated function f.*

Proof of the Theorem 4.3. Let $x_1, x_2 \in [a,b]$ be arbitrary. Let F be defined as above. Then it holds that

$$\begin{aligned} |F(x_2) - F(x_1)| &= \left|\int_\alpha^{x_2} f \, d\xi - \int_\alpha^{x_1} f \, d\xi\right| \\ &= \left|\int_\alpha^{x_2} f \, d\xi + \int_{x_1}^{\alpha} f \, d\xi\right| = \left|\int_{x_1}^{x_2} f \, d\xi\right| \\ &\leq \int_{x_1}^{x_2} \left|f\right| \, d\xi \leq \int_{x_1}^{x_2} \underbrace{\left\|f\right\|_\infty}_{\text{const independent of } \xi} d\xi = \left\|f\right\|_\infty \cdot |x_2 - x_1| \end{aligned}$$

Hence F is Lipschitz continuous with Lipschitz constant $\left\|f\right\|_\infty$. So F is continuous in $[a, b]$.

One-sided derivatives: Let $x \in [a, b)$ and $\varepsilon > 0$ be arbitrary. Choose $\delta > 0$ such that $\forall \xi \in [x, x + \delta)$ it holds that $\left|f(\xi) - f_+(x)\right| < \varepsilon$. For $\xi \in (x, x + \delta)$ it holds that

$$\left|\frac{F(\xi) - F(x)}{\xi - x} - f_+(x)\right| = \frac{1}{|\xi - x|}\left|\underbrace{\int_x^\xi f\,dy}_{F(\xi)-F(x)} - \underbrace{f_+(x)(\xi - x)}_{\int_x^\xi \underbrace{f_+(x)}_{\text{const.}}\,dy}\right|$$

$$= \frac{1}{|\xi - x|}\left|\int_x^\xi (f - f_+(x))\,dy\right| \leq \frac{1}{|\xi - x|}\int_x^\xi \underbrace{\left|f(y) - f_+(x)\right|}_{<\varepsilon}\,dy$$

$$y \in (x, \xi) \subseteq (x, x + \delta)$$

$$< \frac{1}{\xi - x}\varepsilon \cdot \underbrace{\int_x^\xi 1\,dy}_{|\xi - x|} = \varepsilon$$

Hence, $F'_+(x) = f_+(x)$. Analogously, $F'_-(x) = f_-(x)$ for $x \in (a, b]$. □

Theorem 4.4 (Fundamental theorem of differential/integration calculus, variation 2)**.** *Let $f \in \mathcal{R}[a, b]$ and ϕ is an arbitrary antiderviative of f according to Definition 4.3. For $\alpha, \beta \in [a, b]$ arbitrary, it holds that*

$$\int_\alpha^\beta f\,dx = \phi(\beta) - \phi(\alpha)$$

Remark 4.7. *Let f be continuous and ϕ be an antiderivative of f. Hence, $\Phi'(x) = f(x) \forall x \in [a, b]$. Then it holds that*

$$\int_\alpha^\beta \Phi'\,dx = \Phi(\beta) - \Phi(\alpha)$$

"Integral of a derivative of Φ gives $\Phi(\beta) - \Phi(\alpha)$".

Lemma 4.4. *Let $A \subseteq [a, b]$ countable. $f : [a, b] \to \mathbb{R}$ is continuous and f is differentiable in every point $x \in [a, b] \setminus A$. Furthermore let $\left|f'(x)\right| \leq L$ ($L \geq 0$) for all $x \in [a, b] \setminus A$. Then f is Lipschitz continuous on $[a, b]$ with constant L, hence*

$$\left|f(x_2) - f(x_1)\right| \leq L\,|x_2 - x_1|\,\forall x_1, x_2 \in [a, b]$$

Remark 4.8. *Some people call it* differentiable almost everywhere*, but this expression collides with a different definition pronounced the same way from measure theory.*

Proof. Let $x_1, x_2 \in [a, b]$, without loss of generality: $x_1 < x_2$. Let $\varepsilon > 0$ be arbitrary. We define

$$F_\varepsilon(x) = \left|f(x) - f(x_1)\right| - (L + \varepsilon)(x - x_1)$$

for $x \in [x_1, b]$.

Let $\varepsilon > 0$ be arbitrary. We prove: $F_\varepsilon(x) \leq 0 \forall x \in [x_1, b]$. In particular: $F_\varepsilon(x_2) \leq 0$. Hence,

$$\left|f(x_2) - f(x_1)\right| \leq (L + \varepsilon) \underbrace{(x_2 - x_1)}_{|x_2 - x_1|}$$

We prove by contradiction: Assume there exists $\varepsilon > 0$ and $x_\varepsilon > x_1$ such that

$$F_\varepsilon(x_\varepsilon) = \eta > 0 \tag{4}$$

We recognize: Let $A' = [x_1, b] \cap A$ be countable.

1. hence $F_\varepsilon(A') \subseteq \mathbb{R}$ is countable
2. $F_\varepsilon(x_1) = 0, F_\varepsilon(x_\varepsilon) > 0 \implies x_\varepsilon > x_1$
3. F_ε is continuous on $[x_1, b]$. It holds that $0 \in F_\varepsilon([x_1, x_\varepsilon])$ because $0 = F_\varepsilon(x_1)$ and $\eta \in F_\varepsilon([x_1, x_\varepsilon])$ because $\eta = F_\varepsilon(x_\varepsilon)$.

   By the Intermediate Value Theorem, it follows that $[0, \eta] \subseteq F_\varepsilon([x_1, x_\varepsilon])$ where $[0, \eta]$ is uncountable. $F_\varepsilon(A')$ is countable, hence there exists $\gamma \in (0, \eta]$ such that $\gamma = F_\varepsilon(y)$ and $\gamma \notin A'$ $(\gamma > 0)$; compare with inequality (4). Hence, $y \notin A'$. So f in y is differentiable. Let $B := F_\varepsilon^{-1}(\{\gamma\}) \cap ([x_1, x_\varepsilon] \setminus A')$, where we can skip A'. Then $B \neq \emptyset$.

   $B \subseteq [x_1, x_\varepsilon]$ is therefore bounded, $B \neq 0$. Hence, B has a supremum. Let $x = \sup B$. Choose $(y_n)_{n \in \mathbb{N}}$ with $y_n \in B$ and $y_n \to x$ for $n \to \infty$. Because F_ε is continuous, it holds that

   $$\lim_{n \to \infty} \underbrace{F_\varepsilon(y_n)}_{\gamma} = F_\varepsilon(x)$$

   hence $F_\varepsilon(x) = \gamma$. This implies $x \notin A$.

   Furthermore it holds for $w \in (x, x_\varepsilon]$ that $F_\varepsilon(w) > \gamma$. Because assume the opposite ($F_\varepsilon(w) \leq \gamma$ for $w > x$). Furthermore it holds that $F_\varepsilon(x_\varepsilon) = \eta \geq \gamma$. Because of the Intermediate Value Theorem, $\exists y \geq w$ with $F_\varepsilon(y) = \gamma$. This contradicts with the supremum property of x.

Now let $y \in (x, x_\varepsilon]$.

$$\begin{aligned}\varphi(y) &= \frac{F_\varepsilon(y) - F_\varepsilon(x)}{y - x} \\ &\underbrace{=}_{\substack{\text{definition of} \\ F_\varepsilon}} \frac{\left|f(y) - f(x_1)\right| - \left|f(x) - f(x_1)\right|}{y - x} - \frac{(L + \varepsilon)(y - x_1 - x + x_1)}{y - x} \\ &\underbrace{\le}_{\text{reversed triangle ineq.}} \frac{f(y) - f(x)}{y - x} - (L + \varepsilon)\end{aligned}$$

Because $F_\varepsilon(y) > \gamma = F_\varepsilon(x)$ it holds that $\varphi(y) > 0$ for $y > x$. So,

$$\frac{\left|f(y) - f(x)\right|}{y - x} \ge L + \varepsilon$$

$$\left|f'(x)\right| = \lim_{y \to x_+} \left|\frac{f(y) - f(x)}{y - x}\right| \ge L + \varepsilon$$

This contradicts with the boundedness of the derivative by L and f is in $x \notin A$ differentiable.

So, inequality (4) does not hold. Therefore $\forall x_1, x_2$ with $x_1 < x_2$ in $[a, b]$ and $\forall \varepsilon > 0$,

$$\left|f(x_2) - f(x_1)\right| \le (L + \varepsilon)\left|x_2 - x_1\right|$$
$$\implies \left|f(x_2) - f(x_1)\right| \le L\left|x_2 - x_1\right|$$

□

Corollary (Corollary to Lemma 4.4)**.** *Let $f, g : [a, b] \to \mathbb{R}$ differentiable for all points $x \in [a, b] \setminus A$ and A is countable. Furthermore let $f'(x) = g'(x) \forall x \notin A$. Then there exists $K \in \mathbb{R}$ such that $f(x) = g(x) + K \forall x \in [a, b]$.*

Proof. Let $h = f - g$. Then it holds that

$$h'(x) = f'(x) - g'(x) = 0 \forall x \in [a, b] \setminus A$$

By Lemma 4.4 with $L = 0$, it follows that

$$\left|h(x_1) - f(x_2)\right| \le 0 \cdot |x_1 - x_2| = 0$$
$$\implies h(x_1) = h(x_2) \forall x_1, x_2 \in [a, b]$$

Hence, $h(x) = K \in \mathbb{R}$.

$$\implies f(x) = g(x) + h(x) = g(x) + K$$

□

This lecture took place on 2018/04/19.

By reference (*), $\gamma \in [0, \eta)$ (uncountable) and $\gamma \notin f(A)$ (countable).

$$\implies \forall u \in [x_1, b) \text{ with } F_\varepsilon(u) = \gamma$$

it holds that $u \notin A$, hence f is differentiable in u.

Proof of Theorem 4.4. Let $f \in \mathcal{R}[a, b]$, ϕ is an antiderivative of f, hence $\phi'_+ = f_+$, $\phi'_- = f_-$. Let $\alpha \in [a, b]$ be arbitrary. By the Theorem variant 1, $F(x) = \int_\alpha^x f\,d\xi$ is also an antiderivative of f. By Lemma 4.4, $\exists K \in \mathbb{R} : F(x) = \int_\alpha^x f\,d\xi = \phi(x) + K$. Determine K: Let $x = \alpha \implies F(\alpha) = \int_\alpha^\alpha f\,dx = 0 = \phi(\alpha) - K$ hence $K = \phi(\alpha)$. Hence,

$$\int_\alpha^x f\,d\xi = \phi(x) - \phi(\alpha)$$

Let $x = \beta$. □

Remark 4.9 (Remark for the previous corollary)**.** *F, ϕ are differentiable on all points x for which f is continuous (all of them except for countable many). For those x, it holds that $F'(x) = \varphi'(x) = f(x)$.*

Remark 4.10 (Notation)**.** *Let $f \in \mathcal{R}[a, b]$. Then*

$$\int f\,dx$$

- *is some particular antiderivative of f (usually some arbitrary chosen)*
- *the set of* all *antiderivatives of f*

$$\int f\,dx = \{F : F \text{ is antiderivative of } f\}$$

If F_0 is some fixed antiderivative, then

$$\int f\,dx = \{F_0 + K : K \in \mathbb{R}\}$$

Then $\int f\,dx$ is the so-called indefinite integral of f. *Notation:*

$$\int x^k\,dx = \frac{x^{k+1}}{k+1} + c \qquad (k \neq -1)$$

f	F	remark
x^α	$\frac{x^{\alpha+1}}{\alpha+1} + c$	$\alpha \in \mathbb{R} \setminus \{-1\}$; restrict x such that x^α and $x^{\alpha+1}$ are defined
x^{-1}	$\ln x + c$ $(x > 0)$	
$\left(\frac{1}{-x}\right) \cdot (-1) = x^{-1}$	$\ln -x + c$ $(x < 0)$	
e^x	e^x	
$\sin x$	$-\cos x$	
$\cos x$	$\sin x$	
$\sinh x$	$\cosh x$	
$\cosh x$	$\sinh x$	
$\frac{1}{1+x^2}$	$\arctan x$	
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin x$	$\lvert x\rvert < 1$
$-\frac{1}{\sqrt{1-x^2}}$	$\arccos x$	

Table 1: Table of antiderivatives

4.1 Integration methods

In this chapter, we discuss how to determine the antiderivative of a function. Usually they are composites of basic functions. Some of these are given in Table 1.

Remark 4.11. *Let $F, G : [a, b] \to \mathbb{R}$ in $x \in [a, b)$ right-sided differentiable. Then also $F \cdot G$ in x is right-sided differentiable and it holds that*

$$(F \cdot G)'_+(x) = F'_+(x) \cdot G(x) + F(x) \cdot G'_+(x)$$

hence the product law holds.

Analogously, the same holds for the left-sided derivative.

Look up the proof in the course Analysis 1.

4.1.1 Partial integration

Definition 4.4 (Partial integration)**.** *Let f, g be given. Let F, G be its antiderivatives respectively. Then $F \cdot G$ is an antiderivative of $F \cdot g + f \cdot G$.*

This is immediate, because

$$(F \cdot G)'_+ = F'_+ \cdot G + F \cdot G'_+ = f_+ \cdot G + F \cdot g_+ = f_+ G_+ + F_+ \cdot g_+$$

Hence, it holds that

$$\int_a^b (Fg + fG)\, dx = \underbrace{F(b) \cdot G(b) - F(a)G(a)}_{=: F \cdot G|_a^b}$$

Usually, this is rewritten as

$$\int_a^b F \cdot g \, dx = F \cdot G|_a^b - \int_a^b fG \, dx$$

If F = u is continuously differentiable and G = v as well, then $f = u'$ and $g = v'$ and the law has the structure

$$\boxed{\int_a^b uv' \, dx = u \cdot v|_a^b - \int_a^b u'v \, dx}$$

Example 4.1. *Let $a \neq -1$ and $x > 0$.*

$$\int \underbrace{x^a}_{v'} \cdot \underbrace{\ln x}_{u} \, dx = \underbrace{\begin{vmatrix} u = \ln x & u' = \frac{1}{x} \\ v' = x^\alpha & v = \frac{x^{\alpha+1}}{\alpha+1} \end{vmatrix}}_{\text{scribble notes}} \quad \frac{x^{\alpha+1}}{\alpha+1} \cdot \ln x - \int \frac{1}{x} \cdot \frac{x^{\alpha+1}}{\alpha+1} \, dx$$

$$= \frac{x^{\alpha+1}}{\alpha+1} \cdot \ln x - \frac{1}{\alpha+1} \int x^\alpha \, dx = \frac{x^{\alpha+1}}{\alpha+1} \cdot \ln x - \frac{1}{(\alpha+1)^2} x^{\alpha+1}$$

Example 4.2. *Let $k \in \{2, 3, 4, \ldots\}$.*

$$\int \cos^k(x) \, dx = \begin{vmatrix} u = \cos^{k-1}(x) & u' = (k-1) \cdot \cos^{k-2}(x) \cdot (-\sin x) \\ v' = \cos x & v = \sin x \end{vmatrix}$$

$$\cos^{k-1}(x) \sin x + (k-1) \int \cos^{k-2}(x) \cdot \underbrace{\sin^2(x)}_{(1-\cos^2 x)} \, dx$$

$$= \cos^{k-1}(x) \cdot \sin(x) + (k-1) \int \cos^{k-2}(x) \, dx - (k-1) \int \cos^k(x) \, dx$$

Then we can use the following identity:

$$k \int \cos^k(x) \, dx = \cos^{k-1}(x) \cdot \sin(x) + (k-1) \int \cos^{k-2}(x) \, dx$$

This gives a recursive formula:

$$\int \cos^k(x) \, dx = \frac{1}{k} \cos^{k-1}(x) \cdot \frac{k-1}{k} \sin(x) + (k-1) \int \cos^{k-2}(x) \, dx$$

Analogously,

$$\int \sin^k(x) \, dx = -\frac{1}{k} \sin^{k-1}(x) \cdot \cos(x) + \frac{k-1}{k} \int \sin^{k-2}(x) \, dx$$

Let $c_m = \int_0^{\frac{\pi}{2}} \cos^m(x)\,dx$. *Then the following formula holds:*

$$\begin{aligned} c_{2n} &= \frac{2n-1}{2n} \cdot \frac{2n-3}{2n-2} \cdot \frac{2n-5}{2n-4} \cdots \frac{1}{2} \cdot \frac{\pi}{2} \\ &= \prod_{k=1}^{n} \frac{2k-1}{2k} \cdot \frac{\pi}{2} \\ c_{2n+1} &= \prod_{k=1}^{n} \frac{2k}{2k+1} \end{aligned}$$

Proof by induction. Let $n = 1$.

$$\begin{aligned} c_2 &= \int_0^{\frac{\pi}{2}} \cos^2 x\,dx = \frac{1}{2} \cos x \sin x \Big|_0^{\frac{\pi}{2}} + \frac{1}{2} \int_0^{\frac{\pi}{2}} 1\,dx = 0 - 0 + \frac{\pi}{4} \\ &= \underbrace{\prod_{k=1}^{1} \frac{2k-1}{2k}}_{\frac{1}{2}} \cdot \frac{\pi}{2} \\ c_1 &= \int_0^{\frac{\pi}{2}} \cos x\,dx = \sin x|_0^{\frac{\pi}{2}} = 1 - 0 = 1 \end{aligned}$$

$$\underbrace{\prod_{k=1}^{0} \frac{2k}{2k+1}}_{\text{empty product}} = 1$$

We make the induction step $n \to n+1$:

$$\begin{aligned} c_{2(n+1)} &= \frac{1}{2n+2} \cdot \underbrace{\cos^{2n+1}(x)}_{=0 \text{ for } x=\frac{\pi}{2}} \cdot \underbrace{\sin(x)}_{=0 \text{ for } x=0} \Bigg|_0^{\frac{\pi}{2}} + \frac{2n+1}{2n+2} \int_0^{\frac{\pi}{2}} \cos^{2n}(x)\,dx \\ &= \frac{2n+1}{2n+2} \prod_{k=1}^{n} \frac{2k-1}{2k} \cdot \frac{\pi}{2} = \prod_{k=1}^{n+1} \frac{2k-1}{2k} \cdot \frac{\pi}{2} \end{aligned}$$

$c_{2(n+1)+1}$ analogously. □

Theorem 4.5 (Wallis product)**.** *John Wallis (1616–1703), result from 1655*

Let $w_n = \prod_{k=1}^{n} \frac{(2k)^2}{(2k-1)(2k+1)} = \frac{2 \cdot 2}{1 \cdot 3} \cdot \frac{4 \cdot 4}{3 \cdot 5} \ldots$. *Then it holds that* $\lim_{n\to\infty} w_n = \frac{\pi}{2}$.

Proof.

$$\frac{\pi}{2} \cdot \frac{c_{2n+1}}{c_{2n}} = \frac{\pi}{2} \cdot \prod_{k=1}^{n} \frac{\frac{2k}{2k+1}}{\prod_{k=1}^{n} \frac{2k-1}{2k} \cdot \frac{\pi}{2}} = \prod_{k=1}^{n} \frac{(2k)^2}{(2k-1)(2k+1)} = w_n$$

It remains to show that $\lim_{n\to\infty} \frac{c_{2n+1}}{c_{2n}} = 1$ in $[0, \frac{\pi}{2}]$ it holds that $0 \leq \cos x \leq 1$.

$$\implies \cos^{2n+2}(x) \leq \cos^{2n+1}(x) \leq \cos^{2n}(x)$$

So, $c_{2n+2} \leq c_{2n+1} \leq c_{2n}$ for $n \geq 1$.

$$1 \geq \frac{c_{2n+1}}{c_{2n}}$$

$$\implies 1 \geq \frac{c_{2n+1}}{c_{2n}} \geq \frac{c_{2n+2}}{c_{2n}} = \frac{\prod_{k=1}^{n+1} \frac{2k-1}{2k} \frac{\pi}{2}}{\prod_{k=1}^{n} \frac{2k-1}{2k} \frac{\pi}{2}}$$

$$= \frac{2n+2-1}{2n+2} \to 1 \text{ for } n \to \infty$$

Because of the sandwich lemma for convergent sequences, the intermediate expression must also converge to 1, hence

$$\lim_{n\to\infty} \frac{c_{2n+1}}{c_{2n}} = 1 \qquad \wedge \qquad \frac{\pi}{2} \cdot \lim_{n\to\infty} \frac{c_{2n+1}}{c_{2n}} = \underbrace{\lim_{n\to\infty}}_{=1} w_n$$

□

4.1.2 Integration by substitution

Definition 4.5 (Integration by substitution)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous. Let $t : [\alpha, \beta] \to [a, b]$ be continuously differentiable. Let F be an antiderivative of f (F is therefore continuously differentiable). Then $F \circ t : [\alpha, \beta] \to \mathbb{R}$ is also continuously differentiable and the chain rule holds:*

$$(F \circ t)' = (F' \circ t) \cdot t' = (f \circ t) \cdot t'$$

Hence $F \circ t$ is an antiderivative of $(f \circ t) \cdot t'$. We apply it to integration:

$$\int_{\alpha}^{\beta} (f \circ t)(u) \cdot t'(u)\, du = (F \circ t)(\beta) - (F \circ t)(\alpha) = F(t(\beta)) - F(t(\alpha)) = \int_{t(\alpha)}^{t(\beta)} f(x)\, dx$$

Then we get the substitution integration method:

$$\boxed{\int_{t(\alpha)}^{t(\beta)} f(x)\, dx = \int_{\alpha}^{\beta} f(t(u)) \cdot t'(u)\, du}$$

Remark 4.12 (Mnemonic). *Consider the left-hand side and right-hand side simultaneously. Let $x = t(u)$ (expressions inside parentheses). Then $dx = t'(u) \cdot du$ (expressions on the right). Let $u = \alpha \implies x = t(\alpha)$ and $u = \beta \implies x = t(\beta)$ (interval boundaries).*

Example 4.3.

$$\int_0^1 2x\sqrt{1-x^2}\,dx$$

Usually we have some expression, we want to substitute with u.

$$1 - x^2 = u \qquad x = \sqrt{1-u} = t(u)$$

$$x = 0 = t(1) \qquad x = 1 = t(0)$$

$$dx = \frac{1}{2} \cdot \frac{1}{\sqrt{1-u}} \cdot (-1)\,du$$

$$\int_0^1 2x\sqrt{1-x^2}\,dx = \int_1^0 2 \cdot \sqrt{1-u} \cdot u \cdot \frac{1}{2}(-1)\frac{1}{\sqrt{1-u}}\,du = \int_0^1 \sqrt{u}\,du = \left.\frac{u^{\frac{3}{2}}}{\frac{3}{2}}\right|_0^1 = \frac{2}{3}$$

$$\int_0^1 2x\sqrt{\underbrace{1-x^2}_{u}}\,dx = \left|\begin{matrix} u = 1-x^2 & \\ x = 0 & \Leftrightarrow u = 1 \\ x = 1 & \Leftrightarrow xu = 0 \\ 1 \cdot du & = -2x\,dx \end{matrix}\right| = -\int_1^0 \sqrt{u}\,du = \int_0^1 \sqrt{u}\,du$$

In general: we set $h(u) = g(x)$, then it holds that $h'(u)\,du = g'(x)\,dx$.

Theorem 4.6. *Let $f, \tilde{f} \in \mathcal{R}[a,b]$ and $A \subseteq [a,b]$ countable. Furthermore $f(x) = \tilde{f}(x) \forall x \in [a,b] \setminus A$. Then it holds that*

$$\int_a^b \left|f - \tilde{f}\right|\,dx = 0$$

Then it follows especially that

$$\int_a^b f\,dx = \int_a^b \tilde{f}\,dx$$

This lecture took place on 2018/04/24.

Proof. Show: $r \in \mathcal{R}[a,b], r \geq 0$. $\int_a^b r\,dx = 0$ and $r(x) = 0$ for $x \in [a,b] \setminus A$. Then it holds that $\int_a^b r\,dx = 0$. Let r be as above. First, we show: $r_+(x) = \lim_{\xi \to x_+} r(\xi) = 0 \forall x \in [a,b)$ and also $r_-(x) = 0 \forall x \in (a,b]$.

Proof of that: Let $x \in [a,b)$ and $y = r_+(x)$ (exists because $r \in \mathcal{R}[a,b]$). Choose $\delta_n = \frac{1}{n}$. $(x, x + \frac{1}{n}) \cap [a,b)$ is an open interval with uncountable many points, so there is certainly one point in A. So there exists $\xi_n \in ((x, x + \frac{1}{n}) \cap [a,b)) \setminus A$ and $|\xi_n - x| < \delta_n = \frac{1}{n}$. Hence, $\lim_{n\to\infty} \xi_n = x$ and $r(\xi_n) = 0$. Therefore, $\lim_{n\to\infty} r(\xi_n) = 0$ where $r(\xi_n) = y = r_+(x)$.

Analogously, $r_-(x) = 0$ on $(a, b]$.

Let $\varepsilon > 0$ be arbitrary. We let $A_\varepsilon = \{w \in [a,b] \,|\, r(w) > \varepsilon\}$. We show: A_ε is finite.

Assume A_ε would have infinitely many points. Choose a sequence $(w_n)_{n\in\mathbb{N}}$ with $w_n \in A_\varepsilon$ and $w_n \neq w_m$ for $n \neq m$ (works because A_ε is infinite). $(w_n)_{n\in\mathbb{N}}$ is bounded, hence there exists a convergent subsequence $(w_{n_k})_{k\in\mathbb{N}}$ with $x = \lim_{k\to\infty} w_{n_k} \in [a,b]$ and $w_{n_k} \in [a,b]$.

Either (w_{n_k}) contains infinitely many sequence element $w_{n_k} < x$ (variant (a)) or infinitely many $w_{n_k} > x$ (variant (b)). Let variant b hold without loss of generality.

Combine all $w_{n_k} > x$ to one subsequence $(w_{n_{k_l}})_{l\in\mathbb{N}}$. This gives $\lim_{l\to\infty} w_{n_{k_l}} = x$ and $w_{n_{k_l}} > x$, thus $\lim_{l\to\infty} \underbrace{r(w_{n_{k_l}})}_{\geq \varepsilon \text{ because } w_{n_{k_l}} \in A_\varepsilon} = r_+(x) = 0$. This gives a contradiction.

A_ε must be finite.

Consider

$$A_{\frac{1}{n}} = \left\{w_1^n, \ldots, w_{m_n}^n\right\}$$

finite. Let $\varphi_n = \sum_{k=1}^{m_n} r(w_k^n) \cdot \chi_{\{w_k^n\}} \in \tau[a,b]$.

For $x = w_k^n \in A_{\frac{1}{n}}$ it holds that

$$\varphi_n(w_k^n) = \sum_{k=1}^{m_n} r(w_k^n) \cdot \underbrace{\chi_{\{w_k^n\}}(w_j^n)}_{\delta_{jk}} = r(w_j^n)$$

so $\left|\varphi_n(x) - r(x)\right| = 0 \forall x \in A_{\frac{1}{n}}$. Let $x \in [a,b] \setminus A_{\frac{1}{n}}$. Then it holds $0 \leq r(x) < \frac{1}{n}$ and for $x \notin A_{\frac{1}{n}}$ it holds that $\varphi(x) = 0$. Therefore,

$$\left|r(x) - \varphi(x)\right| = r(x) < \frac{1}{n}$$

hence $\left\|r - \varphi_n\right\|_\infty < \frac{1}{n}$. This means that $\varphi_n \to r$ uniformly on $[a,b]$. Therefore

$$\lim_{n\to\infty} \underbrace{\int_a^b \varphi_n \, dx}_{=0} = \int_a^b r \, dx = 0$$

Now we want to finish the proof of our theorem: Let $r(x) = \left|f(x) - \tilde{f}(x)\right| \geq 0$ and $r(x) = 0$ for $x \notin A$. So, $\int_a^b \left|f - \tilde{f}\right| dx = 0$ (first part proven).

$$\left|\int_a^b f \, dx - \int_a^b \tilde{f} \, dx\right| = \left|\int_a^b (f - \tilde{f}) \, dx\right| \leq \int_a^b \left|f - \tilde{f}\right| dx = 0$$

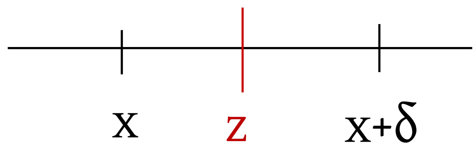

Figure 23: ξ must be sufficiently close enough to z such that $\left|f(\xi) - f_+(z)\right| \leq \frac{\varepsilon}{2}$.

$$\implies \int_a^b f\,dx = \int_a^b \tilde{f}\,dx$$

Second part proven. $\square$

Lemma 4.5. *Let $f \in \mathcal{R}[a,b]$. Then it holds that $f_+ \in \mathcal{R}[a,b]$ and also $f_- \in \mathcal{R}[a,b]$.*

Proof. Only for f_+: First, we show: Let $x \in [a,b)$.

$$f_+(x) = \lim_{\xi \to x_+} f(\xi) = \lim_{\xi \to x_+} f_+(x)$$

(the plus is important on the right-hand side!).

Proof of this: Let $\varepsilon > 0$ be arbitrary. Then there exists $\delta > 0$ such that $\forall \xi \in (x, x+\delta)$: $\left|f(\xi) - f_+(x)\right| < \frac{\varepsilon}{2}$. Now let $z \in (x, x+\delta)$ be arbitrary chosen. For z there exists $\xi \in (z, x+\delta)$ because $f_+(z)$ exists. Compare with Figure 23.

$$\left|f_+(z) - f_+(x)\right| \leq \left|f_+(z) - f(\xi)\right| + \left|f(\xi) - f_+(x)\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}$$

so $f_+(x) = \lim_{z \to x_+} f_+(z)$.

It remains to show: f_+ has left-sided limits. Let $x \in (a,b]$ be arbitrary and $f_-(x) = \lim_{\xi \to x_-} f(\xi)$. We show: $f_-(x) = \lim_{\xi \to x_-} f_+(x)$ (again: the plus is important).

Let $\varepsilon > 0$ be arbitrary. Choose $\delta > 0$ such that $\forall z \in (x-\delta, x)$ it holds that $\left|f(z) - f_-(x)\right| < \frac{\varepsilon}{2}$.

Now let $\xi \in (x-\delta, x)$ (compare with Figure 24) and choose $x > z > \xi$ with the property that $\left|f(z) - f_+(\xi)\right| < \frac{\varepsilon}{2}$ (feasible because f in ξ has a right-sided limit):

$$\left|f_+(\xi) - f_-(x)\right| \leq \underbrace{\left|f_+(\xi) - f(z)\right|}_{<\frac{\varepsilon}{2}} + \underbrace{\left|f(z) - f_-(x)\right|}_{<\frac{\varepsilon}{2}}$$

because of the choice of δ and $z \in (\xi, x) \subseteq (x-\delta, x)$.

Hence, $\lim_{\xi \to x_-} f_+(\xi) = f_-(x)$. Analogously for f_- $\square$

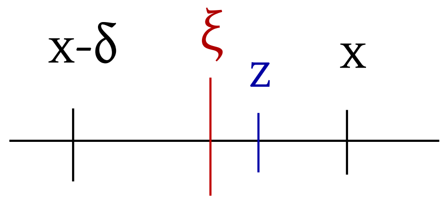

Figure 24: ξ and z

Remark 4.13.

$$\lim_{\xi \to x_+} f_+(\xi) = f_+(x)$$
$$\lim_{\xi \to x_-} f_-(\xi) = f_-(x)$$

from the proof. So f_+ is right-sided continuous and f_- is left-sided continuous.

Lemma 4.6. *Let $f \in \mathcal{R}[a,b]$. Then it holds that*

$$\int_a^b f\,dx = \int_a^b f_+\,dx = \int_a^b f_-\,dx$$

Proof. For f_+:

$$f, f_+ \in \mathcal{R}[a,b]$$

$\forall x \in [a,b]$ with f is continuous in x it holds that

$$f(x) = \lim_{\xi \to x} f(\xi) = \lim_{\xi \to x_+} f(\xi) = f_+(x)$$

f has at most countable many discontinuity points. By Satz 4.6,

$$\int_a^b \left|f - f_+\right|\,dx = 0 \quad \text{and accordingly} \quad \int_a^b f\,dx = \int_a^b f_+\,dx$$

□

4.2 Improper integrals

Let I be an interval in $\mathbb{R}$ with marginal points a and b with $-\infty \leq a < b \leq +\infty$. Let f be a regulated function on I. We define

1. If $I = [a,b)$, $\int_a^b f\,dx = \lim_{\beta \to b_-} \int_a^\beta f\,dx$

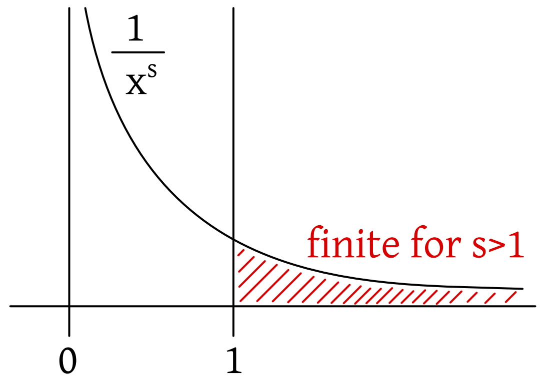

Figure 25: The function $\frac{1}{x^s}$ for $s > 1$

2. If $I = (a, b]$, $\int_a^b f\,dx = \lim_{\alpha\to a_+} \int_\alpha^b f\,dx$
3. If $I = (a, b)$, $\int_a^b f\,dx = \lim_{\alpha\to a_+} \int_\alpha^c f\,dx + \lim_{\beta\to b_-} \int_c^\beta f\,dx$

for an arbitrarily chosen $c \in (a, b)$ under the constraint that the corresponding limits in $\mathbb{R}$ exist.

Standard examples will follow:

Example 4.4. *Let $s > 1$.*

$$\int_1^\infty x^{-s}\,dx = \lim_{\beta\to\infty} \int_1^\beta x^{-s}\,dx = \lim_{\beta\to\infty} \left(\frac{1}{-s+1} x^{-s+1}\right)\Big|_1^\beta$$

$$= \frac{1}{1-s} \cdot \underbrace{\lim_{\beta\to\infty} \frac{1}{\beta^{\underbrace{s-1}_{>0}}}}_{=0} - \frac{1}{1-s} \cdot 1 = \frac{1}{s-1}$$

Compare with Figure 25.

Example 4.5. *Let $s < 1$.*

$$\int_0^1 x^{-s}\,dx = \lim_{\alpha\to 0_+} \int_\alpha^1 x^{-s}\,ds = \lim_{\alpha\to 0_+} \frac{1}{-s+1} x^{-s+1}\Big|_\alpha^1$$

$$= \frac{1}{1-s} - \frac{1}{1-s} \cdot \underbrace{\lim_{\alpha\to 0} \alpha^{\overbrace{1-s}^{>0}}}_{=0} = \frac{1}{1-s}$$

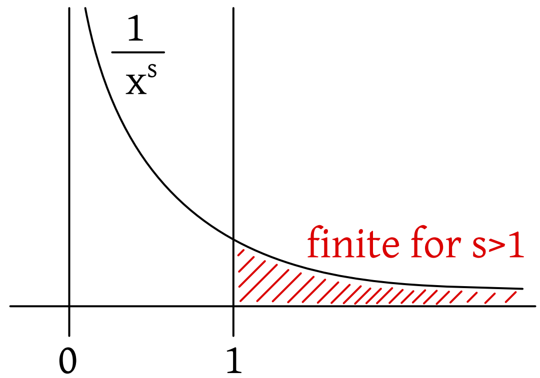

Figure 26: The function x^s for $s < 1$

Compare with Figure 26.

For $s = 1$, neither $\int_0^1 \frac{1}{x}\,dx$ nor $\int_1^\infty \frac{1}{x}\,dx$ exists.

Example 4.6. *For $c > 0$,*

$$\int_0^\infty e^{-cx}\,dx = \lim_{\beta\to\infty}\int_0^\beta e^{-cx}\,dx = \lim_{\beta\to\infty}\left(-\frac{1}{c}\right)\cdot e^{-cx}\Big|_0^\beta - \frac{1}{c}\cdot\underbrace{\lim_{\beta\to\infty} e^{-c\beta}}_{=0} + \frac{1}{c} = \frac{1}{c}$$

Theorem 4.7 (Direct comparison test for improper integrals)**.** *In German, "Majorantenkriterium für uneigentliche Intergale".*

Let f, g be regulated functions on I and it holds that

$$\left|f(x)\right| \le g(x)\forall x \in I$$

Assume $\int_a^b g\,dx$ exists as improper integral. Then also the following improper integrals exist:

$$\int_a^b \left|f\right|\,dx \text{ and } \int_a^b f\,dx$$

In German, g is called Majorante of f (there is no equivalent terminology in English).

Proof. Without loss of generality, let $I = [a, b)$. Let $G(\beta) = \int_a^\beta g\,dx$. We know that $\lim_{\beta\to b_-} G(\beta)$ exists. By Lemma 3.6 (Cauchy criterion for existence of limits): Let $\varepsilon > 0$ be arbitrary, then there exists a right-sided neighborhood U of b ($U = (b-\delta, b)$ if $b < \infty$ and $U = (M, \infty)$ if $b = \infty$) with $u, v \in U$, then it holds that $|G(v) - G(u)| < \varepsilon$.

$$|G(v) - G(u)| = \left|\int_a^v g\,dx - \int_a^u g\,dx\right| = \left|\int_u^a g\,dx + \int_a^v g\,dx\right| = \left|\int_u^v g\,dx\right|$$

Let $F(\beta) = \int_a^\beta \left|f\right| dx$. Analogously as for G, it holds that $F(v) - F(u) = \int_u^v \left|f\right| dx$. Let $u, v \in U$. Then it holds that

$$|F(v) - F(u)| = \left|\int_u^v \left|f\right| dx\right| \leq \left|\int_u^v g\, dx\right| = |G(v) - G(u)| < \varepsilon$$

hence by the Cauchy criterion for F: $\lim_{\beta \to b_-} F(\beta)$ exists, so there exists $\int_a^b \left|f\right| dx$ as improper integral. The same applies for the existence of $\int_a^b f\, dx$. $\square$

Example 4.7. *The cardinal sine function is defined as*

$$\operatorname{sinc}(x) = \frac{\sin x}{x}$$

$$\lim_{x \to 0} \frac{\sin x}{x} = 1 \qquad \operatorname{sinc}(0) = 1$$

So $\operatorname{sinc}(x)$ *is continuous on* $\mathbb{R}$*.*

$$\int_0^\infty \frac{\sin x}{x}\, dx = \underbrace{\int_0^1 \underbrace{\frac{\sin x}{x}}_{\textit{continuous}}\, dx}_{\textit{exists}} + \int_1^\infty \frac{\sin x}{x}\, dx$$

How about $\int_1^\infty \frac{\sin(x)}{x}\, dx$*?*

$$\lim_{\beta \to \infty} \int_1^\beta \frac{\sin x}{x}\, dx = \begin{vmatrix} u = \frac{1}{x} & u' = -\frac{1}{x^2} \\ v' = \sin x & v = -\cos x \end{vmatrix} = \lim_{\beta \to \infty} \left[-\frac{1}{x} \cos x\right]_1^\beta - \int_1^\beta \frac{\cos x}{x^2}\, dx$$

$$= \cos(1) - \lim_{\beta \to \infty} \int_1^\beta \frac{\cos(x)}{x^2}\, dx$$

$$\left|\frac{\cos(x)}{x^2}\right| \leq \frac{1}{x^2} \textit{ on } [1, \beta]$$

and $\int_1^\infty \frac{1}{x^2}\, dx$ *exists. So* $g(x) = \frac{1}{x^2}$ *is a majorant of* $\frac{\cos(x)}{x^2}$ *and by Theorem 4.7,* $\lim_{\beta \to \infty} \int_1^\beta \frac{\cos(x)}{x^2}\, dx$ *eixsts.*

Attention! $\int_0^\infty \left|\frac{\sin(x)}{x}\right| dx$ *does not exist. Is not Lebesgue integrable.*

Definition 4.6. *Let* $x > 0$*. We call* Γ Euler's Gamma function.

$$\Gamma(x) := \int_0^\infty t^{x-1} e^{-t}\, dx$$

Remark 4.14. *The improper integral in the definition of the* Γ*-function exists for all* $x > 0$*.*

This lecture took place on 2018/04/26.

Euler's Γ-function exists for all $x > 0$:

$$\int_0^1 t^{x-1} e^{-t}\, dt$$

for $0 < x \leq 1$:

$$t^{x-1} \cdot \underbrace{e^{-t}}_{\leq 1} \leq t^{x-1} \text{ and } \underbrace{\int_0^1 t^{\overbrace{x-1}^{\in(-1,0]}}\, dt}_{\text{exists}}$$

also exists $\int_0^1 t^{x-1}e^{-t}\, dt$ because of the direct comparison criterion. For $x \geq 1$, $t^{x-1}e^{-t}$ is continuous on $[0,1]$, hence $\int_0^1 t^{x-1}e^{-t}\, dt$ exists.

Claim. *Let $x > 0$ be fixed. $\exists c > 0$ such that*

$$t^{x-1}e^{-t} \leq c \cdot e^{-\frac{t}{2}} \forall t \in [0, \infty)$$

Proof.

$$\lim_{t\to\infty} \underbrace{t^{x-1}}_{\text{polynomially in } t} \cdot \underbrace{e^{-t}}_{\text{exponentially } \to 0} = 0$$

Also there exists $L > 1$, such that $\forall x > L$ it holds that $t^{x-1}e^{-t/2} < 1$ on $[1, L]$ (which is a compact interval) continuous. So there exists $M > 0$ such that $t^{x-1}e^{-\frac{t}{2}} \leq M \forall t \in [1, L]$. Let $c = \max\{M, 1\}$. Therefore it holds on $[1, L]$ and also on (L, ∞).

$$t^{x-1}e^{-\frac{t}{2}} \leq c$$

Multiply with $e^{-\frac{t}{2}} > 0$, then it holds that $t^{x-1} \cdot e^{-t} \leq ce^{-\frac{t}{2}} \forall t \in [1, \infty)$.

$$c \int_1^\infty e^{-\frac{t}{2}}\, dt$$

exists. By the direct comparison test, we get $\int_1^\infty t^{x-1}e^{-t}\, dt$ exists. □

Lemma 4.7. *For all $x > 0$ it holds that*

$$\Gamma(x+1) = x \cdot \Gamma(x) \qquad \textit{(functional equation of the } \Gamma\textit{-function)}$$

Especially with $\Gamma(1) = 1$ it holds that $\Gamma(n+1) = n!$ for all $n \in \mathbb{N}_0$.

Proof.

$$\Gamma(x+1) = \int_0^\infty t^{x+1-1}e^{-t}\, dt = \int_0^\infty t^x e^{-t}\, dt$$

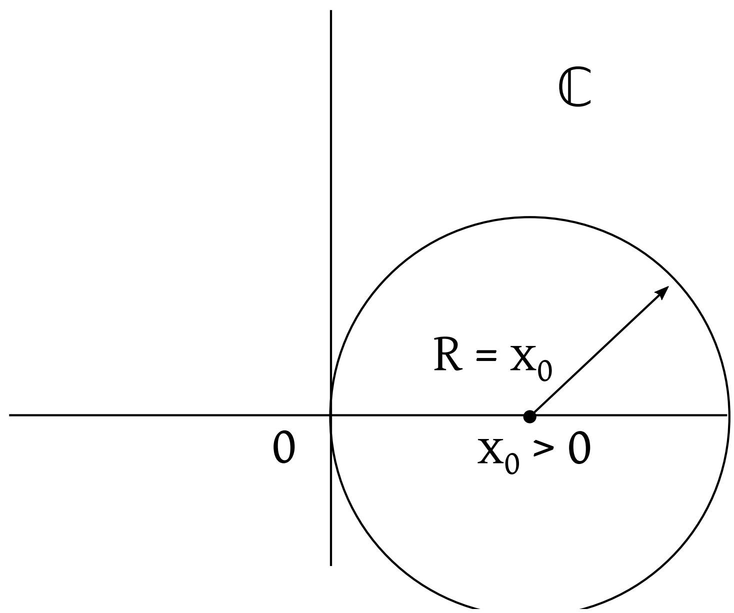

Figure 27: Γ on $\mathbb{C}$

$$= \begin{vmatrix} u = t^x & u' = x \cdot t^{x-1} \\ v' = e^{-t} & v = -e^{-t} \end{vmatrix}$$

$$= \underbrace{-t^x \cdot e^{-t}\Big|_0^\infty}_{\substack{=0 \text{ on the upper bound} \\ =0 \text{ on the lower bound}}} + \int_0^\infty x \cdot t^{x-1} \cdot e^{-t}\, dt = x \int_0^\infty t^{x-1} e^{-t}\, dt = x\Gamma(x)$$

$$\Gamma(1) = \int_0^\infty \underbrace{t^{1-1}}_{=1} \cdot e^{-t}\, dt = -e^{-t}\Big|_0^\infty = 1$$

$$\Gamma(n+1) = n \cdot \Gamma(n) = n \cdot (n-1)\Gamma(n-1) = n \cdot (n-1) \cdot \ldots \cdot 1 \cdot \underbrace{\Gamma(1)}_{=1} = n!$$

□

Remark 4.15. *There exists a power series* $\Gamma(x) = \sum_{n=0}^\infty a_n (x - x_0)^n$. $\Gamma(z)$ *is also defined for* $z \in \mathbb{C}$ *with* $\Re z > 0$. *Compare with Figure 27.*

4.3 Young's inequality

Some important inequalities in integration theory follow.

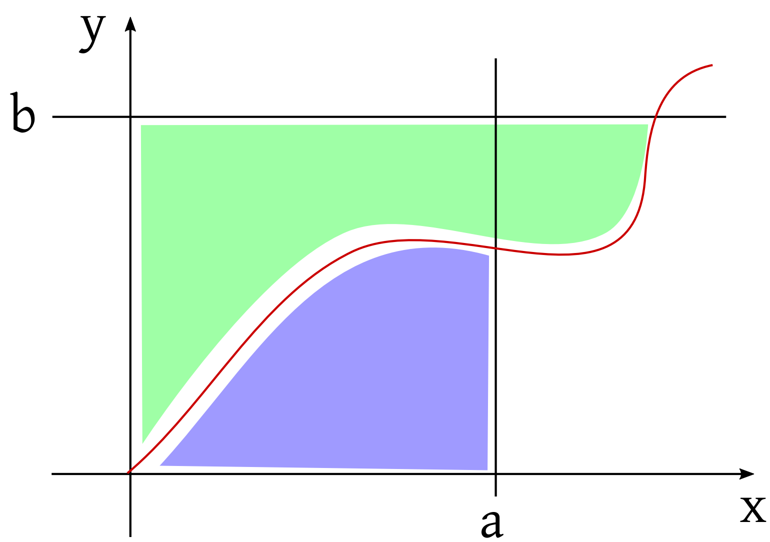

Figure 28: Young's inequality visualized. The blue area denotes $\int_0^\alpha f\,dx$ and $\int_0^b f^{-1}(y)\,dy$ is the green area.

Theorem 4.8 (Young's inequality)**.** *Let $f : [0,\infty) \to [0,\infty)$ be continuous differentiable, strictly monotonically increasing with $f(0) = 0$ and f is unbounded. Then $f : [0,\infty) \to [0,\infty)$ bijective and $f^{-1} : [0,\infty) \to [0,\infty)$ is strictly monotonically increasing and continuous. Let $a, b \geq 0$ be given. Then it holds that*

$$ab \leq \int_0^\alpha f(x)\,dx + \int_0^b f^{-1}(y)\,dy$$

Equality is given if and only if, $b = f(a)$ or $a = f^{-1}(b)$. Compare with Figure 28.

Proof. Let $f : [0,\infty) \to [0,\infty)$ be as above. Let $x_1 \neq x_2$. Without loss of generality $x_1 < x_2$. Then it holds that $f(x_1) < f(x_2) \implies f$ is injective. Surjectivity: $f(0) = 0$, hence $0 \in f([0,\infty])$. Let $\eta > 0$ be arbitrary. Because f is unbounded, there exists $z \in (0,\infty)$ with $f(z) > \eta$. $f(0) = 0 < \eta < f(z)$.

By the Intermediate Value Theorem (f is continuous), there exists $\xi \in (0,z)$ with $f(\xi) = \eta$. So f is surjective.

$$f^{-1} : [0,\infty) \to [0,\infty)$$

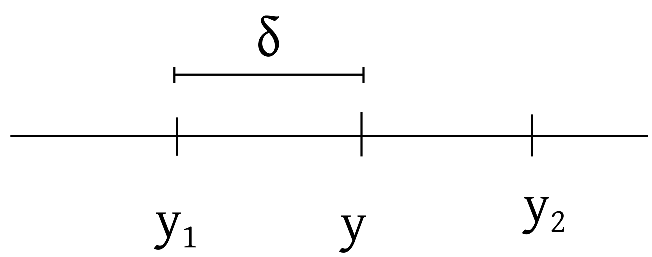

Figure 29: δ, y, y_1 and y_2

Monotonicity: Let $y_1 < y_2$. Then it holds that $x_1 = f^{-1}(y_1) < x_2 = f^{-1}(y_2)$. If this would not be true (hence, $x_2 \leq x_1$) then $y_2 = f(x_2) \leq y_1 = f(x_1)$ gives a contradiction.

Continuity of f^{-1}: Let $\varepsilon > 0$ be arbitrary. Let $y \in (0, \infty)$ be chosen arbitrarily. We show f^{-1} is continuous in y. Let $x = f^{-1}(y) > 0$ and choose $\hat{\varepsilon} = \min\left\{\frac{x}{2}, \frac{\varepsilon}{2}\right\}$.

$$x_1 = x - \hat{\varepsilon} > 0 \qquad x_2 = x + \hat{\varepsilon} > 0$$

Let $y_1 = f(x_1)$, $y_2 = f(x_2)$, $x_1 = f^{-1}(y_1)$ and $x_2 = f^{-1}(y_2)$. By monotonicity of f: $x_1 < x < x_2 \implies y_1 < y < y_2$.

Choose $\delta = \min\{y - y_1, y_2 - y\} > 0$ (compare with Figure 29). Hence $(y - \delta, y + \delta) \subseteq (y_1, y_2) \forall \eta \in (y - \delta, y + \delta)$ it holds that

$$f^{-1}(\eta) < f^{-1}(y + \delta) < f^{-1}(y_2) = x_2 = x + \hat{\varepsilon}$$

$$f^{-1}(\eta) < f^{-1}(y - \delta) < f^{-1}(y_1) = x_1 = x - \hat{\varepsilon}$$

So $f^{-1}(\eta) \in (x - \hat{\varepsilon}, x + \hat{\varepsilon})$, and accordingly

$$\left|\eta - y\right| < \delta \implies \left|f^{-1}(\eta) - \underbrace{f^{-1}(y)}_{=x}\right| < c \leq \frac{\varepsilon}{2} < \varepsilon$$

So f^{-1} is continuous in y and f^{-1} is continuous in y_0 analogously. $\square$

Consider

$$\int_0^b f^{-1}(y)\,dy = \left|\begin{array}{ll} y & = f(x) \\ dy & = f'(x)\,dx \\ y = 0 & \implies x = f^{-1}(0) = 0 \\ y = b & \implies x = f^{-1}(b) \end{array}\right| = \int_0^{f^{-1}(b)} \underbrace{f^{-1}(f(x))}_{=x} \cdot f'(x)\,dx = \int_0^{f^{-1}(b)} x \cdot f'(x)\,dx$$

$$\underbrace{=}_{\text{integration by parts}} x \cdot f(x)\Big|_0^{f^{-1}(b)} - 0 \int_0^{f^{-1}(b)} 1 \cdot f(x)\,dx$$

$$= f^{-1}(b) \cdot b - \int_0^{f^{-1}(b)} f(x)\,dx$$

So

$$I = \int_0^a f(x)\,dx + \int_0^b f^{-1}(y)\,dy = \int_{f^{-1}(b)}^0 f(x)\,dx + b \cdot f^{-1}(b)$$
$$= \int_{f^{-1}(b)}^a f(x)\,dx + b \cdot f^{-1}(b)$$

Case 1 $a = f^{-1}(b)$

$$\implies I = \underbrace{\int_a^a f(x)\,dx}_{=0} + b \cdot a$$

Case 2 $b < f(a)$, and accordingly $f^{-1}(b) < a$

$$\implies \int_{f^{-1}(b)}^a \underbrace{f(x)}_{f(f^{-1}(b)) \text{ for } x > f^{-1}(b)} dx > \overbrace{b}^{\text{minimal value}} \cdot \underbrace{(a - f^{-1}(b))}_{\text{length of integration interval}}$$

Therefore $I > b(a - f^{-1}(b)) + b \cdot f^{-1}(b) = ab$.

Case 3 $b > f(a)$, and accordingly $f^{-1}(b) > a$

$$\int_{f^{-1}(b)}^a f(x)\,dx = \int_a^{f^{-1}(b)} \underbrace{(-f(x))}_{\substack{\text{monotonically decreasing} \\ > -f(f^{-1}(b)) \forall x \in [a, f^{-1}(b))}} dx > -f(f^{-1}(b)) \cdot (f^{-1}(b) - a)$$

$$= -b(f^{-1}(b) - a)$$
$$I > -b(f^{-1}(b) - a) + b \cdot f^{-1}(b) = ab$$

Remark 4.16. *Young's inequality also holds without requiring differentiability of f (but the proof is more complex).*

Lemma 4.8 (Special case of Young's inequality)**.** *Let $A, B \geq 0$ and $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1 \iff p + q = p \cdot q$. Then p and q are called* conjugate exponents*. Then it holds that $AB \leq \frac{A^p}{p} + \frac{B^q}{q}$.*

Proof.

$$f(x) = x^{p-1} \text{ in Young's inequality}$$

$$y = x^{p-1} \iff x = y^{\frac{1}{p-1}}$$

$$\frac{1}{p-1} = q - 1 \text{ is immediate, because}$$

$$\frac{1}{p-1} = q - 1 \iff 1 = pq - p - q + 1 \iff p + q = pq$$

So $f^{-1}(y) = y^{\frac{1}{p-1}} = y^{q-1}$. By Young's inequality:

$$AB \leq \int_0^A x^{p-1}\,dx + \int_0^B y^{q-1}\,dy$$

$$= \frac{x^p}{p}\bigg|_0^A + \frac{y^q}{q}\bigg|_0^B = \frac{A^p}{p} + \frac{B^q}{q}$$

□

Remark 4.17.

$$AB = \frac{A^p}{p} + \frac{B^q}{q}$$

Equality holds if and only if $B = A^{p-1} \iff B^q = A^{\overbrace{pq - q}^{p}} = A^p$.

4.4 Hölder's ineqaulity

Theorem 4.9 (Hölder's inequality)**.** *Let I be an interval with boundary values a and b. $-\infty \leq a < b \leq +\infty$. Let p and q be conjugate exponents. Let f_1 and f_2 be regulated function on I such that*

$$\int_a^b \left|f_1(x)\right|^p\,dx < \infty$$

$$\int_a^b \left|f_2(x)\right|^q\,dx < \infty$$

both exist.

We let $\left\|f_1\right\|_p := \left(\int_a^b \left|f_1(x)\right|^p \, dx\right)^{\frac{1}{p}}$ *and* $\left\|f_2\right\|_q := \left(\int_a^b \left|f_2(x)\right|^q \, dx\right)^{\frac{1}{q}}$. *They are called* L^p*-norm of* f_1 *and* L^q*-norm of* f_2.
Then it holds that

$$\int_a^b \left|f_1(x) \cdot f_2(x)\right| dx < \infty$$

exists and

$$\int_a^b \left|f_1(x) \cdot f_2(x)\right| dx \leq \left\|f_1\right\|_p \cdot \left\|f_2\right\|_q$$

Proof. Assume that $\left\|f_1\right\|_p > 0$ and $\left\|f_2\right\|_q > 0$. Let $A = \frac{|f_1(x)|}{\|f_1\|_p}$ and $B = \frac{|f_2(x)|}{\|f_2\|_q}$. By Lemma 4.8,

$$\frac{\left|f_1(x)\right|}{\left\|f_1\right\|_p} \cdot \frac{\left|f_2(x)\right|}{\left\|f_2\right\|_q} \leq \frac{1}{q} \cdot \frac{\left|f_1(x)\right|^p}{\left\|f_1\right\|_p^p} + \frac{1}{q} \cdot \frac{\left|f_2(x)\right|^q}{\left\|f_2\right\|_q^q}$$

We integrate the inequality,

$$\frac{1}{\left\|f_1\right\|_p \cdot \left\|f_2\right\|_q} \cdot \int_a^b \left|f_1(x) \cdot f_2(x)\right| dx$$

$$\leq \frac{1}{p} \cdot \frac{1}{\left\|f_1\right\|_p^p} \cdot \underbrace{\int_a^b \left|f_1(x)^p\right| dx}_{=\|f_1\|_1^p} + \frac{1}{q} \cdot \frac{1}{\left\|f_2\right\|_q^q} \underbrace{\int_a^b \left|f_2(x)\right|^q dx}_{=\|f_2\|_q^q} = \frac{1}{p} + \frac{1}{q} = 1$$

$$\frac{1}{\left\|f_1\right\|_p \cdot \left\|f_2\right\|_q} \cdot \int_a^b \left|f_1(x) \cdot f_2(x)\right| dx \implies \int_a^b \left|f_1(x) f_2(x)\right| dx \leq \left\|f_1\right\|_p \cdot \left\|f_2\right\|_q$$

Special case: Let $\left\|f_1\right\|_p = 0$

$$\implies \left(\int_a^b \left|f_1(x)\right|^p dx\right)^{\frac{1}{p}} = 0 \implies \int_a^b \underbrace{\left|f_1(x)\right|^p}_{\geq 0} dx = 0$$

By Theorem 4.6, $f_1(x) = 0 \forall x \in [a, b] \setminus A$ and A is at most countable.

$$\implies f_1(x) \cdot f_2(x) = 0 \forall x \in [a, b] \setminus A$$

$$\implies \int_a^b \left|f_1(x) \cdot f_2(x)\right| dx = 0$$

$$\implies 0 = 0 \text{ in Hölder's inequality}$$

□

Remark 4.18 (Special case of Hölder's inequality)**.** *Let $p = q = 2$, $\frac{1}{2} + \frac{1}{2} = 1$.*

$$\int_a^b \left|f_1(x) \cdot f_2(x)\right| \, dx \leq \left\|f_1\right\|_2 \left\|f_2\right\|_2$$

is called Cauchy-Schwarz inequality for L^2 functions.

$$\int_a^b f_1(x) f_2(x) \, dx = \langle f_1, f_2 \rangle_2 = \langle f_1, f_2 \rangle_{L^2}$$

is an inner product on a proper space of functions.

5 Elaboration on differential calculus

We consider a metric space X and functions $f : X \to \mathbb{C}$. We define a concept of uniform convergence of such sequences:

$$f_n : X \to \mathbb{C} \quad (n \in \mathbb{N}) \text{ and } f : X \to \mathbb{C}$$

We say, $(f_n)_{n\in\mathbb{N}}$ converges uniformly towards f if $\forall \varepsilon > 0 \forall N \in \mathbb{N}$ such that $\forall x \in X$ and $\forall n \geq N$ it holds that

$$\underbrace{\left|f_n(x) - f(x)\right|}_{\text{absolute value in } \mathbb{C}} < \varepsilon$$

$$\iff \sup\left\{\left|f_n(x) - f(x)\right| : x \in X\right\} < \varepsilon$$

Remark 5.1. *Do not use $\left\|f\right\|_\infty$ for the definition of uniform convergence, because f_n and f must not be necessarily bounded. Hence,*

$$\left\|f\right\|_\infty = \left\{\left|f(x)\right| : x \in X\right\}$$

must not be finite.

Theorem 5.1. *Let X be a metric space, $f_n : X \to \mathbb{C}$ be a sequence of continuous functions and $f : X \to \mathbb{C}$ such that $f_n \to f$ uniform on X. Then f is also continuous on X.*

This lecture took place on 2018/05/03.

Proof. Let $\varepsilon > 0$ be arbitrary. Choose $x \in X$. Show: f is continuous in x.

Compare with Figure 30.

Because of uniform convergence $f_n \to f$, there exists $N \in \mathbb{N}$ such that $\left|f_N(z) - f(z)\right| < \frac{\varepsilon}{3} \forall z \in X$. Let N be fixed. Because f_N is continuous in x, there exists $\delta > 0$ such that $d(x, \xi) < \delta \implies \left|f_N(\xi) - f_N(x)\right| < \frac{\varepsilon}{3}$.

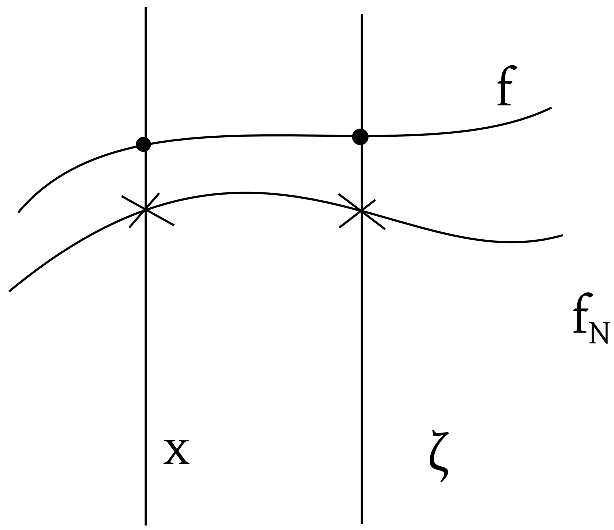

Figure 30: Uniform convergence of f_N to f

We consider now $\xi \in X$ with $d_X(x, \xi) < \delta$. Then it holds that

$$\begin{aligned}\left|f(x) - f(\xi)\right| &= \left|f(x) - f_N(x) + f_N(x) - f_N(\xi) + f_N(\xi) - f(\xi)\right| \\ &\leq \underbrace{\left|f(n) - f_N(x)\right|}_{<\frac{\varepsilon}{2}} + \underbrace{\left|f_N(x) - f_N(\xi)\right|}_{<\frac{\varepsilon}{3}} + \underbrace{\left|f_N(\xi) - f(\xi)\right|}_{<\frac{\varepsilon}{3}} \\ &= \varepsilon\end{aligned}$$

by uniform convergence, by continuity and by uniform convergence respectively.

Thus, f is continuous in x. □

Theorem 5.2. *Let $P(z) = \sum_{k=0}^{\infty} a_k z^k$ be a power series in $\mathbb{C}$ with convergence radius $\rho_P > 0$. Furthermore, let $0 < r < \rho_P$. Let $P_n(z) = \sum_{k=0}^{n} a_k z^k$ (n-th partial sum of P). Then $P_n \to P$ uniformly on $\overline{K_r(0)}$.*

Proof. Approximation theorem for power series. Lettl Analysis 1, lecture notes, section 5, theorem 10.

Let $0 < r < \rho_P$. Choose $\overline{r}$ with $r < \overline{r} < \rho_P$. Then it holds for $z \in \overline{K_r(0)}$ that

$$|P(z) - P_n(z)| < \frac{\overline{r}}{\overline{r} - r} \cdot \left(\frac{r}{\overline{r}}\right)^n$$

Remark 5.2.

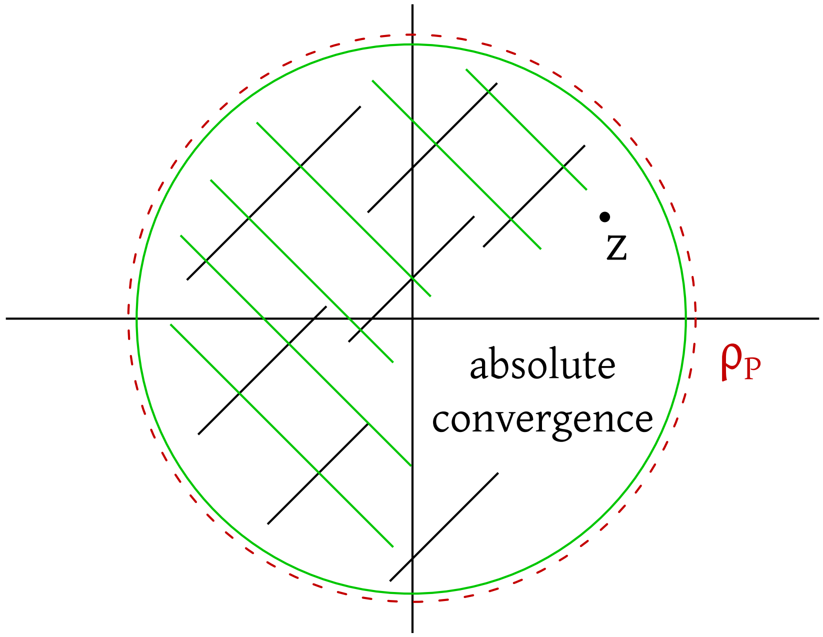

Figure 31: We cannot make a general statement about convergence/divergence. But on every small closed sphere P converges absolutely for every z

$$\frac{r}{\overline{r}} < 1$$

hence $\left(\frac{r}{\overline{r}}\right)^n$ is arbitrary small, for every n sufficiently large.

$$\implies \sup\left\{\left|P(z) - P_n(z) : z \in \overline{K_r(0)}\right|\right\} \leq \underbrace{\frac{\overline{r}}{\overline{r} - r}}_{\text{fixed}} \cdot \underbrace{\left(\frac{r}{\overline{r}}\right)^n}_{\substack{\text{arbitrary small for } n \\ \text{sufficiently large}}}$$

Hence, $P_n \to P$ uniform on $\overline{K_r(0)}$. □

Corollary. *P is continuous on $K_{\rho_P}(0)$.*

Theorem 5.3. *Let $I \subseteq \mathbb{R}$ be an interval. Let $f_n : I \to \mathbb{R}$ be continuously differentiable on $I \forall n \in \mathbb{N}$. It holds that*

1. *$\exists g : I \to \mathbb{R}$ such that $f_n' \to g$ uniform on I*
2. *$\exists f : I \to \mathbb{R}$ such that $\forall x \in I$ it holds that $f(x) = \lim_{n\to\infty} f_n(x)$ ("pointwise convergence").*

Then it holds that f is continuously differentiable on I and $g = f'$.

Proof. g is continuous as uniform limit of continuous f_n' (Theorem 5.1). For f_n, the Fundamental Theorem of Differential Calculus can be applied (f_n' is continuous, hence a regulated function). Let $x_0 \in I$. Then it holds that

$$f_n(x) = f_n(x_0) + \int_{x_0}^{x} f_n'(\xi)\, d\xi$$

Convergence for $n \to \infty$:

$$f_n(x) \to f(x) \qquad f_n(x_0) \to f(x_0)$$

(Pointwise convergence)

$$\int_{x_0}^{x} f_n'(\xi)\, d\xi \to \int_{x_0}^{x} g(\xi)\, d\xi$$

Therefore, for $n \to \infty$,

$$f(x) = f(x_0) + \int_{x_0}^{x} g(\xi)\, d\xi$$

The right-hand side is continuously differentiable by x according to the Fundamental Theorem, variant 1, with

$$\left(f(x_0) + \int_{x_0}^{x} g(\xi)\, d\xi\right)'(x) = g(x)$$

Hence, by $f(x) = f(x_0) + \int_{x_0}^{x} g(\xi)\, d\xi$ it follows that

$$f'(x) = g(x) \qquad \forall x \in I$$

□

To finish our proof, we need a result we missed in the section about Integrals.

Lemma 5.1. *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of regulated functions on $[a, b]$ and $f_n \to f$ uniform on $[a, b]$. Then it holds that*

$$\int_a^b \left|f_n - f\right| dx \to 0 \qquad \text{for } n \to \infty \qquad \text{especially } \int_a^b f_n\, dx \to \int_a^b f\, dx$$

Proof. f as a uniform limit of regulated functions is a regulated function. The proof has been done in the practicals.

Let $N \in \mathbb{N}$ large enough such that

$$\forall n \geq N \forall x \in [a, b] : \left|f_n(x) - f(x)\right| < \frac{\varepsilon}{b-a}$$

Then it holds that

$$\int_a^b \left|f_n(x) - f(x)\right| dx < \int_a^b \frac{\varepsilon}{b-a}\, dx = \frac{\varepsilon}{b-a}(b-a) = \varepsilon$$

Hence,

$$\lim_{n\to\infty}\int_a^b \left|f_n(x)-f(x)\right|\,dx = 0$$

$$\underbrace{\left|\int_a^b f_n\,dx - \int_a^b f\,dx\right|}_{\implies\to 0} \le \underbrace{\int_a^b \left|f_n - f\right|\,dx}_{\to 0}$$

So,

$$\int_a^b f\,dx = \lim_{n\to\infty}\int_a^b f_n\,dx$$

□

5.1 Higher derivatives and Taylor's Theorem

Definition 5.1. *Let $f : I \to \mathbb{R}$, $I \subseteq \mathbb{R}$ is an interval. We define inductively:*

$$f^{(0)}(x) = f(x)$$

Assume $f^{(n-1)}$ is defined continuously on I and differentiable in $x \in I$. Then we let

$$f^{(n)}(x) = \left(f^{(n-1)}\right)'(x)$$

$f^{(n)}(x)$ is called n-th derivative of f in x.
Notational remark:

$$f^{(0)} = f \qquad f^{(1)} = f' \qquad f^{(2)} = f'' \qquad f^{(3)} = f''' \qquad f^{(4)} = f''''$$

Furthermore, we let

$C^n(I) := \left\{f : I \to \mathbb{R} : f^{(k)}(x) \text{ exists } \forall x \in I \text{ and } x \mapsto f^{(k)}(x) \text{ is continuous } \forall 0 \le k \le n\right\}$

We call C the space of n-times continuously differentiable functions on I.

Remark 5.3. *$C^n(I)$ is a vector space. If $I = [a, b]$ is compact, then*

$$\left\|f\right\|_{C^n} = \max\left\{\sup\left|f^{(k)}(x)\right| : x \in I : 0 \le k \le n\right\}$$

defines a norm on $C^n(I)$ with $\sup\left|f^{(k)}(x)\right| : x \in I = \left\|f^{(k)}\right\|_\infty$.

Remark 5.4 (New topic)**.** *Let $f \in C^n(I)$ and $x_0 \in I$. Find an appropriate polynomial T which approximated f in an environment of x_0 in the "best" way.*

Definition 5.2. *Let $P(x) = \sum_{k=0}^n a_k x^k$ be a polynomial with $a_n \neq 0$ (hence degree of P is n).*

$$P \in \mathbb{R}[x] \ldots \text{ set of all polynomials with coefficients in } \mathbb{R}$$

This set of polynomials is a ring.

$x_0 \in \mathbb{R}$ is called k-times root of P ($k \in \mathbb{N}$) if $Q \in \mathbb{R}[x]$ exists such that $P(x) = (x - x_0)^k Q(x)$ with $Q(x_0) \neq 0$.

Remark 5.5. *$P(x) = (x - x_0)^k \cdot Q(x)$ means that division of P by $(x - x_0)^k$ gives no remainder. Recall that division with remainder means that $\exists \hat{Q}, \hat{R}$ that are polynomials of degree $\hat{R} < k$,*

$$P(x) = (x - x_0)^k \cdot \hat{Q}(x) + \hat{R}(x)$$

$\hat{Q}, \hat{R}$ is unique. If $P(x) = (x - x_0)^k \cdot Q(x) \implies \hat{R} = 0, \hat{Q} = Q$.

Lemma 5.2. *Let $P(x) = \sum_{l=0}^{n} a_l x^l$ with $a_n \neq 0$. Let $1 \leq k \leq n$. Then it holds that $x_0 \in \mathbb{R}$ is a k-times root of polynomial $P \iff P^{(j)}(x_0) = 0$ for $j = 0, \dots, k-1$ and $P^{(k)}(x_0) \neq 0$.*

Proof. Proof by complete induction.

Induction begin Consider $k = 1$. Direction $\implies$.

Let x_0 be a simple root of P, then it holds that $P(x) = (x - x_0) \cdot Q(x)$ and $Q(x_0) \neq 0$. Hence, $P(x_0) = (x_0 - x_0) \cdot Q(x_0) = 0$ and $P'(x) = Q(x) + (x - x_0) \cdot Q'(x)$. Thus, $P'(x_0) = Q(x_0) + (x_0 - x_0) \cdot Q'(x_0) = Q(x_0) \neq 0$.

Direction $\impliedby$.

Let $P(x_0) = 0$ and $P'(x_0) \neq 0$. Division with remainder: $P(x) = (x - x_0) \cdot Q(x) + R(x)$ with degree$(R) \leq$ degree$(x - x_0) = 1$. Thus, R is constant. We insert x_0. This gives $P(x_0) = (x_0 - x_0) \cdot Q(x_0) + R$ with $P(x_0) = 0$ and $(x_0 - x_0) = 0$. Hence, $R = 0$ is the zero polynomial and $P(x) = (x - x_0) \cdot Q(x)$. It remains to show that $Q(x_0) \neq 0$. $P'(x) = 1 \cdot Q(x_0) + (x - x_0) \cdot Q'(x_0)$. We insert $x = x_0 \implies 0 \neq P'(x_0) = Q(x_0) + (x_0 - x_0) \cdot Q'(x)$. Thus is holds that $Q(x_0) = P'(x_0) \neq 0$.

Induction step $k \to k + 1$

Claim (Auxiliary claim)**.** *Let $P(x) = (x - x_0) \cdot \tilde{P}(x)$. Let $P, \tilde{P}$ be polynomials. Then it holds $\forall j \in \mathbb{N}$ that*

$$P^{(j)}(x) = (x - x_0) \cdot \tilde{P}^{(j)}(x) + j \cdot \tilde{P}^{(j-1)}(x)$$

Proof. Proof by complete induction.

Let $j = 1$.

$$P'(x) = 1 \cdot \underbrace{\tilde{P}(x)}_{\tilde{P}^{(0)}(x)} + (x - x_0) \cdot \underbrace{\tilde{P}'(x)}_{\tilde{P}^{(1)}(x)}$$

Consider $j \to j + 1$.

$$\begin{aligned} P^{(j+1)}(x) &= \left(P^{(j)}\right)'(x) \\ &\underbrace{=}_{\substack{\text{induction} \\ \text{assumption}}} ((x - x_0) \cdot \tilde{P}^{(j)}(x) \\ &+ j\tilde{P}^{(j-1)}(x))'(x - x_0)\tilde{P}^{(j+1)}(x) + \tilde{P}^{(j)}(x) + j \cdot \tilde{P}^{(j)}(x) \\ &= (x - x_0)\tilde{P}^{(j+1)}(x) + (j + 1) \cdot \tilde{P}^{j}(x) \end{aligned}$$

□

We continue with the induction step after verifying our auxiliary claim.
Direction $\implies$.

Let x_0 be an $k+1$ times zero of P. Hence $P(x) = (x-x_0)^{k+1} \cdot Q(x)$. $Q(x_0) \neq 0$. Let $\tilde{P}(x) = (x-x_0)^k \cdot Q(x)$. We can apply the induction assumption on $\tilde{P}$. Hence

$$\tilde{P}^{(j)} = 0 \qquad \text{for } j = 0, \dots, k-1 \qquad \text{and} \qquad \tilde{P}^{(k)}(x_0) \neq 0$$

$$P(x) = (x - x_0) \cdot \tilde{P}(x)$$

By the auxiliary claim, $P^{(j)}(x) = (x-x_0) \cdot \tilde{P}^{(j)}(x) + j \cdot \tilde{P}^{(j-1)}(x)$. Therefore

$$P^{(j)}(x_0) = j \cdot \tilde{P}^{(j-1)}(x) = \begin{cases} 0 & \text{for } j = 0, \dots, k \\ (k+1)\tilde{P}^{(k)}(x_0) \neq 0 & \text{for } j = k+1 \end{cases}$$

Hence, our claim about the derivatives is true (all derivatives are zero).

Direction $\impliedby$.

Let $P^{(j)}(x_0) = 0$ for $j = 0, \dots, k$ and $P^{(k+1)}(x_0) \neq 0$ and induction assumption holds for k. Division with remainder and $P^{(0)}(x_0) = 0 \implies P(x) = (x-x_0) \cdot \tilde{P}(x)$. By our auxiliary claim, we get

$$P^{(j)}(x) = (x - x_0) \cdot \tilde{P}^{(j)}(x) + j\tilde{P}^{(j-1)}(x)$$

we insert $x = x_0$ and use $P^{(j)}(x_0) = 0$ for $j = 0, \dots, k$

$$\implies \tilde{P}^{(j)}(x_0) = 0 \qquad \text{for } j = 0, \dots, k-1$$

By the induction assumption, $\tilde{P}(x) = (x-x_0)^k Q(x)$ with $Q(x_0) \neq 0$

$$\implies P(x) = (x - x_0) \cdot \tilde{P}(x) = (x - x_0)^{k+1} Q(x)$$

□

This lecture took place on 2018/05/08.

Definition 5.3. *Let $I \subseteq \mathbb{R}$ be an interval, $f \in C^n(I)$. We let*

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

$T_f^n(x;x_0)$ is a polynomial in x with degree$(T_f^n) \leq n$. *$T_f^n(x;x_0)$ is called* Taylor polynomial of f of order n in x_0.

Brook Taylor (1685–1731)

Lemma 5.3. *The premise is the same like in Definition 5.3. The Taylor polynomial of $T_f^n(x; x_0)$ is the only polynomial of degree $\neq n$ which satisfies*

$$(T_f^n)^{(k)}(x_0) = f^{(k)}(x_0) \qquad \text{for } k = 0, \ldots, n$$

Proof. Claim:

$$(T_f^n)^{(k)}(x; x_0) = \sum_{l=k}^{n} \frac{f^{(l)}(x_0)}{(l-k)!}(x - x_0)^{l-k} \qquad \text{for } 0 \leq k \leq n$$

Proof of the claim by complete induction:

Induction base $n = 0$

$$(T_f^n)^{(0)}(x; x_0) = \sum_{l=0}^{n} \frac{f^{(l)}(x_0)}{l!}(x - x_0)^l$$

Induction step $k \to k+1$ Let $(T_f^n)^{(k)}(x; x_0)$

$$= \sum_{l=k}^{n} \frac{f^{(l)}(x_0)}{(l-k)!}(x - x_0)^{(l-k)}$$

by induction hypothesis. Then,

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

We apply insertion: $x = x_0$ into $(T_f^n)^{(k)}(x; x_0)$

$$(T_f^n)^{(k)}(x; x_0) = \sum_{l=k}^{n} \frac{f^{(l)}(x_0)}{(l-k)!}(x - x_0)^{l-k} = \frac{f^{(k)}(x_0)}{0!} = f^{(k)}(x_0)$$

We need to prove uniqueness: Let $T, \tilde{T}$ be polynomials with $T^{(k)}(x_0) = \tilde{T}^{(k)}(x_0) = f^{(k)}(x_0)$ for $k = 0, \ldots, n$. Assume $T \neq \tilde{T}$, hence $T - \tilde{T} \neq 0$ (where 0 is the zero polynomial). For $P = T - \tilde{T}$ it holds that

$$P^{(k)}(x_0) = T^{(k)}(x_0) - \tilde{T}^{(k)}(x_0) = 0 \qquad (\text{for } 0 \leq k \leq n)$$

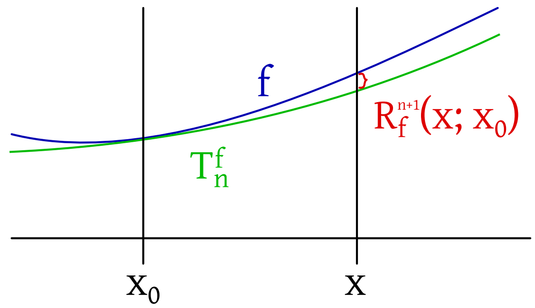

Figure 32: Visualization of the remainder term of a Taylor polynomial

By Lemma 5.2 it holds that x_0 is an $n+1$-times root of P. Thus, there exists a polynomial $Q \neq 0$ with $Q(x_0) \neq 0$ such that

$$\underbrace{P(x)}_{\text{degree } \leq n} = \underbrace{(x - x_0)^{(n+1)} \cdot Q(x)}_{\text{degree } \geq n+1}$$

This is a contradiction. Hence it holds that $T - \tilde{T} = 0$. □

Definition 5.4. *Let $f \in C^n(I)$, $x_0 \in I$. Furthermore let $T_f^n(x; x_0)$ be the Taylor polynomial of n-th degree of f in x_0. We let $R_f^n(x; x_0) = f(x) - T_f^n(x; x_0)$. We call $R_f^{n+1}(x; x_0)$ the* approximation error *of the Taylor polynomial. Also called* remainder term of $n+1$-th order. *Compare with Figure 32.*

Theorem 5.4. *Let $f^{(n+1)}(I)$, $x \in I$, $x_0 \in I$. Then it holds that*

$$R_f^{n+1}(x; x_0) = \frac{1}{n!} \int_{x_0}^{x} (x - t)^n f^{(n+1)}(t)\, dt$$

We call it the integral form of the remainder term*.*

Proof. Complete induction over n.

Induction base $n = 0$

$$T_f^0(x; x_0) = f(x_0)$$

$$
\begin{aligned}
R_f^1(x;x_0) &= \underbrace{f(x) - f(x_0)}_{f\in C^1} \\
&= \int_{x_0}^{x} f'(t)\,dt \\
&= \frac{1}{0!}\int_{x_0}^{x} (x-t)^0 f^{(1)}(t)\,dt
\end{aligned}
$$

Induction step $n-1 \to n$

$$
\begin{aligned}
R_f^n(x;x_0) &= f(x) - T_f^{n-1}(x;x_0) \\
&\underbrace{=}_{\text{ind. hypothesis}} \frac{1}{(n-1)!}\int_{x_0}^{x} (x-t)^{n-1} f^{(n)}(t)\,dt \\
&= \begin{vmatrix} u' = (x-t)^{n-1} & v = f^{(n)}(t) \\ u = -\frac{1}{n}(x-t)^n & v' = f^{(n+1)}(t) \end{vmatrix} \\
&= \frac{1}{(n-1)!}\underbrace{\left[-\frac{1}{n}(x-t)^n \cdot f^{(n)}(t)\right]}_{=\frac{1}{n!}(x-x_0)^n\cdot f^{(n)}(x_0)}\Bigg|_{t=x_0}^{x} + \underbrace{\frac{1}{(n-1)!}\int \frac{1}{n}(x-t)^n \cdot f^{(n+1)}(t)\,dt}_{=\frac{1}{n!}\int_{x_0}^{x}(x-t)^n\cdot f^{(n+1)}(t)\,dt}
\end{aligned}
$$

So,

$$
\begin{aligned}
f(x) \underbrace{-T_f^{n-1}(x;x_0) - \frac{1}{n!}f^{(n)}(x_0)(x-x_0)^n}_{-T_f^n(x;x_0)} \\
= \frac{1}{n!}\int_{x_0}^{x}(x-t)^n \cdot f^{(n+1)}(t)\,dt
\end{aligned}
$$

Therefore,

$$
R_f^{(n+1)}(x;x_0) = \frac{1}{n!}\int_{x_0}^{x}(x-t)^n f^{(n+1)}(t)\,dt
$$

□

Theorem 5.5 (Lagrange form of the remainder term)**.** *Let $f \in C^{n+1}(I)$, $n \in \mathbb{N}_0$, $x, x_0 \in I$, $x \neq x_0$. Then there exists some ξ between x_0 and x (hence, $\xi \in (x_0, x)$ if $x > x_0$ or $\xi \in (x, x_0)$ if $x < x_0$) such that*

$$
R_f^{(n+1)}(x;x_0) = \frac{1}{(n+1)!}f^{(n+1)}(\xi)(x-x_0)^{n+1}
$$

Proof. Idea: we apply the Mean Value Theorem for definite integrals on the Taylor remainder.

Case 1 Let $x_0 < x$.

$$
\begin{aligned}
R_f^{n+1}(x;x_0) &= \frac{1}{n!}\int_{x_0}^{x} \underbrace{(x-t)^n}_{\substack{\geq 0\\ \text{regulated function}}} \underbrace{f^{(n+1)}(t)}_{\text{continuous in } t}\, dt \\
&\underbrace{=}_{\text{MVT}} \frac{1}{n!} f^{(n+1)}(\xi)\cdot \int_{x_0}^{x}(x-t)^n\, dt \\
&= \frac{1}{n!} f^{(n+1)}(\xi)\left[-\frac{1}{n+1}(x-t)^{n+1}\right]_{t=x_0}^{x} \\
&= \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-x_0)^{n+1}
\end{aligned}
$$

where MVT is the Mean Value Theorem for definite integrals (Theorem 4.2).

Case 2 Let $x < x_0$ and n odd.

$$
\begin{aligned}
R_f^{n+1}(x;x_0) &= -\frac{1}{n!}\int_{x}^{x_0} \underbrace{(x-t)^n}_{=(-1)^n(t-x)^n} \cdot f^{(n+1)}(t)\, dt \\
&= \frac{1}{n!}\int_{x}^{x_0} \underbrace{(t-x)^n}_{\geq 0} \cdot \underbrace{f^{(n+1)}(t)}_{\text{continuous}}\, dt \\
&= \frac{f^{(n+1)}(\xi)}{n!}\int_{x}^{x_0}(t-x)^n\, dt \\
&= \frac{f^{(n+1)}(\xi)}{n!}\left[\frac{1}{n+1}(t-x)^{n+1}\right]\Bigg|_{x}^{x_0} \\
&= \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x_0-x)^{n+1}
\end{aligned}
$$

$n+1$ is even

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

Case 3 Let $x < x_0$ and n even.

$$\begin{aligned}
R_f^{n+1}(x;x) &= -\frac{1}{n!}\int_x^{x_0} \underbrace{(x-t)^n}_{\geq 0} \cdot \underbrace{f^{(n+1)}(t)}_{\text{continuous}}\, dt \\
&= -\frac{1}{n!} f^{(n+1)}(\xi) \cdot \int_x^{x_0} (x-t)^n\, dt \\
&= -\frac{1}{n!} f^{(n+1)}(\xi) \cdot \left[-\frac{1}{n+1}(x-t)^{n+1} \right]\Bigg|_x^{x_0} \\
&= \frac{1}{(n+1)!} f^{(n+1)}(\xi)(x-x_0)^{n+1}
\end{aligned}$$

□

Any extreme value satisfies that its derivative is zero. But not every point with derivative zero is an extreme value. We now consider conditions to select extreme values from all value satisfying derivative zero.

Corollary (Sufficient conditions for existence of extreme values)**.** *Let I be an open interval. Let $x_0 \in I$ and $f \in C^{n+1}(I)$. Assume*

$$f^{(1)}(x_0) = f^{(2)}(x_0) = \cdots = f^{(n)}(x_0) = 0$$

and $f^{(n+1)}(x_0) \neq 0$. Then f in x_0 has

1. *a strict local maximum if n is even and $f^{(n+1)}(x_0) < 0$*
2. *a strict local minimum if n is odd and $f^{(n+1)}(x_0) > 0$*
3. *no extreme value in x_0 if n is even.*

Proof. **Case a** Let $f^{(n+1)}(x_0) < 0$ and $f^{(n+1)}$ is continuous, then $\exists \varepsilon > 0$ such that $(x_0 - \varepsilon, x_0 + \varepsilon) \subseteq I$ (I is open) and $f^{(n+1)}(\xi) < 0 \forall \xi \in (x_0 - \varepsilon, x_0 + \varepsilon)$. Now let $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$. Then by Theorem 5.5,

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

for $k = 1, \ldots, n$. So,

$$f(x) - f(x_0) = \frac{\overbrace{f^{(n+1)}(\xi)}^{<0}}{\underbrace{(n+1)!}_{>0}} \underbrace{(x-x_0)^{\overbrace{n+1}^{\text{even}}}}_{>0 \text{ for } x \neq x_0}$$

hence $f(x) - f(x_0) < 0$, and accordingly

$$f(x) < f(x_0) \qquad \forall x \in (x_0 - \varepsilon, x_0 + \varepsilon), x \neq x_0$$

So f is a strict local maximum.

Case b Analogously.

Case c We apply the same idea as in Case a up to the point, where we consider $f(x) - f(x_0)$.

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

$f^{(n+1)}(\xi)$ has the same sign as $\underbrace{f^{(n+1)}(x_0)}_{\neq 0}$ $\forall \xi \in (x_0 - \varepsilon, x_0 + \varepsilon)$. This is feasible due to continuity of $f^{(n+1)}$ for sufficiently small ε.

$$f(x) - f(x_0) = \underbrace{\frac{f^{(n+1)}(\xi)}{(n+1)}}_{\text{has constant sign indep. of } x} \cdot \underbrace{(x - x_0)}_{\text{changes its sign}}{}^{\overbrace{n+1}^{\text{odd}}}$$

Therefore $f(x) - f(x_0)$ changes its sign for $x = x_0$. Hence f has no extreme value in $x = x_0$.

$\square$

Theorem 5.6 (Qualitative Taylor equation)**.** *Let $f \in C^n(I)$, $x, x_0 \in I$. Then there exists some function $r \in C(I)$ with $r(x_0) = 0$ such that*

$$f(x) = T_f^n(x; x_0) + (x - x_0)^n \cdot r(x)$$

and accordingly,

$$R_f^{n+1}(x; x_0) = (x - x_0)^n \cdot r(x)$$

Remark 5.6. *For some function r with $\lim_{x \to x_0} r(x) = 0$, we also denote $o(x - x_0)$ instead of $r(x)$. This general notation is called Landau's Big-Oh notation.*

$$f(x) = T_f^n(x; x_0) + (x - x_0)^n \cdot o(x - x_0)$$

Proof. Let $r(x) = \frac{f(x) - T_f^n(x; x_0)}{(x - x_0)^n}$ for $x \neq x_0$ and $r(x_0) := 0$. Then f is continuous and T_f^n is continuous in every point $x \neq x_0$. It remains to show that r is continuous in $x = x_0$.

$$r(x) = \frac{1}{(x - x_0)^n} (\underbrace{f(x) - T_f^{n-1}(x; x_0)}_{R_f^n(x; x_0)} - \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n)$$

$$\underbrace{=}_{\text{Lagrange}} \frac{1}{(x - x_0)} \left[\frac{1}{n!}(x - x_0)^n \cdot f^{(n-1)}(\xi) - \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n \right]$$

$\xi \in (x_0, x)$

$$= \frac{1}{n!}\left[f^{(n)}(\xi) - f^{(n)}(x_0)\right] \to 0 \text{ for } x \to x_0 \text{ because } f^{(n)} \text{ is continuous}$$

as $x_0 < x < x$, hence $\xi \to x_0$ for $x \to x_0$

So $\lim_{x\to x_0} r(x) = 0 = r(x_0)$, so r in x_0 is continuous. □

This lecture took place on 2018/05/15.

5.2 Taylor series

Assume $f : I \to \mathbb{R}$ is infinitely often differentiable on I, $x_0 \in I$. Then there exists $T_f^n(x; x_0)$ for arbitrary $n \in \mathbb{N}$.

$$T_f(x; x_0) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k$$

$T_f(x; x_0)$ defines the *Taylor series* on f in x_0. Power series in $\xi = x - x_0$. T_f has a convergence radius,

$$\rho(T_f) = \left[\limsup_{k\to\infty} \sqrt[k]{\frac{\left|f^{(k)}(x_0)\right|}{k!}}\right]^{-1}$$

If $\rho(T_f) > 0$, then it holds that

$$f(x) = T_f(x; x_0) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x)}{k!}(x - x_0)^k$$

in $(x_0 - \rho(T_f), x_0 + \rho(T_f))$? Compare with Figure 33.

Example 5.1 (Counterexample)**.** *Let $f : \mathbb{R} \to \mathbb{R}$.*

$$f(x) = \begin{cases} e^{-\frac{1}{x}} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

It holds for $x > 0$,

$$f^{(n)}(x) = \frac{P(x)}{Q(x)} \cdot e^{-\frac{1}{x}}$$

where P, Q are polynomials. So the function value (of an infinitely often differentiable function) must not equate with its Taylor series value.

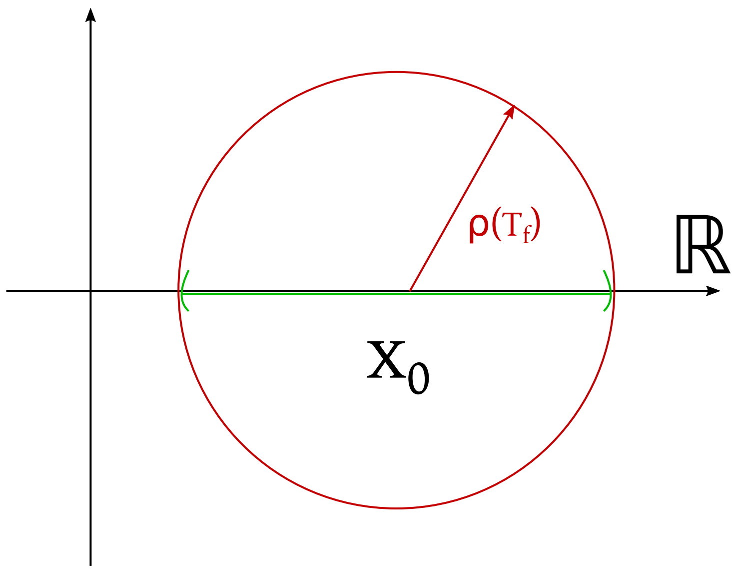

Figure 33: Taylor series

Proof. Proof by complete induction over n.

Case $n = 0$ immediate with $P = Q = 1$.

Case $n \mapsto n + 1$

$$
\begin{aligned}
f^{(n+1)}(x) &= \Bigg[\underbrace{\frac{P(x)}{Q(x)} \cdot e^{-\frac{1}{x}}}_{f^{(n)}(x) \text{ by induction hypothesis}} \Bigg]' \\
&= \frac{P' \cdot Q - Q' \cdot P}{Q^2} \cdot e^{-\frac{1}{x}} + \frac{P}{Q} \cdot \frac{1}{x^2} \cdot e^{-\frac{1}{x}} \\
&= \frac{(P'Q - Q'P)x^2 + PQ}{Q^2x^2} \cdot e^{-\frac{1}{x}}
\end{aligned}
$$

It holds that $\lim_{x \to 0_+} \frac{P(x)}{Q(x)} \cdot e^{-\frac{1}{x}} = 0$. Immediately, $\lim_{x \to 0^-} f^{(n)}(x) = 0$, hence $f^{(n)}(0) = 0 \forall n \in \mathbb{N}$. f is arbitrarily often continuously differentiable on $\mathbb{R}$. Thus,

$$
T_f(x; 0) = \sum_{k=0}^{\infty} \frac{\overbrace{f^{(k)}(0)}^{=0}}{k!} x^k = 0
$$

but $f(x) \neq 0$ on $\mathbb{R}$. Thus, it holds that $f \neq T_f(x;0)$. But it holds that $R_f = f - T_f(x;0) = f$.

$$\left|R_f(x)\right| \leq c_n \left|x\right|^n \qquad \forall n \in \mathbb{N}$$

$\square$

Theorem 5.7. *Let $f(x) = \sum_{k=0}^{\infty} a_k(x-x_0)^k$ be an analytical[2] function with convergence radius $\rho(f) > 0$. Then f is infinitely often continuously differentiable on $I := (x_0 - \rho(f), x_0 + \rho(f))$ and it holds that $a_k = \frac{f^{(k)}(x_0)}{k!}$, hence the given power series is the Taylor series of the function.*

Proof. See Analysis 1 lecture notes, chapter 8, theorem 1 by G. Lettl.

f is differentiable on $I = (x_0 - \rho(f), x_0 + \rho(f))$ and it holds that $f'(x) = \sum_{k=0}^{\infty} k a_k (x - x_0)^{k-1}$. Thus, f' is also analytical and the power series of f' converges on $K(x_0) \implies \rho(f') \geq \rho(f)$ (if you consider the Cauchy-Hadamard Theorem, then $\rho(f') = \rho(f)$).

Induction: $f^{(n)}(x)$ is analytical on I and it holds that

$$f^{(n)}(x) = \sum_{k=n}^{\infty} k \cdot (k-1) \dots (k-n+1) \cdot a_k \cdot (x - x_0)^{k-n}$$

We insert: $x = x_0$

$$f^{(n)}(x_0) = n \cdot (n-1) \dots 1 \cdot a_n \implies a_n = \frac{f^{(n)}(x_0)}{n!}$$

$\square$

Revision: Expansion on a different point (ξ_0 instead of x_0):

$$f(z) = \sum_{k=0}^{\infty} a_k (z - x_0)^k = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (z - x_0)^k$$

with $a_k = \frac{f^{(k)}(x_0)}{k!}$. $f(z) = \sum_{k=0}^{\infty} \tilde{a}_k (z - \xi_0)^k$ with $\tilde{a}_k = \frac{f^{(k)}(\xi_0)}{k!}$. Compare with Figure 34.

6 Multidimensional differential calculus

Let V, W be vector space over $\mathbb{K}$ ($\mathbb{R}, \mathbb{C}$).

$$\underbrace{\mathcal{L}(V, W)}_{\mathrm{Hom}(V,W)} = \{\varphi : V \to W : \varphi \text{ is linear}\}$$

[2]Reminder: A function is analytical if it is locally given by a convergent power series.

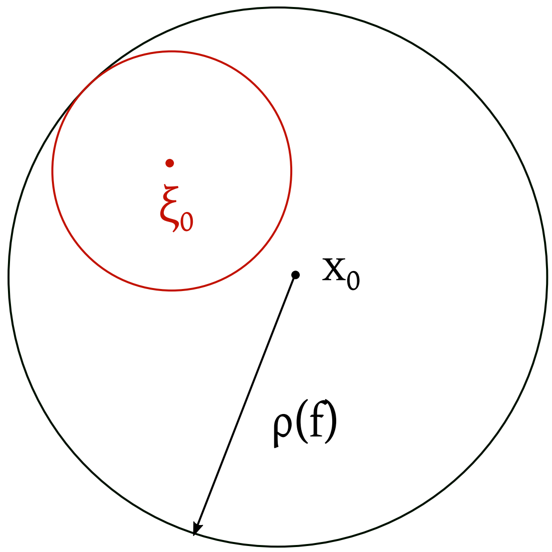

Figure 34: Expansion on a different point

Hom(V, W) has vector space properties. $\varphi, \psi \in \mathcal{L}(V, W), \lambda, \mu \in \mathbb{K}$. Then it holds that $\lambda\varphi + \mu\psi \in \mathcal{L}(V, W)$. In general, it is feasible that to define a norm on $\mathcal{L}(V, W)$. Hence, $\|\cdot\| : \mathcal{L}(V, W) \to [0, \infty)$ with

1. $\|\varphi\| = 0 \iff \varphi = 0$ (zero mapping)
2. $\forall \lambda \in \mathbb{K}, \varphi \in \mathcal{L}(V, W)$ it holds that $\|\lambda\varphi\| = |\lambda| \cdot \|\varphi\|$.
3. $\forall \varphi, \psi \in \mathcal{L}(V, W)$ it holds that $\|\varphi + \psi\| \leq \|\varphi\| + \|\psi\|$.

6.1 Frobenius and matrix norm

Example 6.1. *Let $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n) \cong \mathbb{R}^{n\times n}$. (identify linear maps with its matrix representation in regards of the canonical basis)*

$$A \in \mathbb{R}^{n\times m} \qquad A = (a_{ij})_{\substack{i=1,\dots,n \\ j=1,\dots,m}}$$

$$\|A\|_F = \left(\sum_{i=1}^{n}\sum_{j=1}^{m} |a_{ij}|^2\right)^{\frac{1}{2}} \qquad \text{"Forbenius norm"}$$

It basically works by appending the next column to the previous one. Hence, this gives a column vector. We square every entries, sum it up and take its square root (a common norm procedure). A norm on $\mathbb{R}^{n\times m}$ (a matrix) is called matrix norm *($\mathbb{C}^{n\times m}$).*

6.2 Operator norm and bounded linear operators

Definition 6.1. *Let V, W be normed vector spaces over $\mathbb{K}$. A linear map $\varphi : V \to W$ is called* bounded *if $\exists m \geq 0 : \left\|\varphi(x)\right\|_W < m \cdot \|x\|_V$ (we call this the* boundedness criterion*) such that $\left\|\varphi(x)\right\|_W \leq m \cdot \|x\|_V$ for all $x \in V$.*

The set $\mathcal{L}_b(V, W) = \{\varphi : V \to W : \varphi \text{ is linear and bounded}\}$ is a subvectorspace of $\mathcal{L}(V, W)$. We let

$$\left\|\varphi\right\| = \inf\left\{m \geq 0 : \left\|\varphi(x)\right\|_W \leq m \cdot \|x\|_V \ \forall x \in V\right\}$$

and call $\left\|\varphi\right\|$ the operator norm *on φ in regards of $\|\cdot\|_V$ and $\|\cdot\|_W$.*

Regarding the subvector space property:

Let $\varphi, \psi \in \mathcal{L}_b(V, W), \lambda, \mu \in \mathbb{K}$. Show that $\lambda\varphi + \mu\psi \in \mathcal{L}_b(V, W)$.

$$\begin{aligned}\left\|(\lambda\varphi + \mu\psi)(x)\right\|_W &= \left\|\lambda \cdot \varphi(x) + \mu \cdot \psi(x)\right\|_W \\ &\leq |\lambda| \left\|\varphi(x)\right\|_W + \left|\mu\right| \left\|\varphi(x)\right\|_W \\ &\underbrace{\leq}_{\text{because } \varphi,\psi \text{ are bounded}} |\lambda|\, m \,\|x\|_V + \left|\mu\right| m' \,\|x\|_V \\ &= \underbrace{\left(|\lambda|\, m + \left|\mu\right| m'\right)}_{=\tilde{m} \geq 0} \|x\|_V\end{aligned}$$

hence $\lambda\varphi + \mu\psi \in \mathcal{L}_b(V, W)$. $\mathcal{L}_b(V, W) \neq \emptyset$.

Lemma 6.1. *Let V, W be normed vector spaces. Then it holds for any $\varphi \in \mathcal{L}_b(V, W)$*

1. *$\left\|\varphi(x)\right\|_W \leq \left\|\varphi\right\| \cdot \|x\|_V \ \forall x \in V$. Hence, $m = \left\|\varphi\right\|$ satisfies the boundedness criterion, hence informally* inf *equals* min *in Definition 6.1.*

2. $$\left\|\varphi\right\| = \sup\left\{\frac{\left\|\varphi(x)\right\|_W}{\|x\|_V} : x \in V \setminus \{0\}\right\} = \sup\left\{\left\|\varphi(x)\right\|_W : x \in V \text{ with } \|x\|_V = 1\right\}$$

3. *$\|\cdot\|$ is a norm on $\mathcal{L}_b(V, W)$.*

Proof. 1. Let $m_n \geq 0$ with m_n satisfies the boundedness criterion, hence

$$\left\|\varphi(x)\right\|_W \leq m_n \cdot \|x\|_V \ \forall x \in V$$

and $m_n \to \left\|\varphi\right\|$. The inequality retains in the limit. Thus, $\left\|\varphi(x)\right\|_W \leq \left\|\varphi\right\| \cdot \|x\|_V$.

2. Let $\tilde{m} = \sup\left\{\frac{\|\varphi(x)\|_W}{\|x\|_V} : x \neq 0\right\}$. Hence, $\frac{\|\varphi(x)\|_W}{\|x\|_V} \leq \tilde{m} \forall x \in V$, because $\tilde{m}$ is an upper bound. So, $\left\|\varphi(x)\right\|_W \leq \tilde{m}\,\|x\|_V$. Thus $\tilde{m}$ satisfies the boundedness criterion and $\left\|\varphi\right\| \leq \tilde{m}$.

   On the opposite: Let m such that the boundedness criterion is satisfied $\implies \left\|\varphi(x)\right\|_W \leq m \cdot \|x\|_V \;\forall x \in V, x \neq 0$ and accordingly, $\frac{\|\varphi(x)\|}{\|x\|_V} \leq m$. Hence, m is upper bound of $\left\{\frac{\|\varphi(x)\|}{\|x\|_V} : X \neq 0\right\}$, hence $m \geq \tilde{m} = \sup\{\cdot\}$. Hence, $m \geq \tilde{m} = \sup\left\{\frac{\|\varphi(x)\|_W}{\|x\|_V} : x \neq 0\right\}$. The statement above also holds for the infimum of $m - s$, hence $\left\|\varphi\right\| \geq \tilde{m}$, hence $\left\|\varphi\right\| = \tilde{m} = \sup\left\{\frac{\|\varphi(x)\|_W}{\|x\|_V} : x \neq 0\right\}$. Because $\{x \in V : \|x\| = 1\} \subseteq \{x \in V : x \neq 0\}$ it holds that $\sup \left\|\varphi(x)\right\|_W : \|x\| = 1 = \sup\left\{\frac{\|\varphi(x)\|_W}{\|x\|_V} : \|x\|_V = 1\right\} \leq \sup\left\{\frac{\|\varphi(x)\|_W}{\|x\|_V} : x \neq 0\right\} = \left\|\varphi\right\|$.

   On the opposite: Let $x \neq 0$. Then $\tilde{x} = \frac{x}{\|x\|_V}$ defines a *unit vector*.

$$\|\tilde{x}\|_V = \left\|\frac{x}{\|x\|_V}\right\| = \frac{1}{\|x\|_V} \cdot \|x\|_V = 1$$

   and it holds that

$$\frac{\left\|\varphi(x)\right\|_W}{\|x\|_V} = \frac{1}{\|x\|_V}\left\|\varphi(x)\right\|_W = \left\|\frac{1}{\|x\|_V}\varphi(x)\right\|_W \underbrace{=}_{\varphi \text{ is linear}} = \left\|\varphi(\frac{x}{\|x\|_V})\right\|_W = \left\|\varphi(\tilde{x})\right\|$$

$$\implies \forall x \neq 0 : \frac{\left\|\varphi(x)\right\|_W}{\|x\|_V} = \left\|\varphi(\tilde{x})\right\| \leq \sup\left\{\left\|\varphi(z)\right\|_W : \|z\|_V = 1\right\}$$
$$\implies \sup\left\{\frac{\left\|\varphi(x)\right\|_W}{\|x\|_V} : x \neq 0\right\} \leq \sup\left\{\left\|\varphi(z)\right\|_W : \|z\|_V = 1\right\}$$

3. Show that $\left\|\varphi\right\|$ is a norm.

$$\left\|\varphi\right\| = 0 \iff \forall x \in V : \left\|\varphi(x)\right\|_W \leq 0 \cdot \|x\|_W$$

   hence $\varphi(x) = 0 \forall x \in V$ and accordingly, $\varphi = 0$ in $\mathcal{L}(V, W)$.

$$\left\|\lambda\varphi\right\| = \sup\left\{\left\|\lambda\varphi(x)\right\|_W : \|x\|_V = 1\right\} = \sup\left\{|\lambda|\left\|\varphi(x)\right\|_W : \|x\|_V = 1\right\}$$
$$= |\lambda| \sup\left\{\left\|\varphi(x)\right\|_W : \|x\|_V = 1\right\} = |\lambda|\left\|\varphi\right\|$$

   Triangle inequality: Let $\varphi, \psi \in \mathcal{L}_b(V, W)$.

$$\left\|\varphi(x) + \psi(x)\right\|_W \underbrace{\leq}_{\text{triangle inequality in } W} \left\|\varphi(x)\right\|_W + \left\|\psi(x)\right\|_W$$
$$\leq \left\|\varphi\right\| \cdot \|x\|_V + \left\|\psi\right\| \cdot \|x\|_V = (\left\|\varphi\right\| + \left\|\psi\right\|) \cdot \|x\|$$

By (1.), $\|\varphi\| + \|\psi\|$ satisfies the boundedness criterion for the linear map $\varphi + \psi$. Hence, $\|\varphi + \psi\| \leq \|\varphi\| + \|\psi\|$.

□

Remark 6.1. • *$\|A\|_F$ is no operator norm on $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$.*

• *Boundedness of linear mappings is required to define $\|\varphi\|$.*

We consider special case $V = \mathbb{R}^m, W = \mathbb{R}^n$.

$$\|\cdot\|_V = \|\cdot\|_\infty \qquad \|\cdot\|_W = \|\cdot\|_\infty$$

Let $A \in \mathbb{R}^{n\times m}$ ($\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$). Then it holds that

$$\begin{aligned}
\|Ax\|_\infty &= \max\left\{|(Ax)_i| : i = 1, \dots, n\right\} \\
&= \max\left\{\left|\sum_{j=1}^m a_{ij}x_j\right| : i = 1, \dots, n\right\} \\
&\leq \max\left\{\sum_{j=1}^m \left|a_{ij}\right| \cdot \underbrace{\left|x_j\right|}_{\leq\|x\|_\infty} : i = 1, \dots, n\right\} \\
&\leq \max\left\{\|x\|_\infty \cdot \sum_{j=1}^m \left|a_{ij}\right| : i = 1, \dots, n\right\} \\
&= \underbrace{\max\left\{\sum_{j=1}^m \left|a_{ij}\right| : i = 1, \dots, n\right\}}_{=m} \cdot \|x\|_\infty \\
&= m \cdot \|x\|_\infty
\end{aligned}$$

Hence the boundedness criterion is satisfied. A is bounded in regards of $\|\cdot\|_\infty$ in the preimage and image space. By the norm equivalence theorem, it follows that A is bounded in regards of arbitrary norms on $\mathbb{R}^m$, and accordingly $\mathbb{R}^n$.

This lecture took place on 2018/05/17.

Further remarks:

Remark 6.2. *A linear map $A : \mathbb{R}^m \to \mathbb{R}$ is always bounded. Thus,*

$$\|Ax_1 - Ax_2\|_{\mathbb{R}^n} = \|A(x_1 - x_2)\|_{\mathbb{R}^n} \leq \|A\| \, \|x_1 - x_2\|_{\mathbb{R}^m}$$

So every linear map $A : \mathbb{R}^m \to \mathbb{R}^n$ is Lipschitz continuous *with Lipschitz constant $\|A\|$.*

The considerations above hold for arbitrary finite-dimensional normed vector spaces V and W (over $\mathbb{K} = \mathbb{R}, \mathbb{C}$).

Lemma 6.2. *Let X, Y and Z be normed vector spaces. Let $A : X \to Y$ be linear and bounded. Let $B : Y \to Z$ be linear and bounded. Then*

$$B \cdot A : X \to Z$$

is also bounded and it holds that

$$\|B \cdot A\| \leq \|B\| \cdot \|A\|$$

Proof. Let $x \in X$ be arbitrary and $BAx = B(Ax) \in Z$ and

$$\|BAx\|_Z = \|B(Ax)\|_Z \overbrace{\leq}_{B \text{ is bounded}} \|B\| \, \|Ax\|_Y \underbrace{\leq}_{A \text{ is bounded}} \|B\| \cdot \|A\| \, \|x\|_X$$

Hence $m = \|B\| \, \|A\|$ satisfies the boundedness criterion for the linear map $B \cdot A : X \to Z$. Because $\|B \cdot A\|$ is the smallest constant for which the boundedness criterion holds, it follows that $\|BA\| \leq \|B\| \, \|A\|$. □

6.3 Landau notation

Definition 6.2 (Landau $\mathbb{O}$ symbols)**.** *Let $h, g : D \subseteq \mathbb{R}^n \to \mathbb{R}$, D is open, $a \in D$.*

1. *We denote $h = \mathcal{O}(g)$ in a (German pronunciation: h ist groß̈ $\mathcal{O}$ von g) iff $\exists U \subseteq D$ environment of a in D and $\exists r : U \to \mathbb{R}$ with r bounded such that $h(x) = r(x) \cdot g(x) \forall x \in U$. Thus,*

$$\left|\frac{h(x)}{g(x)}\right| = |r(x)| \leq M \qquad \forall x \in U$$

   ($g(x) = 0$ iff $h(x) = 0$)

2. *We denote $h = o(g)$ in a (German pronunciation: h ist klein o von g) iff $\exists U \subseteq D$, with U being the environment of a, and $r : U \to \mathbb{R}$ such that $\lim_{x \to a} r(x) = 0$ and $h(x) = r(x) \cdot g(x) \forall x \in U$. In that sense,*

$$\lim_{x \to a} \frac{h(x)}{g(x)} = 0$$

   Most often,

$$\mathcal{O}(\underbrace{\|x - x_0\|^n}_{g(x)})$$

   is used and $a = x_0$.

6.4 Multidimensional derivative of a function

Definition 6.3 (Definition of the derivative of a function)**.**

$$f : D \subseteq \mathbb{R}^m \to \mathbb{R}^n \qquad D \textit{ is open}, x_0 \in D$$

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix} \in \mathbb{R}^n$$

$$f_i(x) = f_i(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}) = f_i(x_1, x_2, \dots, x_m) \in \mathbb{R}$$

Remark 6.3. *First trial to define of a the derivative:*

$$f'(x_0) := \lim_{x \to x_0} \frac{\overbrace{f(x) - f(x_0)}^{\in \mathbb{R}^n}}{\underbrace{x - x_0}_{\in \mathbb{R}m}}$$

Does not work because of incompatibility of dimensions (and we cannot divide vectors).

Remark 6.4. *We use Taylor's Theorem to characterize $f'(x_0)$. A Taylor polynomial of 1st degree is given by*

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + R_f^2(x; x_0)$$

and $R_f^2(x; x_0) = r(x)(x - x_0)$ with $\lim_{x \to x_0} r(x) = 0$. Hence,

$$f(x) - f(x_0) - f'(x_0)(x - x_0) = \mathcal{O}(x - x_0)$$

or we insert the absolute operators:

$$\left| f(x) - f(x_0) - f'(x_0)(x - x_0) \right| = |r(x)| \cdot |x - x_0| = o(|x - x_0|)$$

where $(f(x) - f(x_0)) \in \mathbb{R}^n$ for the multidimensional case and $(x - x_0) \in \mathbb{R}^m$ for the multidimensional case. Thus,

$$= \mathcal{O}(|x - x_0|)$$

Definition 6.4. *Let $f : D \subseteq \mathbb{R}^m \to \mathbb{R}^n$, D is open and $x_0 \in D$. We say, "f is differentiable in x_0" (specifically, "Frechét differentiable") if there exists $A \in \mathbb{R}^{n \times m}$ such that*

$$\left\| f(x) - f(x_0) - A(x - x_0) \right\|_{\mathbb{R}^n} = o(\|x - x_0\|_{\mathbb{R}^m})$$

We call this condition differentiability condition*. A is a linear approximation of f in x_0. Because of norm equivalence on $\mathbb{R}^n$, and accordingly $\mathbb{R}^m$, it is irrelevant which norm $\|\cdot\|_{\mathbb{R}^n}$ and $\|\cdot\|_{\mathbb{R}^m}$ is chosen.*

Lemma 6.3. *Let f be, as in Definition 6.4, differentiable in x_0. Then the linear approximation A, by the differentiability condition, is uniquely determined.*

Proof. Assume $A, B \in \mathbb{R}^{n\times m}$ satisfy the differentiability condition. Let $r > 0$ such that $K_r(x_0) \subseteq D$ (feasible because D is open and $x_0 \in D$). Furthermore let $v \in \mathbb{R}^m$ and $\|v\| < r$, hence $x = x_0 + v \in K_r(x_0) \subset D$ and $v = x - x_0$.

$$\|(A-B)v\| = \|Av - Bv\| = \|A(x-x_0) - B(x-x_0)\|$$

$$= \left\| f(x) - f(x_0) - B(x-x_0) - (f(x) - f(x_0) - A(x_0 - x_0))\right\|$$

$$\leq \left\| f(x) - f(x_0) - B(x-x_0)\right\| + \left\| f(x) - f(x_0) - A(x-x_0)\right\|$$

by the differentiability criterion

$$r(x) \cdot \|x - x_0\| + \tilde{r}(x)\,\|x - x_0\|$$

with $\lim_{x\to x_0} r(x) = \lim_{x\to x_0} \tilde{r}(x) = 0$. Hence, for $\hat{r}(x) = r(x) + \tilde{r}(x)$ it holds that

$$\|(A-B)v\| \leq \hat{r}(x)\,\|x-x_0\| = \hat{r}(x)\,\|v\| = \mathcal{O}(\|v\|) \text{ in } x_0$$

(Thus $\lim_{x\to x_0} \hat{r}(x) = 0$)

Show: $(A-B)w = 0 \forall w \in \mathbb{R}^m$. Assume $\exists w \in \mathbb{R}^m, w \neq 0$ with $(A-B)w \neq \varnothing$. For $|\alpha| < \frac{r}{\|w\|}$ (with r as radius of the sphere) it holds that

$$\|\alpha w\| = |\alpha|\,\|w\| < \frac{r}{\|w\|} \cdot \|w\| = r$$

Let $v = \alpha w$. Then it holds that

$$\|(A-B)v\| = |\alpha|\,\|(A-B)w\| \leq \hat{r}(x)\,\|\alpha w\| = |\alpha|\,\hat{r}(x) \cdot \|w\|$$

$$\implies \|(A-B)w\| \leq \underbrace{\hat{r}(x)}_{\to 0 \text{ for } x\to x_0} \cdot \underbrace{\|w\|}_{\text{constant}}$$

$$\implies (A-B)w = 0$$

This contradicts with our assumption.

Therefore, $Aw = Bw \forall w \in \mathbb{R}^m$, hence $A = B$. □

6.5 Frechét derivative

Definition 6.5 (Part 2 of Definition 6.4)**.** *If f is differentiable in x_0, then we call the uniquely determined linear map A the "Frechét derivative" of f in x_0 and denote $A = Df(x_0)$. An alternative notations are $f'(x_0)$ and $D_{x_0}f$.*

Lemma 6.4. *Let $f : D \to \mathbb{R}^n$, $D \subseteq \mathbb{R}^m$ open, $x_0 \in D$. If f is differentiable in x_0, then f is also continuous in x_0.*

Proof. Let $x \in D$. Then it holds that

$$\begin{aligned}
\left\|f(x) - f(x_0)\right\|_{\mathbb{R}^n} &= \left\|f(x) - f(x_0) - Df(x_0)\cdot(x - x_0) + Df(x_0)(x - x_0)\right\| \\
&\leq \left\|f(x) - f(x_0) - Df(x_0)(x - x_0)\right\| + \left\|Df(x_0)(x - x_0)\right\| \\
&\leq r(x)\cdot \underbrace{\|x - x_0\|}_{\to 0 \text{ for } x\to x_0} + \overbrace{\left\|\underbrace{Df(x_0)}_{\in\mathbb{R}^{n\times m}}\right\|}^{\text{constant}}\cdot \underbrace{\|x - x_0\|}_{\to 0 \text{ for } x\to x_0} \\
&\to 0 \text{ for } x \to x_0
\end{aligned}$$

Hence f is continuous in x_0. □

Lemma 6.5. *Let $f, g : D \subseteq \mathbb{R}^m \to \mathbb{R}^n$. Let f and g be differentiable in $x_0 \in D$. Let $\lambda \in \mathbb{R}$. Then it holds that*

1. *$f + g$ is differentiable in x_0 with*

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

2. *λf is differentiable in x_0 and*

$$D(\lambda f)(x_0) = \lambda Df(x_0)$$

Thus differentiability is a linear operation on the vector space's appropriate differentiable functions.

Proof. Let $F := \left\|(f + g)(x) - (f + g)(x_0) - \underbrace{[Df(x_0) + Dg(x_0)]}_{D(f+g)(x_0)}(x - x_0)]\right\|$. Show that $F = o(\|x - x_0\|)$.

$$\begin{aligned}
F &\leq \left\|f(x) - f(x_0) - Df(x_0)(x - x_0)\right\| + \left\|g(x) - g(x_0) - Dg(x_0)(x - x_0)\right\| \\
&= o(\|x - x_0\|) + o(\|x - x_0\|) \\
&= o(\|x - x_0\|)
\end{aligned}$$

For λf it holds analogously. □

Lemma 6.6. *Let $C : D \to \mathbb{R}^n$. $c(x) = k \in \mathbb{R}^n$ is constant. Then c be differentiable in every point $x_0 \in D$ and it holds that $DC(x_0) = 0 \in \mathbb{R}^{n\times m}$. Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be linear. Then A is differentiable in every point $x_0 \in \mathbb{R}^m$ and it holds that $DA(x_0) = A$.*

Let $f(x) = k + Ax : \mathbb{R}^m \to \mathbb{R}^n$ be linear affine, then f is differentiable in every point $x_0 \in \mathbb{R}^m$ with $Df(x_0) = A$.

Proof.

$$\|c(x) - c(x_0) - 0 \cdot (x - x_0)\|$$

where 0 denotes the zero-matrix.

$$= \|k - k\| = \|0\| = 0 = o(\|x - x_0\|)$$

Hence 0 satisfies the differentiability condition for c.

$$\implies 0 = Dc(x_0)$$

in the linear case

$$\|Ax - Ax_0 - A(x - x_0)\| = \|Ax - Ax_0 - Ax + Ax_0\| = 0 = o(x - x_0)$$

hence $DA(x_0) = A$. Affine: use Lemma 6.5.

$$D(k + A)(x_0) = \underbrace{Dk(x_0)}_{=0} + \underbrace{DA(x_0)}_{=A} = A$$

This is analogous to the one-dimensional case:

$$(k + ax)' = a$$

□

6.6 Chain rule

Theorem 6.1 (Chain rule in multiple dimensions)**.** *Let $D \subseteq \mathbb{R}^l$ be open. Let $E \subseteq \mathbb{R}^m$ be open. Let $f : D \to \mathbb{R}^m$ such that $f(D) \subseteq E$ and $g : E \to \mathbb{R}^n$. Compare with Figure 35.*

Let f in x_0 be differentiable and g in $y_0 = f(x_0)$ is differentiable. Then also $g \circ f : D \to \mathbb{R}^n$ is differentiable in x_0 and it holds that

$$\underbrace{D(g \circ f)(x_0)}_{\in \mathbb{R}^{n \times l}} = \underbrace{Dg(f(x_0))}_{\in \mathbb{R}^{n \times m}} \underbrace{Df(x_0)}_{\in \mathbb{R}^{m \times l}}$$

(The dimensions match.)

Proof. Let $\varepsilon > 0$ be arbitrary, but $\varepsilon < 2$. Show that

$$\frac{1}{\|x - x_0\|} \left\| g(f(x)) - g(f(x_0)) - Dg(y_0) \cdot Df(x_0)(x - x_0) \right\| < \varepsilon$$

for sufficiently small $\|x - x_0\|$ $(x \neq x_0)$.

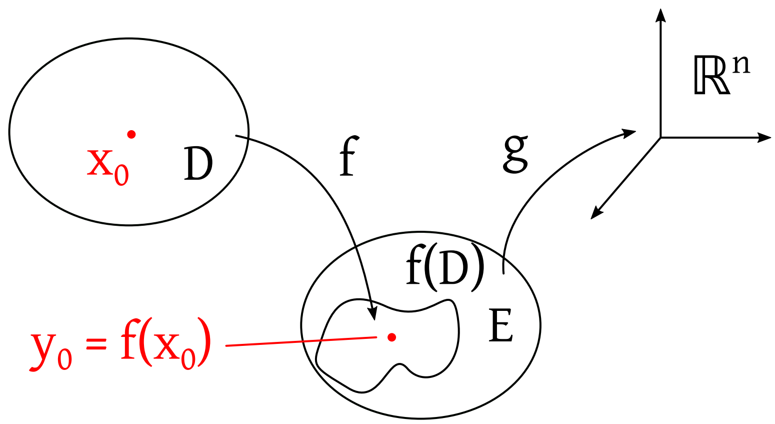

Figure 35: Chain rule in multiple dimensions

$$
\begin{aligned}
&\frac{1}{\|x-x_0\|}\left\|g(f(x))-g(f(x_0))-Dg(y_0)\cdot Df(x_0)(x-x_0)\right\| \\
&=\frac{1}{\|x-x_0\|}\|g(f(x))-g(f(x_0))-Dg(f(x_0))\cdot(f(x)-f(x_0)) \\
&+Dg(f(x_0))(f(x)-f(x_0))-Dg(f(x_0))Df(x_0)(x-x_0)\|
\end{aligned}
$$

recognize that we have a common factor $Dg(f(x_0))$

$$
\begin{aligned}
&\leq \frac{1}{\|x-x_0\|}\left\|g(f(x))-g(f(x_0))-Dg(f(x_0))(f(x)-f(x_0))\right\| \\
&+\frac{1}{\|x-x_0\|}\left\|Dg(y_0)\right\|\left\|f(x)-f(x_0)-Df(x_0)(x-x_0)\right\| \\
&=: (I)+(II)
\end{aligned}
$$

Choose $\delta_1 > 0$ such that $\|x-x_0\| < \delta_1 \implies$

$$\frac{1}{\|x-x_0\|}\cdot\left\|f(x)-f(x_0)-Df(x_0)(x-x_0)\right\| < \frac{\varepsilon}{2}\frac{1}{\left\|Dg(y_0)\right\|+1}$$

is feasible, because f is differentiable in x_0.

$$\frac{\varepsilon}{2}\cdot\frac{1}{\left\|Dg(y_0)\right\|+1} < \frac{2}{2} = 1$$

so it also holds that

$$\left\| f(x) - f(x_0) - Df(x_0)(x - x_0) \right\| < 1 \cdot \|x - x_0\|$$

By the reverse triange inequality,

$$\begin{aligned}
\left\| f(x) - f(x_0) - Df(x_0)(x - x_0) \right\| &\geq \left\| f(x) - f(x_0) \right\| - \left\| Df(x_0)(x - x_0) \right\| \\
&\geq \left\| f(x) - f(x_0) \right\| - \left\| Df(x_0) \right\| \cdot \|x - x_0\| \\
\implies \frac{\left\| f(x) - f(x_0) \right\|}{\|x - x_0\|} &\leq \left\| Df(x_0) \right\| + 1
\end{aligned}$$

This lecture took place on 2018/05/24.

$$\left\| f(x) - f(x_0) \right\| - \left\| Df(x_0) \right\| \|x - x_0\| < 1 \, \|x - x_0\|$$

hence, for $x \neq x_0$

$$\frac{\left\| f(x) - f(x_0) \right\|}{\|x - x_0\|} < \left\| Df(x_0) \right\| + 1$$

g is differentiable in $y_0 = f(x_0)$. Hence, we can choose $\delta_g > 0$ such that $\forall y \in E$ with $\left\| y - y_0 \right\| < \delta_g$ it holds that

$$\left\| g(y) - g(y_0) - Dg(y_0) \cdot (y - y_0) \right\| < \frac{\varepsilon}{2\left(\left\| Df(x_0) + 1 \right\|\right)} \left\| y - y_0 \right\|$$

Because f is continuous in x_0, there exists $\delta_2 > 0$ such that $x \in D$ and $\|x - x_0\| < \delta_2 \implies \left\| f(x) - f(x_0) \right\| < \delta_g$. Now let $\delta = \min(\delta_1, \delta_2) > 0$. Then it holds that

$$I = \frac{1}{\|x - x_0\|} \left\| g(f(x)) - g(f(x_0)) - Dg(f(x_0)) \cdot (f(x) - f(x_0)) \right\|$$

Let $y = f(x), y_0 = f(x_0)$. Because $\left\| f(x) - f(x_0) \right\| < \delta_g$ gives $\|x - x_0\| < \delta_2$

$$\implies I < \frac{\varepsilon}{2\left(\left\| Df(x_0) \right\| + 1\right)} \underbrace{\frac{\left\| f(x) - f(x_0) \right\|}{\|x - x_0\|}}_{< \left\| Df(x_0) \right\| + 1} < \frac{\varepsilon}{2}$$

$$\begin{aligned}
II = \left\| Dg(y_0) \right\| \frac{1}{\|x - x_0\|} \left\| f(x) - f(x_0) - Df(x_0)(x - x_0) \right\| \\
< \left\| Dg(y_0) \right\| \cdot \frac{\varepsilon}{2} \cdot \frac{1}{\left\| Dg(y_0) \right\| + 1} < \frac{\varepsilon}{2}
\end{aligned}$$

hence, $I + II < \varepsilon$ for $\|x - x_0\| < \delta$. □

6.7 Differentiability on D

Definition 6.6. *Let $f : D \subseteq \mathbb{R}^m \to \mathbb{R}^n$ be differentiable in every point $x \in D$. Then we are used to say "f is differentiable* on *D".*

In this case, we call the map

$$x \mapsto Df(x)$$

$$D \subseteq \mathbb{R}^m \to \mathbb{R}^{n\times m}$$

the mapping function of f (dt. Abbildungsfunktion). If this function is continuous (in terms of $\|\cdot\|_{\mathbb{R}}$ or $\|\cdot\|_{\mathbb{R}^{n\times m}}$ ... operator norm), then f is called continuously differentiable on D.

Remark 6.5. *To define the differentiability of f, we require x_0 to be an accumulation point of D. So x_0 might also be a point on the boundary of D.*

6.7.1 Determination of $Df(x_0) \in \mathbb{R}^{n\times m}$

Definition 6.7. *Let $f : D \subseteq \mathbb{R}^m \to \mathbb{R}^n$ be given. D is open, $x_0 \in D$, $v \in \mathbb{R}^m$ is arbitrary, but $v \neq 0$. We consider $t \mapsto f(x_0 + tv)$ defined on $(-\frac{r}{\|v\|}, \frac{r}{\|v\|})$ for $r > 0$ such that $K_r(x_0) \subseteq D$.*

$$\left(-\frac{r}{\|v\|}, \frac{r}{\|v\|}\right) \subseteq \mathbb{R} \to \mathbb{R}^n$$

Compare with Figure 36.

We define $df(x_0; v) = \lim_{t\to 0} \frac{1}{t}(f(x_0+tv) - f(x_0))$ if this limit exists. $df(x_0, v)$ is called directional derivative *of f in x_0 in direction v. It is also called* Gateaux derivative of f in x_0 in direction v.

Remark 6.6. *How does it go together?*

Derivative $Df(x_0)$ and $df(x_0; \cdot)$. Assumption: Let f be differentiable in x_0. We define $l_{x_0,v}(t) = x_0 + tv$ where tv is the linear part.

$$l_{x_0,v} : \left(-\frac{r}{\|v\|}, \frac{r}{\|v\|}\right) \to D$$

$l_{x_0,v}$ is linear affine from $\mathbb{R}$ to $\mathbb{R}^m$.

$$Dl_{x_0,v}(0) = V \in \mathbb{R}^{m\times 1}$$

with V as linear part of $l_{x_0,v}$.

$$l_{x_0,v}(0) = x_0$$

$$f(x_0 + tv) = f \circ l_{x_0,v}(t)$$

Therefore it holds that (chain rule)

$$D(f \circ l_{x_0,v})(0) = Df(l_{x_0,v}(0)) \cdot Dl_{x_0,v}(0) = Df(x_0) \cdot v$$

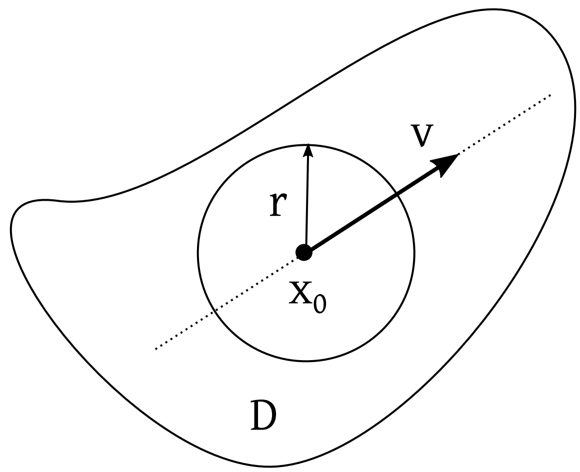

Figure 36: Setting in Definition 6.7. The line is given by $g = \{x_0 + tv : t \in \mathbb{R}\}$

On the other side, it holds that

$$\begin{aligned}
0 &= \underbrace{\lim_{t\to 0} \frac{1}{|t|} \left| f_{x_0,v}(t) - f_{x_0,v}(0) - Df_{x_0,v}(0)\cdot t \right|}_{=0} \\
&= \lim_{t\to 0} \frac{1}{|t|} \left| f(x_0 + tv) - f(x_0) - Df_{x_0,v}(0)\cdot t \right| \\
&= \lim_{t\to 0} \left| \frac{f(x_0+tv) - f(x_0)}{t} - Df_{x_0,v}(0) \right| = 0
\end{aligned}$$

therefore it holds that

$$Df_{x_0,v}(0) = \lim_{t\to 0} \frac{f(x_0+tv) - f(x_0)}{t} = df(x_0; v)$$

Lemma 6.7. *Let $f : D \subseteq \mathbb{R}^m \to \mathbb{R}^n$ in $x_0 \in D$ (Frechét) differentiable with derivative $Df(x_0)$. Then also the directional derivative $df(x_0; v)$ for every direction $v \in \mathbb{R}^m \setminus \{0\}$ and it holds that*

$$df(x_0; v) = Df(x_0) \cdot v$$

Remark 6.7. *$v \mapsto df(x_0; v)$ is linear. We can derive the structure of the derivative matrix. Let f as above. Let $\mathcal{B} = \{e_1, \dots, e_m\}$ be the canonical basis in $\mathbb{R}^m$. Then it holds* *that: $Df(x_0)\cdot e_j$ is the j-th column of $Df(x_0)$ for $j = 1,\dots,m$. On the other hand,*

$$Df(x_0)\cdot e_j = df(x_0; e_j) = \lim_{t\to 0}\frac{1}{t}\left[f(x_0+te_j) - f(x_0)\right]$$

$$= \begin{bmatrix} \lim_{t\to 0}\frac{1}{t}\left(f_1(x_0+te_j) - f_1(x_0)\right) \\ \vdots \\ \lim_{t\to 0}\frac{1}{t}\left(f_n(x_0+te_j) - f_n(x_0)\right) \end{bmatrix} \text{ for } f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix} \in \mathbb{R}^n$$

Remark 6.8 (Notation)**.** *Consider x instead of x_0*

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = [x_1,\dots,x_m]^t$$

Instead of $f(x) = f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}\right)$ we also write $f(x_1, x_2, \dots, x_m)$.

6.8 Partial derivative and Jacobi matrix

Definition 6.8. *Let $f : D \to \mathbb{R}^n$, $x \in D$ as above. f is differentiable in x. Then we let*

$$\frac{\partial f}{\partial x_j}(x) = df(x; e_j) = \lim_{t\to 0}\frac{1}{t}\left[f(x+te_j) - f(x)\right]$$
$$= \lim_{t\to 0}\frac{1}{t}\left[f(x_1,\dots,x_j+t,\dots,x_m) - f(x_1,\dots,x_j,\dots,x_m)\right]$$

and we call $\frac{\partial f}{\partial x_j}(x)$ the partial derivative *of f of variable x_j in point x.*

Notations for $\frac{\partial f}{\partial x_j}$:

$$f_{x_j} \qquad f_j \qquad \partial_j f$$

The second notation is ambiguous. We will prefer the last one.

Remark 6.9.

$$\frac{\partial f}{\partial x_j}(x_0) = \begin{bmatrix} df_1(x; e_j) \\ df_2(x; e_j) \\ \vdots \\ df_n(x; e_j) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_j}(x) \vdots \\ \frac{\partial f_n}{\partial x_j}(x) \end{bmatrix} \in \mathbb{R}^n$$

Because $\frac{\partial f}{\partial x_j}(x)$ is the j-th column of $Df(x_0)$, we get

$$(Df(x))_{i,j} = \frac{\partial f_i}{\partial x_j}(x) = \partial_j f_i(x)$$

We say, $Df(x) = (\partial_j f_i(x))_{\substack{i=1,\dots,n\\ j=1,\dots,m}}$ *is the* Jacobi matrix *of* f.

Remark 6.10. *The Jacobi matrix can exist even though the derivative does not exist. If the derivative exists, the Jacobi matrix exists for sure.*

Remark 6.11.

$$\frac{\partial f}{\partial x_j} = \lim_{t \to 0} \frac{1}{t} \left[f(x_1, \dots, x_j + t, \dots, x_m) - f(x_1, \dots, x_j, \dots, x_n) \right]$$

Thus, consider x_j *as derivation variable and all* x_k *for* $k \neq j$ *as constant parameters.*

Example 6.2. *Consider* $f : \mathbb{R}^3 \to \mathbb{R}^2$ *wit*

$$f(x_1, x_2, x_3) = \begin{bmatrix} x_1 x_3^2 + \sin(x_1 x_3) \\ \frac{x_2^2}{x_1^2 + 1} \end{bmatrix}$$

$$\partial_1 f(x_1, x_2, x_3) = \begin{bmatrix} 1 \cdot x_3^2 \\ -x_2^2 (x_1^2 + 1)^{-2} \cdot 2x_1 \end{bmatrix} = \begin{bmatrix} x_3^2 \\ -2 \frac{x_1 x_2^2}{(x_1^2+1)^2} \end{bmatrix}$$

$$\partial_2 f(x_1, x_2, x_3) = \begin{bmatrix} x_3 \cos(x_2 x_3) \\ \frac{2x_2}{x_1^2 + 1} \end{bmatrix}$$

$$\partial_3 f(x_1, x_2, x_3) = \begin{bmatrix} 2x_1 x_3 + x_2 \cos(x_2 x_3) \\ 0 \end{bmatrix}$$

Jacobi-Matrix

$$Df(x) = \left[\frac{\partial f_i}{\partial x_j} \right]_{\substack{i=1,\dots,2 \\ j=1,\dots,3}} = \begin{bmatrix} x_3^2 & x_3 \cos(x_2 x_3) & 2x_1 x_3 + x_2 \cos(x_2 x_3) \\ -\frac{2x_1 x_2^2}{(x_1^2+1)^2} & \frac{2x_2}{x_1^2+1} & 0 \end{bmatrix}$$

Remark 6.12. *Existence of partial derivatives of* f *does not suffice to ensure Frechét-differentiability.*

This lecture took place on 2018/05/29.

Usually, we always have to point out which norm is used to define differentiability. Of course, in $\mathbb{R}$ itself, all norms are equivalent.

Remark 6.13. *Let* $f : D \subseteq \mathbb{R}^m \to \mathbb{R}^n$ *be given. Let* $\|\cdot\|_{1,m}$ *and* $\|\cdot\|_{2,m}$ *be two equivalent norms on* $\mathbb{R}^m$ *(norm equivalence theorem) and* $\|\cdot\|_{1,n}$ *and* $\|\cdot\|_{2,n}$ *are equivalent norms on* $\mathbb{R}^n$.

Let f *in* $x_0 \in D$ *be differentiable in regards of* $\|\cdot\|_{1,m}$ *and* $\|\cdot\|_{1,n}$. *Then also* f *is differentiable in* x_0 *in regards of* $\|\cdot\|_{2,m}$ *and* $\|\cdot\|_{2,n}$.

Rationale: *Let* $c\|x\|_{2,m} \leq \|x\|_{1,m} \leq C\|x\|_{2,m}$ *and* $k\left\|y\right\|_{2,n} \leq \left\|y\right\|_{1,n} \leq K\left\|y\right\|_{2,n}$*, then*

$$\begin{aligned}
&\frac{\left\|f(x) - f(x_0) - A(x - x_0)\right\|_{2,n}}{\|x - x_0\|_{2,m}} \\
\leq\ &\frac{\frac{1}{k}\left\|f(x) - f(x_0) - A(x - x_0)\right\|_{1,n}}{\frac{1}{c}\|x - x_0\|_{1,m}} \\
=\ &\frac{c}{k}\underbrace{\frac{\left\|f(x) - f(x_0) - A(x - x_0)\right\|_{1,n}}{\|x - x_0\|_{1,m}}}_{\xrightarrow{x \to x_0} 0}
\end{aligned}$$

Let $f : D \subseteq \mathbb{R}^m \to \mathbb{R}^n$, $f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix}$*. Then* f *is differentiable in* $x_0 \in D \iff f_k : D \to \mathbb{R}$ *is differentiable for all* $k \in \{1, \dots, n\}$*.*

Rationale: *Let* f *be differentiable in* x_0*. Let* $A = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ *be the derivative of* f*. Let* a_k *be the rows of* A*.* f *is differentiable in* x_0*, so*

$$\frac{\left\|f(x) - f(x_0) - A(x - x_0)\right\|_\infty}{\|x - x_0\|} \xrightarrow{x \to x_0} 0$$

$$\iff \frac{\left|f_n(x) - f_k(x) - a_k(x - x_0)\right|}{\|x - x_0\|} \xrightarrow{x \to x_0} 0$$

where a_k *is a row vector and* $x - x_0$ *is a column vector and* $k \in \{1, \dots, n\}$*.*

$$\iff f_k \text{ is differentiable in } x_0$$

Example 6.3 (Counterexample)**.** *We define* $f : \mathbb{R}^2 \to \mathbb{R}$*.*

$$f(x_1, x_2) = \begin{cases} \frac{x_1^2 x_2}{x_1^2 + x_2^2} & \text{for } \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ 0 & \text{for } \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{cases}$$

Partial derivatives exist in every point $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

$$\partial_1 f(x_1, x_2) = \frac{2x_1 x_2 (x_1^2 + x_2^2) - x_1^2 x_2 2x_1}{(x_1^2 + x_2^2)^2} = \frac{2x_1 x_2^3}{(x_1^2 + x_2^2)^2}$$

$$\partial_2 f(x_1,x_2) = \frac{x_1^2(x_1^2+x_2^2) - x_1^2 x_2 \cdot 2x_2^2}{(x_1^2+x_2^2)^2} = \frac{x_1^2(x_1^2 x_2^2)}{(x_1^2+x_2^2)^2}$$

$$\partial_1 f(0,0) = \lim_{x_1\to 0} \frac{1}{x_1}[\underbrace{f(x_1,0)}_{=0} - \underbrace{f(0,0)}_{=0}] = 0$$

$$\partial_2 f(0,0) = \lim_{x_2\to 0} \frac{1}{x_2}[\underbrace{f(0,x_2)}_{=0} - \underbrace{f(0,0)}_{=0}] = 0$$

If f would be differentiable in $\begin{bmatrix}0\\0\end{bmatrix}$ then $Df(0) = [00]$, $df(0,v) = Df(0)\cdot v$. Choose $v = \begin{bmatrix}1\\1\end{bmatrix} \implies df(0;v) = [00]\begin{bmatrix}1\\1\end{bmatrix} = 0$. But!

$$\begin{aligned} df(0;v) &= \lim_{t\to 0}\frac{1}{t}[f(tv) - f(0)] = \lim_{t\to 0}\frac{1}{t}[f(t,t) - f(0,0)] \\ &= \lim_{t\to 0}\frac{1}{t}\left[\frac{t^2\cdot t}{t^2+t^2} - 0\right] = \lim_{t\to 0}\frac{t^3}{2t^3} = \frac{1}{2} \end{aligned}$$

6.9 Total differentiability, nabla operator and gradients

Remark 6.14. *Notation: $f : D \subseteq \mathbb{R}^m \to \mathbb{R}$. Often, we denote $df(x_0)$ ($\in \mathbb{R}^{1\times m}$, row vector) instead of $Df(x_0)$ and we call $df(x_0)$ the* total differential of f. *We let*

$$\nabla f(x_0) = df(x_0)^t = [Df(x_0)]^t = \begin{bmatrix}\partial_1 f(x_0)\\ \partial_2 f(x_0)\\ \vdots \\ \partial_n f(x_0)\end{bmatrix}$$

$\nabla f(x_0)$ *is called* gradient of f in x_0. *We call ∇ the* nabla operator. *It is also denoted* $\mathrm{grad}(f)$ *instead of ∇f. It holds that*

$$df(x_0;v) = df(x_0)\cdot v = \nabla f(x_0)^t \cdot v = \langle \nabla f(x_0), v\rangle_{\mathbb{R}^m}$$

Remark 6.15. *Let $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ continuously differentiable be given on D. Consider $\Gamma_s = \left\{x \in D \,\middle|\, f(x) = s\right\}$. Compare with Figure 37.*

Assume Γ_s is a graph of a family of curves (dt. Parametrisierte Kurve) $\gamma_s : I \to D$. Let I be an interval. $\Gamma_s = \left\{\gamma_s(t) \,\middle|\, t \in I\right\}$. We assume, that γ_s is regular, hence γ is differentiable and $\gamma_s'(t) = \begin{bmatrix}\gamma_1'(t)\\ \gamma_2'(t)\end{bmatrix} \neq \begin{bmatrix}0\\0\end{bmatrix} \forall t \in I$. $\gamma_s'(t)$ is tangential vector on Γ_s in point $x = \gamma_s(t)$.

It holds that

$$\langle \nabla f(\gamma_s(t)), \gamma_s'(t)\rangle = \langle \nabla f(x), v\rangle = 0$$

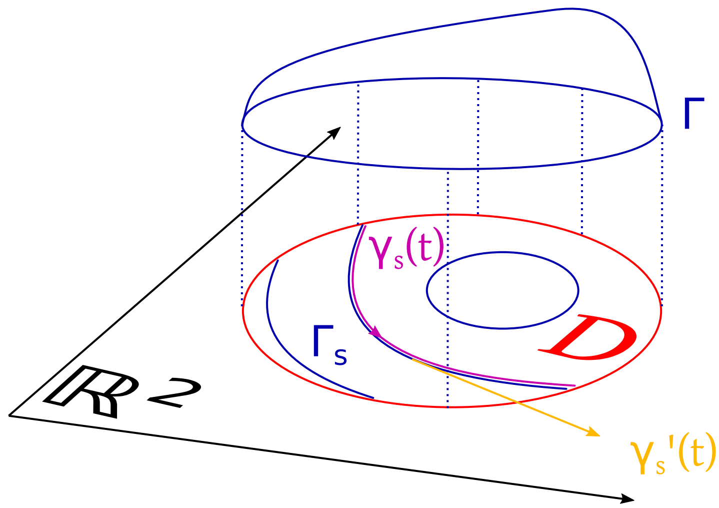

Figure 37: Γ_S is the niveau level of f

$$f(\underbrace{\gamma_s(t)}_{\in\Gamma_s}) = s \qquad \ldots \textit{constant}$$

$$\implies \frac{d}{dt}[f(\gamma_s(t))] = 0$$

where

$$\frac{d}{dt}f(\gamma_s(t)) = \underbrace{Df(x)}_{\nabla f(x)^t} \cdot \underbrace{D\gamma_s(x)}_{\gamma_s'(t)} = \nabla f(x)^t \cdot \gamma_s'(t)$$

$$= \langle \nabla f(x), \gamma_s'(t) \rangle$$

Compare with Figure 38.

Theorem 6.2. *This theorem establishes the relation of differentiability and partial derivatives.*

Let $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}^m$ is open. Assume $\forall x \in D$ exist all partial derivatives $\partial_i f(x)$ for $j = 1, \ldots, m$ and the functions $x \mapsto \partial_j f(x)$ on $D \to \mathbb{R}$ are continuous for $j = 1, \ldots, m$. Then f is differentiable in every point $x_0 \in D$ with $Df(x_0) = [\partial_1 f(x_0), \ldots, \partial_m f(x_0)]$. Then f is also continuously differentiable.

Proof. Proof idea: We approximate x_0 (starting from $x_0 \in D$) along lines parallel to the coordinate system axes. Compare with Figure 39.

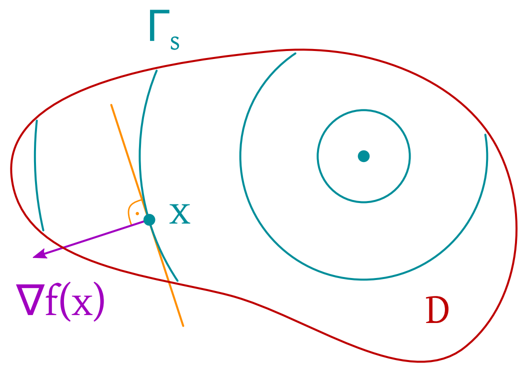

Figure 38: $Df(x) = \nabla f(x)^T$

$$x_0 := \begin{bmatrix} x_1^0 \\ \vdots \\ x_m^0 \end{bmatrix} \qquad x := \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \qquad \xi_0 := x_0 = \begin{bmatrix} x_1^0 \\ x_2^0 \\ \vdots \\ x_m^0 \end{bmatrix} \qquad \xi_1 := \begin{bmatrix} x_1 \\ x_2^0 \\ \vdots \\ x_m^0 \end{bmatrix} \qquad \xi_2 := \begin{bmatrix} x_1 \\ x_2 \\ x_3^0 \\ \vdots \\ x_m^0 \end{bmatrix}$$

$$\xi_k := \begin{bmatrix} x_1 \\ \vdots \\ x_{k_0} \\ x_{k+1} \\ \vdots \\ x_m^0 \end{bmatrix} \qquad \dots \qquad \xi_m := \begin{bmatrix} x_0 \\ \vdots \\ x_m \end{bmatrix} = x$$

where ξ_i are "intermediate" points. It holds that $\xi_k + (x_{k+1} - x_{k+1}^0) \cdot e_{k+1} = \xi_{k+1}$ for $k = 0, \dots, m-1$. Define $\varphi_k : [0,1] \to \mathbb{R}$.

$$\varphi_k(t) = f(\xi_k + t(x_{k+1} - x_{k+1}^0) \cdot e_{k+1})$$

then it holds that $\varphi_k(0) = f(\xi_k); \varphi_k(1) = f(\xi_{k+1})$. $\varphi_k'(t) =?$

$$\underbrace{f(\xi_k + t(x_{k+1} - x_{k+1}^0) \cdot e_{k+1})}_{=\varphi_k(t)} = f(x_1, \dots, x_k, x_{k+1}^0 + t(x_{k+1} - x_{k+1}^0), x_{k+2}^0, \dots, x_m^0)$$

$$\varphi_k'(t) = \frac{d}{dt}\left[f(x_1, \dots, x_k, \underbrace{x_{k+1}^0 + t(x_{k+1} - x_{k+1}^0)}_{(k+1)-\text{th variable}}, x_{k+2}^0, \dots, x_m^0) \right]$$

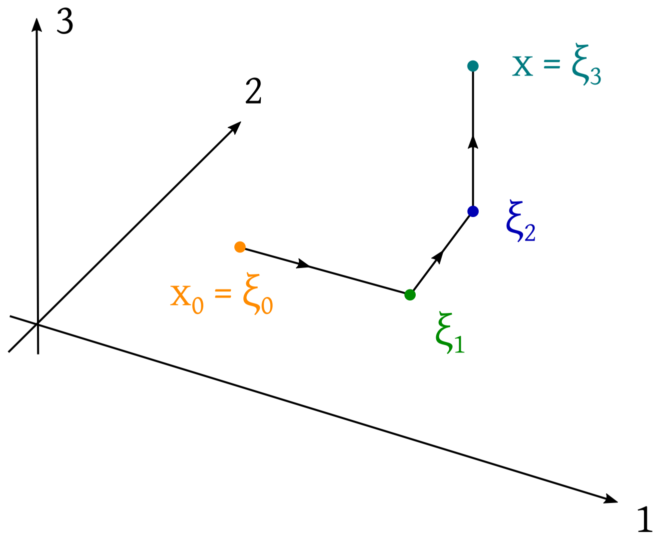

Figure 39: x_0 approximates x along lines parallel to the coordinate system axes

$$= \partial_{k+1} f(x_1, \dots, x_k, x_{k+1}^0 + t(x_{k+1} - x_{k+1}^0), x_{k+2}^0, \dots, x_m^0) \cdot (x_{k+1} - x_{k+1}^0)$$
$$= \partial_{k+1} f(\xi_k + t(x_{k+1} + t(x_{k+1} - x_{k+1}^0) \cdot e_{k+1})) \cdot (x_{k+1} - x_{k+1}^0)$$

φ_k is continuously differentiable on $[0,1]$ because $\partial_{k+1} f$ is continuous. By the mean value theorem of differential calculus, it follows that some $\tau_{k+1} \in (0,1)$ exists such that

$$\varphi(1) - \varphi(0) = \varphi'(\tau_{k+1}) \cdot (1-0)$$
$$\implies f(\xi_{k+1}) - f(\xi_k) = \partial_{k+1} f(\xi_k + \tau_{k+1}(x_{k+1} - x_{k+1}^0) \cdot e_{k+1}) \cdot (x_{k+1} - x_{k+1}^0)$$

For differentiability, we have to show:

$$\lim_{x \to x_0} \frac{1}{\|x - x_0\|} \left| f(x) - f(x_0) - [\partial_1 f(x_0), \dots, \partial_m f(x_0)] \begin{bmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \\ \vdots \\ x_m - x_m^0 \end{bmatrix} \right| = 0 \qquad (*)$$

Choose $\|x\| = \|x\|_\infty$ on $\mathbb{R}^m$.

$$\frac{1}{\|x - x_0\|} \left| \underbrace{f(x)}_{\xi_m} - \underbrace{f(x_0)}_{\xi_0} - \sum_{k=1}^{m} \partial_k f(x_0)(x_k - x_k^0) \right|$$

$$= \frac{1}{\|x - x_0\|} \left| \underbrace{\sum_{k=1}^{m} (f(\xi_k) - f(\xi_{k-1})}_{\text{telescoping sum}} - \partial_k f(x_0)(x_k - x_k^0)) \right|$$

$$\underbrace{\leq}_{\text{triangle ineq.}} \frac{1}{\|x - x_0\|} \sum_{k=1}^{m} \left| \partial_k f(\xi_{k-1} + \tau_k (x_k - x_k^0) \cdot e_k)(x_k - x_k^0) - \partial_k f(x_0)(x_k - x_k^0) \right|$$

$$= \sum_{k=1}^{m} \underbrace{\frac{\left| x_k - x_k^0 \right|}{\|x - x_0\|}}_{\leq 1} \cdot \left| \partial_k f(\xi_{k-1} + \tau_k (x_k - x_k^0) \cdot e_k) - \partial_k f(x_0) \right| \qquad (\#)$$

Choose $\varepsilon > 0$ arbitrary. By continuity of $\partial_k f(x)$, there exists $\delta > 0$ such that $\left\| y - y_0 \right\|_\infty < \delta \implies \left| \partial_k f(y) - \partial_k f(x_0) \right| < \frac{\varepsilon}{m}$ for every $k \in \{1, \dots, m\}$. Choose x such that $\|x - x_0\| \leq \delta$. We consider

$$\xi_{k-1} + \tau_k (x_k - x_k^0) \cdot e_k - x_0 = \begin{bmatrix} x_1 \\ \vdots \\ x_{k-1} \\ x_k^0 \\ \vdots \\ x_m \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \tau_k (x_k - x_k^0) \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} x_1^0 \\ \vdots \\ x_{k-1}^0 \\ x_k^0 \\ \vdots \\ x_m^0 \end{bmatrix} = \begin{bmatrix} x_1 - x_1^0 \\ \vdots \\ x_{k-1} - x_{k-1}^0 \\ \tau_k (x_k - x_k^0) \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\implies \left\| \xi_{k-1} + \tau_k (x_k - x_k^0) \cdot e_k - x_0 \right\|_\infty$$

$$= \max \left\{ \left| x_1 - x_1^0 \right|, \left| x_2 - x_2^0 \right|, \dots, \left| x_{k-1} - x_{k-1}^0 \right|, \underbrace{\tau_k}_{\in (0,1)} \left| x_k - x_k^0 \right| \right\}$$

$$\leq \|x - x_0\|_\infty < \delta$$

$$\implies \left| \partial_k f(\xi_{k-1} + \tau_k (x - x_0^k) e_k) - \delta_k f(x_0) \right| < \frac{\varepsilon}{m}$$

$$(\#) \leq 1 \cdot \sum_{k=1}^{m} \frac{\varepsilon}{m} = m \cdot \frac{\varepsilon}{m} = \varepsilon$$

Thus, f is differentiable in x_0.

Because $Df(x) = [\partial_1 f(x), \dots, \partial_m f(x)]$ depends continuously on x, f is continuously differentiable on D $\square$

Corollary. *Let $f : D \subseteq \mathbb{R}^m \to \mathbb{R}^n$.*

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix}$$

such that all partial derivatives $\partial_k f_i(x)$ exist and are continuous in x (for $k = 1, \dots, m$; $i = 1, \dots, n$). Then f is continuously differentiable on D.

Rationale. Use f is continuously differentiable on $D \iff f_i$ is continuously differentiable for $i = 1, \dots, n$ and use Theorem 6.2. □

Example 6.4 (Counterexample)**.**

$$f(x_1, x_2) = \frac{x_1^2 x_2}{x_1^2 + x_2^2} \qquad \partial_1 f(x) = \frac{2x_1 x_2^3}{(x_1^2 + x_2^2)^2} \qquad \partial_2 f(x) = \frac{x_1^2(x_1^2 - x_2^2)}{(x_1^2 + x_2^2)}$$

$$\partial_1 f(0) = \partial_2 f(0) = 0$$

$$\partial_1 f\left(\begin{bmatrix} \varepsilon \\ \varepsilon \end{bmatrix}\right) = \frac{2\varepsilon\varepsilon^3}{(\varepsilon^2 + \varepsilon^2)^2} = \frac{2\varepsilon^4}{4\varepsilon^4} = \frac{1}{2} \not\to 0$$

Therefore, $\partial_1 f$ is not continuous in 0*.*

This lecture took place on 2018/06/05.

6.10 Optimality criteria for multi-dimensional functions

Necessary (but not sufficient) optimality criteria:

Definition 6.9. *Let $D \subseteq \mathbb{R}^n$, $f : D \to \mathbb{R}$. We say that f in $x_0 \in D$ has a local maximum, if some neighborhood U of x_0 exists such that*

$$\forall x \in D \cap U : f(x) \leq f(x_0)$$

and accordingly for strict maxima:

$$\forall x \in (D \cap U) \setminus \{x_0\} : f(x) < f(x_0)$$

Analogously for a minimum with $f(x) \geq f(x_0)$ and strict minima $f(x) > f(x_0)$.

Lemma 6.8. *Let $f : D \to \mathbb{R}$ be given. Let $D \subseteq \mathbb{R}^n$ be open. Let $x_0 \in D$ be a local maximum or a local minimum and let f in x_0 be differentiable. Then it holds that $Df(x_0) = [0, \dots, 0]$ or $\nabla f(x_0) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.*

Proof. We consider one-dimensional $l_{e_k,x_0} : (-\varepsilon, \varepsilon) \to D$ with $l_{e_k,x_0}(t) \mapsto x_0 + t \cdot e_k$ and $l_{e_k,x_0}(0) \mapsto x_0$. f has a maximum in x_0 (without loss of generality) $\implies f \circ l_{e_k,x_0}$ (differentiable in $t = 0$) is a local maximum in $t = 0$. This is necessary for the optimality of $t = 0$ for $f \circ l_{e_k,x_0}$ that $(f \circ l_{e_k,x_0})'(0) = 0 \implies Df(x_0) \cdot e_k = \partial_k f(x_0) = 0 \implies Df(x_0) = [0, \dots, 0]$. □

Remark 6.16. *By Lemma 6.8 it follows immediately that x_0 is optimal.*

$$\implies \underbrace{Df(x_0)}_{=0} \cdot v = df(x_0; v) = 0 \qquad \forall v \in \mathbb{R}^n$$

6.11 Diffeomorphism

Definition 6.10. *Let $U, V \subseteq \mathbb{R}^n$ be open. We call $f : U \to V$ a* diffeomorphism *if*

1. *$f : U \to V$ is bijective*
2. *f is continuously differentiable on U*
3. *$f^{-1} : V \to U$ is continuously differentiable on V*

Let $f : U \to V$, $x_0 \in U$ and $y_0 = f(x_0) \in V$. We say: f is a local diffeomorphism in x_0 if open neighborhoods $x_0 \in U' \subset U$ and $y_0 \in V' \subseteq V$ such that $f : U' \to V'$ is a diffeomorphism.

Lemma 6.9. *Let $g : D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ and D is open. g is continuously differentiable on D. Furthermore let $g : D \to g(D)$ be bijective, $g^{-1} : g(D) \to D$ and $Dg(x)$ is regular for all $x \in D$. Then $g(D) \to D$ is continuously differentiable with*

$$Dg^{-1}(y) = Dg^{-1}(g(x)) = [Dg(x)]^{-1}$$

Proof. Will be provided later on (page 119). □

6.12 Local inversion theorem

Theorem 6.3 (Local Inversion Theorem, Theorem of Inverse Maps)**.** *Let $D \subseteq \mathbb{R}^n$ be open. Let $f : D \to \mathbb{R}^n$ be continuously differentiable. Furthermore let $Df(x_0)$ regular for $x_0 \in D$. Then f is a local diffeomorphism in x_0. Furthermore it holds that:*

$$Df^{-1}(f(x_0)) = [Df(x_0)]^{-1}$$

Proof.

$$\tilde{f}(\xi) = f(\xi + x_0) - f(x_0)$$
$$\tilde{f} : D - x_0 = \{\xi = x - x_0 : x \in D\} \to \mathbb{R}^n$$
$$\tilde{f}(0) = f(x_0) - f(x_0) = 0$$
$$D\tilde{f}(0) = Df(x_0) \text{ is regular}$$

Claim. *f is a local diffeomorphism $\iff$ $\tilde{f}$ is a local diffeomorphism.*

Proof. **Direction $\impliedby$** Let $\tilde{f}$ be a local diffeomorphism.

$$\tilde{f}(\xi) = \eta \iff \underbrace{f(\xi + x_0)}_{=x} = \underbrace{\eta + f(x_0)}_{y} \iff f(x) = y$$

and $\xi = \tilde{f}^{-1}(\eta)$. Also,

$$x = \xi + x_0 = \tilde{f}^{-1}(y - f(x_0)) + x_0$$

Thus, f is invertible and

$$f^{-1}(y) = \tilde{f}^{-1}(y - f(x_0)) + x_0$$

hence f^{-1} is continuously differentiable, hence f is a local diffeomorphism.

Direction $\implies$ Analogous.

$\square$

$$F(x) = \underbrace{[Df(x_0)]^{-1}}_{\in\mathbb{R}^{n\times n}} \tilde{f}(x)$$

$$v \mapsto [Df(x_0)]^{-1} \cdot v \text{ linear} \implies \text{diffeomorphism}$$

$[Df(x_0)]^{-1} \cdot \tilde{f}$ is a local diffeomorphism $\iff$ $\tilde{f}$ is a local diffeomorphism.

$$DF(0) = [Df(x_0)]^{-1} \cdot \underbrace{D\tilde{f}(0)}_{Df(x_0)} = E$$

Thus, it suffices to show that F is a local diffeomorphism where $F(0) = 0$ and $Df(0) = I$ (the unit matrix).

Now the interesting part of the proof begins: Let $y \in \mathbb{R}^n$ arbitrary. We define

$$\varphi(x) = y + x - F(x)$$

It holds that $y = F(x) \iff y + x - F(x) = x$, hence $\varphi_Y(x) = x$, hence x is a fixed point of φ_Y.

F is continuously differentiable with $DF(0) = I$. Thus for every $\varepsilon > 0 \exists \delta > 0$ such that

$$\|x - 0\| \leq \delta \implies \|DF(x) - DF(0)\| = \|DF(x) - I\| \leq \varepsilon$$

Choose $r > 0$ such that $\|x\| \leq 2r \implies \|DF(x) - I\| \leq \frac{1}{2}$. Additionally let r be sufficiently small such that

$$\overline{K_{2r}(0)} \subseteq \tilde{D} = D - x_0$$

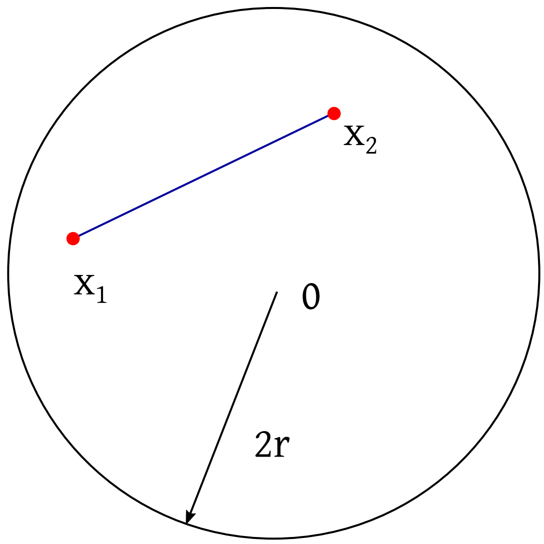

Figure 40: We consider a sphere with radius $2r$. The blue line denotes $l_{x_1,x_2}(t)$. $l_{x_1,x_2}(t) \in \overline{K_{2r}(0)}$ because $\overline{K_{2r}(0)}$ is convex

Recall that $\tilde{D}$ is the domain of F. Let $x_1, x_2 \in \overline{K_{2r}(0)}$ and let

$$l_{x_1,x_2} : \substack{t \mapsto (1-t)x_1 + tx_2 \\ [0,1] \to \overline{K_{2r}(0)}}$$

Compare with Figure 40.

$$l_{x_1x_2}(0) = x_1 \qquad l_{x_1x_2}(1) = x_2$$

$$\begin{aligned}
\left\|\varphi_y(x_2) - \varphi_y(x_1)\right\| &= \left\|\varphi_y \circ l_{x_1x_2}(1) - \varphi_y \circ l_{x_1x_2}(0)\right\| \\
&= \left\| \int_0^1 \frac{d}{dt}[\varphi_Y \circ l_{x_1x_2}(t)]\, dt \right\| \\
&= \left\| \int_0^1 D\varphi_Y(l_{x_1x_2}(t)) \cdot \underbrace{(x_2 - x_1)}_{\substack{\text{inner derivative} \\ l_{x_1x_2}(t)}} dt \right\| \\
&\leq \int_0^1 \left\|D\varphi_Y(l_{x_1x_2}(t))\right\| \cdot \|(x_2 - x_1)\|\, dt \\
&\leq (*)
\end{aligned}$$

$$D\varphi_Y(x) = I - DF(x)$$

hence $\forall x \in \overline{K_{2r}(0)}$ it holds that $\left\|D\varphi_Y(x)\right\| = \|I - DF(x)\| \leq \frac{1}{2}$.

$$(*) \leq \frac{1}{2}\|x_2 - x_1\| \underbrace{\int_0^1 1 \cdot dt}_{=1} = \frac{1}{2}\|x_2 - x_1\|$$

Hence $\forall x_1, x_2 \in \overline{K_{2r}(0)}$ it holds that

$$\left\|\varphi_Y(x_2) - \varphi_Y(x_1)\right\| \leq \frac{1}{2}\|x_2 - x_2\|$$

Therefore φ_Y is a contraction with constant $\nu = \frac{1}{2}$.

$K_{2r}(0) \subseteq \mathbb{R}^n$ is a complete, metric space. It remains to show: $\varphi_Y : \overline{K_{2r}(0)} \to \overline{K_{2r}(0)}$. This only holds if y is sufficiently small.

Let $y \in K_r(0)$ and $x \in \overline{K_{2r}(0)}$.

$$\left\|\varphi_Y(x)\right\| = \left\|\varphi_Y(x) - \varphi_Y(0) + \varphi_Y(0)\right\| \leq \left\|\varphi_Y(x) - \varphi_Y(0)\right\| + \left\|\varphi_Y(0)\right\| \underbrace{\leq}_{\text{contraction}} \frac{1}{2}\underbrace{\|x - 0\|}_{\leq 2r} + \underbrace{\left\|y\right\|}_{<r} < 2r$$

So, for $y \in K_r(0)$ and $x \in \overline{K_{2r}(0)}$, $\varphi_Y(x) \in K_{2r}(0)$. By Banach Fixed Point Theorem, there exists a uniquely determined x such that

$$x = \varphi_Y(x) \iff y = F(x)$$

Let $U = F^{-1}(K_r(0)) \cap K_{2r}(0)$ where $F^{-1}(K_r(0))$ is open, because F is continuous. Hence, U is open and $0 \in U$. Therefore U is an open neighborhood of 0.

Show: $F^{-1} : F(U) \to U$ is continuous [3]: Let $x_1 = F^{-1}(y_1), x_2 = F^{-1}(y_2); y_1, y_2 \in F(U)$.

$$\varphi_0(x) = x + 0 - F(x) = x - F(x)$$

$$\begin{aligned}
x_2 - x_1 &= \underbrace{x_2 - F(x_2)}_{\varphi_0(x_2)} + F(x_2) \underbrace{- x_1 + F(x_1)}_{-\varphi_0(x_1)} - F(x_1) \\
&= \varphi_0(x_2) - \varphi_0(x_1) + F(x_2) - F(x_1) \\
\|x_2 - x_1\| &\leq \left\|\varphi_0(x_2) - \varphi_0(x_1)\right\| + \|F(x_2) - F(x_1)\| \\
&\underbrace{\leq}_{\substack{\varphi_0 \text{ is} \\ \text{a contraction}}} \frac{1}{2}\|x_2 - x_1\| + \|F(x_2) - F(x_1)\|
\end{aligned}$$

[3] This proof was added on 2018/06/07 as we initially forgot this condition

$$\left\| F^{-1}(y_2) - F^{-1}(y_1) \right\| \leq 2 \left\| y_2 - y_1 \right\|$$

hence F^{-1} is Lipschitz continuous on $F(U)$, so also continuous.

We show: $DF(x)$ is invertible for all $x \in U$.

Let $v \in \ker(DF(x)) \iff \|v\| = \|DF(x) \cdot v - v\|$ where $DF(x) \cdot v = 0$.

$$= \|(DF(x) - I) \cdot v\| \leq \underbrace{\|DF(x) - I\|}_{\leq \frac{1}{2}} \cdot \|v\| \leq \frac{1}{2} \|v\|$$

$$\implies \frac{1}{2} \|v\| \leq 0 \implies \|v\| \leq 0 \implies v = 0$$

Hence $\ker(DF(x)) = \{0\}$, so $DF(x)$ is regular.

$$\forall x \in U = F^{-1}(K_r(0)) \cap K_{2r}(0) : y = F(x) \in K_r(0)$$

On the opposite, $\forall y \in K_r(0)$ there exists some uniquely determined $x \in K_{2r}(0) \cap F^{-1}(K_r(0))$ with $y = F(x)$. So $F : U \to K_r(0)$ is bijective and continuously differentiable. Furthermore, $DF(x)$ is regular $\forall x \in U$. By Lemma 6.9, F is a local diffeomorphism in $x = 0$. Thus, f is a local diffeomorphism in x_0.

So the central idea of this proof was to rewrite F such that we can apply Banach's Fixed Point Theorem. □

6.13 Implicit functions

Theorem 6.4 (Implicit function theorem)**.** *Let $U \subseteq \mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$ $f : U \to \mathbb{R}^m$ is continuously differentiable.*

Notation: $\begin{pmatrix} x \\ y \end{pmatrix} \in U; x \in \mathbb{R}^n, y \in \mathbb{R}^m$

$$f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = f(x, y)$$

For $(x_0, y_0) \in U$ it holds that $f(x_0, y_0) = 0$. What about

$$Df(x_0, y_0) = \begin{bmatrix} \partial_{x_1} f_1 & \dots & \partial_{x_n} f_1 & \partial_{y_1} f_1 & \dots & \partial_{y_m} f_1 \\ \vdots & & \vdots & \vdots & & \vdots \\ \partial_{x_1} f_m & \dots & \partial_{x_n} f_m & \partial_{y_1} f_m & \dots & \partial_{y_m} f_m \end{bmatrix}$$

where the left half is given by $D_x f(x_0, y_0) \in \mathbb{R}^{m \times n}$ and the right half is given by $D_y f(x_0, y_0) \in \mathbb{R}^{m \times m}$.

Assumption: Let $D_y f(x_0, y_0)$ be regular.

Then there exists some neighborhood D of x_0 in $\mathbb{R}^n$ and a function $g : D \to E$. E is a neighborhood of y_0 in $\mathbb{R}^m$ such that $D \times E \subseteq U$ and $f(x, y) = 0 \iff y = g(x)$ for $(x, y) \in D \times E$. Hence, $f(x, g(x)) = 0 \forall x \in D$.

This lecture took place on 2018/06/07.

Remark 6.17. *So it holds: $g(x_0) = y_0$.*

Proof.

$$F : U \subset \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$$

$$F\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x & (\in \mathbb{R}^n) \\ f(x,y) & (\in \mathbb{R}^m) \end{bmatrix} \in \mathbb{R}^{n+m}$$

$$DF\left(\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}\right) = \begin{bmatrix} I & 0 \\ D_x f(x_0, y_0) & D_y f(x_0, y_0) \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}$$

is the Jacobi matrix of f.

$$\det\left(DF(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix})\right) = \underbrace{\det(I)}_{=1} \cdot \underbrace{\det(D_y f(x_0, y_0))}_{\neq 0} \neq 0$$

so $DF(\begin{pmatrix} x_0 \\ y_0 \end{pmatrix})$ is regular. By the local inversion theorem (Theorem 6.3), F is a local Diffeomorphism. Thus, $\exists V : \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \in V \subseteq D$ such that $F : V \to F(V)$ is a diffeomorphism. V is open, hence $\exists r, r' > 0$ such that

$$(x_0, y_0) \in K_r(x_0) \times K_{r'}(y_0) \subseteq V$$

here $\mathbb{R}^{n+m}$ is identified as $\mathbb{R}^n \times \mathbb{R}^m$.

$$F^{-1} : F(V) \to V \qquad F^{-1}(\begin{bmatrix} \xi \\ \eta \end{bmatrix}) = ?$$

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} \xi \\ \eta \end{bmatrix} \iff \begin{bmatrix} x \\ y \end{bmatrix} = F^{-1}(\begin{bmatrix} \xi \\ \eta \end{bmatrix})$$

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ f(x,y) \end{pmatrix}$$

hence $x = \xi$ and $F^{-1}(\begin{bmatrix} \xi \\ \eta \end{bmatrix}) = \begin{bmatrix} \xi \\ G(\xi, \eta) \end{bmatrix}$. Define $g(x) := G(x, 0)$. $g : K_r(x_0) \to \mathbb{R}^m$.

$$\begin{bmatrix} x \\ 0 \end{bmatrix} = F\left(F^{-1}\left(\begin{bmatrix} x \\ 0 \end{bmatrix}\right)\right) = F\left(\begin{bmatrix} x \\ G(x,0) \end{bmatrix}\right) = F\left(\begin{bmatrix} x \\ g(x) \end{bmatrix}\right) = \begin{bmatrix} x \\ f(x, g(x)) \end{bmatrix}$$

$$\implies f(x, g(x)) = 0 \forall x \in K_r(x_0)$$

Uniqueness: Let $(x, y) \in K_r(x_0) \times K_{r'}(y_0) \subset V$ and $f(x,y) = 0$. Hence

$$F(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} x \\ f(x,y) \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix}$$

$$\implies \begin{bmatrix} xy \end{bmatrix} = F^{-1}(F(\begin{bmatrix} x \\ y \end{bmatrix})) = F^{-1}(\begin{bmatrix} x \\ 0 \end{bmatrix}) = \begin{bmatrix} x \\ G(x,0) \end{bmatrix} = \begin{bmatrix} x \\ g(x) \end{bmatrix}$$
$$\implies y = g(x)$$

□

Example 6.5.
$$f(x,y) = x^2y - x^3y^2 - 2$$
$$f(-1,1) = 1 + 1 - 2 = 0$$

Hence $(x_0, y_0) = (-1, 1)$ *is root of* F.

$$Df(x,y) = \begin{bmatrix} \underbrace{2xy - 3x^2y^2}_{D_xf}, \underbrace{x^2 - 2x^3y}_{D_yf} \end{bmatrix}$$

$$D_yf(\begin{bmatrix} -1 \\ 1 \end{bmatrix}) = 3 \neq 0$$

hence $f(x,y)$ *is in neighborhood of* $(-1,1)$ *resolvable by* y.

$$x^2y - x^3y^2 - 2 = 0$$
$$x^3y^2 - x^2y + 2 = 0$$
$$y = \frac{x^2 \pm \sqrt{x^4 - 8x^3}}{2x^3} = \frac{x^2(1 \pm \sqrt{1 - \frac{8}{x}})}{2x^3} = \frac{1}{2x}\left(1 \pm \sqrt{1 - \frac{8}{x}}\right)$$

It has to holds that $g(-1) = 1$, *thus*

$$\frac{1}{-2}\left(1 + \sqrt{1 - \frac{8}{-1}}\right) = -\frac{1}{2}(1+3) = -2$$

This is apparently wrong, thus we consider the second result for y:

$$\frac{1}{-2}\left(1 - \sqrt{1 - \frac{8}{-1}}\right) = -\frac{1}{2}(1-3) = -1$$

So $g(x) = \frac{1}{2x} \cdot \left(1 - \sqrt{1 - \frac{8}{x}}\right)$ *is the desired function.* $D_xf(-1,1) = -2 - 3 = -5$, *hence* $f(x,y) = -x^3y^2 + x^2y - 2$ *is also uniquely resolvable by* x *in a neighborhood of* $(-1,1)$.

Proof of Lemma 6.9. Let choose $x_0 \in D$ and show that g^{-1} is differentiable in $y_0 = g(x_0)$. We use the same construction as in Proof 6.13 (Implicit function theorem proof). Without loss of generality: $x_0 = 0$; $g(x_0) = y_0 = 0$ and $Dg(0) = I$. Let $v \in \mathbb{R}^n$ be sufficiently small such that $w = g^{-1}(v)$ is defined.

By differentiability of g,

$$g(w) = \underbrace{Dg(0)}_{I} \cdot w + \underbrace{R(w)}_{o(\|w\|)} = w + R(w)$$

$$g^{-1}(v) = w = g(w) - R(w) = g(g^{-1}(v)) - R(g^{-1}(v)) = v + R^*(v)$$

with $R^*(v) = -R(g^{-1}(v))$. Show that $R^*(v) = o(\|v\|)$.

$$g^{-1}(v) = v + R^*(v) \tag{5}$$

By differentiability of g,

$$\exists r > 0 : \|R(w)\| \leq \frac{1}{2} \|w\|$$

for all $\|w\| \leq r$ (because $\frac{\|R(w)\|}{\|w\|} \to 0$ for $w \to 0$). Continuity of g^{-1} combined with $g^{-1}(0) = 0$

$$\implies \forall \|v\| < \delta : \left\|g^{-1}(v)\right\| = \left\|g^{-1}(v) - g^{-1}(0)\right\| \leq r$$

$$\implies \|R^*(v)\| = \left\|R(g^{-1}(v))\right\| \leq \frac{1}{2} \left\|g^{-1}(v)\right\|$$

for all v with $\|v\| \leq \delta$.

By Equation (5),

$$\left\|g^{-1}(v)\right\| = \|v + R^*(v)\| \leq \|v\| + \|R^*(v)\| \leq \|v\| + \frac{1}{2} \left\|g^{-1}(v)\right\|$$

If $\|v\| \leq \delta$

$$\left\|g^{-1}(v)\right\| \leq 2 \|v\| \qquad \forall \|v\| \leq \delta$$

So,

$$\frac{\|R^*(v)\|}{\|v\|} \leq 2 \frac{\left\|R(g^{-1}(v))\right\|}{\left\|g^{-1}(v)\right\|} = 2 \frac{\|R(w)\|}{\|w\|}$$

$$v \to 0 \implies w = g^{-1}(v) \to 0$$

because g^{-1} is continuous, hence

$$v \to 0 : \lim_{v \to 0} \frac{\left\|R(g^{-1}(v))\right\|}{\left\|g^{-1}(v)\right\|} = \lim_{w \to 0} \frac{R(w)}{\|w\|} = 0$$

because $R = o(\|w\|)$. Thus, $R^*(v) = o(\|v\|)$, so g^{-1} is differentiable in $y_0 = 0$.

We know that g^{-1} is differentiable in every point $y \in g(D)$.

$$\underbrace{y}_{=\mathrm{id}(y)} = g(g^{-1}(y)) \forall y \in E = g(D)$$

We derive both sides of the equation and apply the chain rule:

$$D\,\mathrm{id}(y) = I = Dg(g^{-1}(y)) \cdot Dg^{-1}(y)$$

$$\implies Dg^{-1}(y) = \underbrace{[Dg(\underbrace{g^{-1}(y)}_{\text{continuous}})]^{-1}}_{\text{continuous}}$$

Is the inverse also continuous (does the inverse depend continuously on the coefficients? Yes, you can see it by considering Cramer's Rule which provides a formula with a sum)? So the inverse is a continuous operation on $\mathrm{GL}(n)$. Therefore $Dg^{-1}(y)$ depends continuously on y and $Dg^{-1}(y) = [Dg(x)]^{-1}$ with $x = g^{-1}(y)$ and accordingly, $y = g(x)$. □

6.14 Higher partial derivatives and multi-dimensional Taylor Theorem

Remark 6.18 (Concept idea)**.** *Let $D \subseteq \mathbb{R}^n$ be open. Let $f : D \to \mathbb{R}$ be continuously differentiable. Hence $\partial_{x_i} f(x) \in \mathbb{R}$ and $\partial_{x_i} f : D \to \mathbb{R}$ is continuous. If $\partial_{x_i} f$ is also (continuously) differentiable, then its partial derivatives can be determined. In this case, we define*

$$\partial_{X_j,X_i} f := \partial_{X_j}[\partial_{X_i} f]$$

Continuation for further higher derivatives:

$$\partial_{X_{i_k},X_{i_{k-1}},\ldots,X_{i_1}} f = \partial_{X_{i_k}}(\partial_{X_{i_{k-1}}\ldots X_{i_1}} f)$$

The index k in $\partial_{X_{i_k}X_{i_{k-1}}\ldots X_{i_1}}$ is the order of the partial derivative.

Example 6.6.

$$\begin{aligned}
f(x, y) &= x^2y - x^3y^2 - 2 \\
\partial_X f &= 2xy - 3x^2y^2 \\
\partial_Y f &= x^2 - 2x^3y \\
\partial_{YX} f &= 2x - 6x^2y \\
\partial_{XY} f &= 2x - 6x^2y = \partial_{YX} \\
\partial_{XX} f &= 2y - 6xy^2 \\
\partial_{YY} f &= -2x^3 \\
\partial_{XXX} f &= -6y^2 \\
\partial_{YYY} f &= 0 \\
\partial_{XYX} f &= 2 - 12xy \\
\partial_{XXY} f &= 2 - 12xy = \partial_{XYX} \\
\partial_{YXY} f &= -6x^2 \\
\partial_{XYY} f &= -6x^2 = \partial_{YXY}
\end{aligned}$$

It seems that the derivative is independent of the order of the variables.

Definition 6.11 (Hesse matrix)**.** *Specifically for second derivatives:*

$$D^2 f(x) = \begin{bmatrix} \partial_{X_1,X_1} f & \partial_{X_2,X_1} f & \dots & \partial_{X_n,X_1} f \\ \partial_{X_1,X_2} f & \partial_{X_2,X_2} f & \dots & \partial_{X_n,X_2} f \\ \vdots & \vdots & \ddots & \vdots \\ \partial_{X_1,X_n} f & \partial_{X_2,X_n} f & \dots & \partial_{X_n,X_n} f \end{bmatrix}$$

is called Hessian matrix*. Named after Otto Hesse (1811–1874).*

Definition 6.12. *Let $f : D \to \mathbb{R}$ and let all partial derivatives of f up to order $k \in \mathbb{N}$ exist and be continuous functions in $x \in D$. Then f is called* k-times continuously differentiable *and*

$$C^k(D) := \{f : D \to \mathbb{R} : f \text{ is } k\text{-times continuously differentiable}\}$$

$C^k(D)$ is a vector space.

Theorem 6.5 (Symmetry of second derivatives, Schwarz' theorem, Clairaut's theorem, Young's theorem)**.** *Hermann Amadeus Schwarz (1843–1921).*

Let $f \in C^2(D)$. Let $D \subseteq \mathbb{R}^n$ be open. Then it holds that

$$\partial_{X_iX_j} f = \partial_{X_jX_i} f \text{ on } D$$

Index