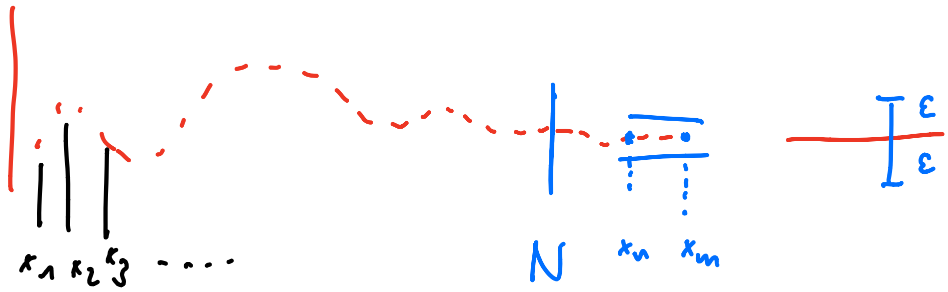
Sequences and convergence

Def:

$(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}^d$ is called a <u>Cauchy sequence</u> if

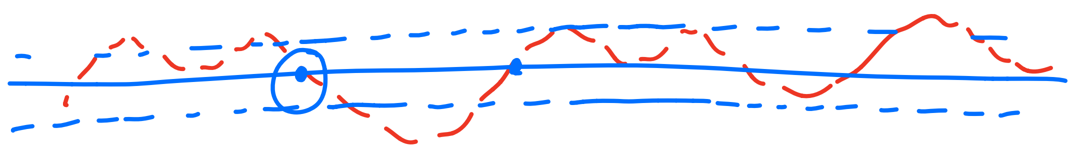
$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall n,m > N : |x_n - x_m| < \varepsilon$$



A point $x \in \mathbb{R}^d$ is called <u>accumulation point</u> of the sequence $(x_n)_n$ if

typo corrected

$$\forall \varepsilon \; \forall N \in \mathbb{N} \; \exists n > N : |x_n - x| < \varepsilon$$



typo corrected

A sequence $(x_n)_n$ converges to $x \in \mathbb{R}^d$ if

$$\forall \varepsilon > 0 \; \exists N \; \forall n > N : |x_n - x| < \varepsilon$$

<u>Notation</u>: $\lim_{n\to\infty} x_n = x$, $x_n \to x$

Observations

- a sequence can have many acc. points (or no acc. point)
- even if the sequence has just one acc. point, it is not necc. a Cauchy sequence.
- If $(x_n)_n$ converges to x, then x is the only acc. point and the sequence is Cauchy.

Example: $x_n = \frac{1}{n}$ on $]0,1]$

$(x_n)_n$ is Cauchy, but does not converge on $]0,1]$.

It does converge to 0 on $[0,1]$.

Max, sup, Min, inf

Assume we are on $\mathbb{R}$ (or more general, on a space that has a total ordering). Let $U \subset \mathbb{R}$ be a subset.

- $x \in \mathbb{R}$ is called a maximum element of U if
  $x \in U$ and $\forall u \in U : u \leq x$.

  - 1 is max. of $[0,1]$
  - $]0,1[$ does not have a max.

- x is called an upper bound of U if
  $\forall u \in U : u \leq x$

  5 is an upper bound of $[0,1]$

- x is called supremum of U if it is the smallest upper bound.

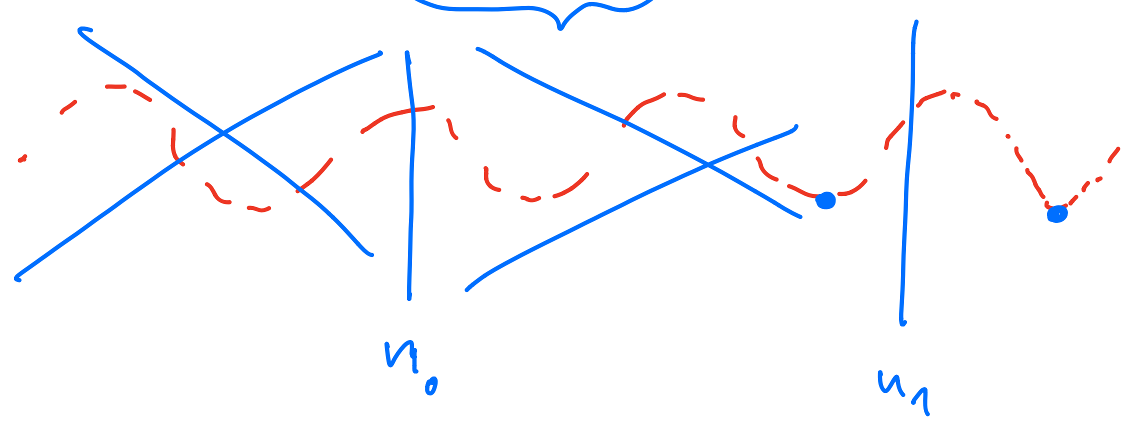
  1 is the sup of $]0,1[$

Analogously, min, lower bound, infimum.

Lim inf, lim sup

For a sequence $(x_n)_n \subset \mathbb{R}$ we define:

$$\liminf_{n\to\infty} x_n := \lim_{n\to\infty} \left(\inf_{m \ge n} x_m \right)$$



$$\limsup_{n\to\infty} x_n := \lim_{n\to\infty} \left(\sup_{m \ge n} x_m \right)$$

For a bounded sequence $(x_n)_n$ (i.e. there exists an upper bound $u \in \mathbb{R}$ s.th. $\forall n \in \mathbb{N}: x_n \le u$, and a lower bound $\ell \in \mathbb{R}$ $\forall n \in \mathbb{N}: x_n \ge \ell$)

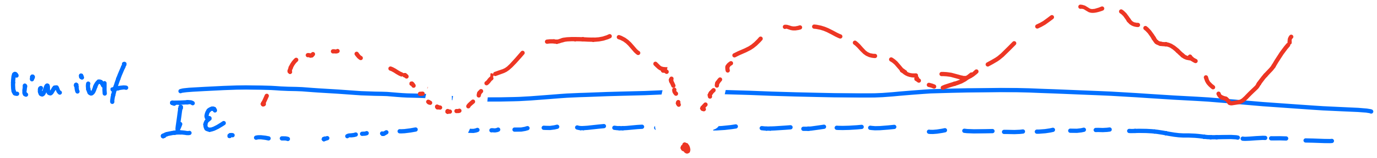
the limsup is the largest accumulation point of $(x_n)_n$.

liminf smallest

typo corrected

Then the liminf is the largest $\gamma \in \mathbb{R}$ such that

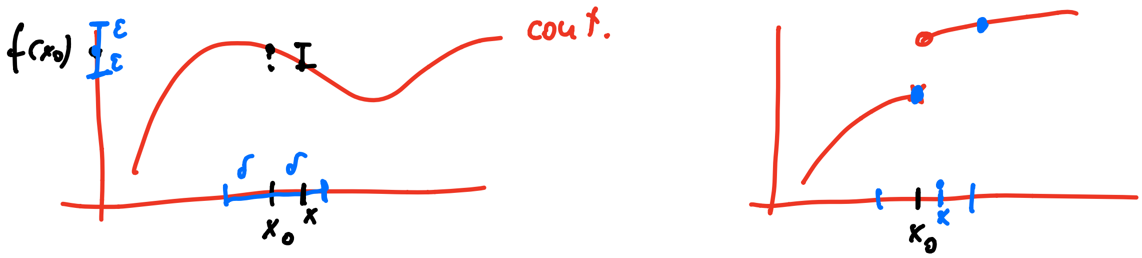
$$\forall \varepsilon > 0 \; \exists N \; \forall n > N : x_n > \gamma - \varepsilon .$$



Continuity

Def A function $f: X \to Y$ between two metric spaces (X, d), (Y, d) is called continuous at $x_0 \in X$ if

$$\forall \varepsilon > 0 \ \exists \delta > 0 \ \forall x \in X: d(x, x_0) < \delta \Rightarrow d(f(x), f(x_0)) < \varepsilon$$



Alternative definition: $f: X \to Y$ is called cont. at x_0 if for every sequence $(x_n)_n \subset X$ we have:

$$x_n \to x_0 \Rightarrow f(x_n) \to f(x_0)$$

A function $f: X \to Y$ is called continuous if it is continuous for every $x_0 \in X$:

$$\forall x_0 \in X \ \forall \varepsilon > 0 \ \exists \delta > 0 \ \forall x \in X: d(x, x_0) < \delta \Rightarrow d(f(x), f(x_0)) < \varepsilon$$

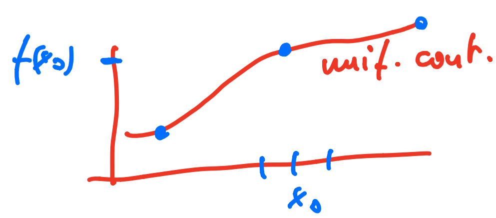
A function $f: X \to Y$ is called Lipschitz continuous with Lipschitz constant L if

$$\forall x, y \in X: \quad d(f(x), f(y)) \leq L \cdot d(x, y)$$

Intuition: "bounded derivative"

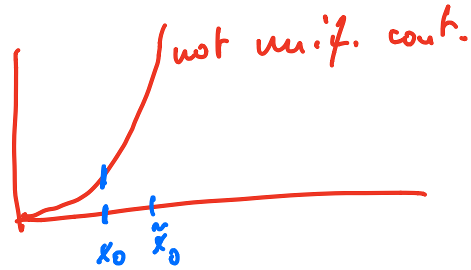
A function $f: X \to Y$ is called <u>uniformly continuous</u> if

$$\forall \varepsilon > 0 \ \exists \delta > 0 \ \underline{\forall x_0 \in X} \ \forall x \in X: d(x, x_0) < \delta \Rightarrow d(f(x), f(x_0)) < \varepsilon.$$



Given ε, I can choose δ that works for all x_0

Intuition: bounded derivative



Cannot choose δ to be the same for all x_0

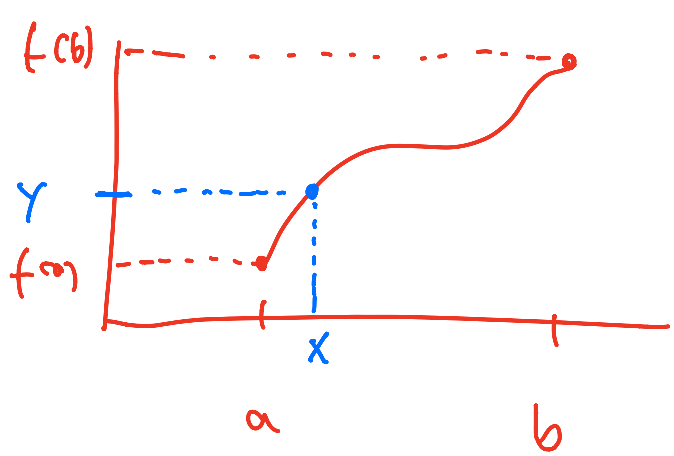
Intuition: unbounded derivative

Important theorems for cont. functions

<u>Intermediate value theorem</u>:

If $f: [a,b] \to \mathbb{R}$ is continuous, then f attains all values between $f(a)$ and $f(b)$:

$$\forall y \in [f(a), f(b)] \ \exists x \in [a,b]: \ f(x) = y.$$



Application: If you want to find x with $f(x)=0$:

- find a with $f(a) < 0$,
  b with $f(b) > 0$

Then there must exist $x \in [a,b]$ with $f(x) = 0$.

<u>Invertible functions</u>: $D \subset \mathbb{R}$, $f: D \to \mathbb{R}$ continuous, strictly monotone ($a < b \Rightarrow f(a) < f(b)$). Then f is invertible and the inverse is continuous as well.

- Invertible follows from monotonicity

- Continuity of the inverse follows directly from cont. of f.

Additional note: in the video I forgot the following important characterization of continuous functions:

A function f between two metric spaces (X,d), (Y,δ) is continuous if and only if pre-images of open sets are open:

$$\underbrace{B \subset Y \text{ open}}_{\text{in } Y} \Rightarrow \underbrace{f^{-1}(B) := \{x \in X \mid f(x) \in B\} \text{ open}}_{\text{in } X}$$

Sequences of functions

Def: Consider functions: $f_n : D \to \mathbb{R}$, $D \subset \mathbb{R}$.

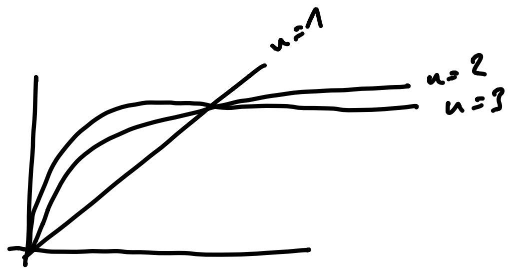
We say that the sequence $(f_n)_{n\in\mathbb{N}}$ converges pointwise to $f : D \to \mathbb{R}$ if

$$\forall x \in D : \quad f_n(x) \to f(x)$$

$y_n := f_n(x)$, $y := f(x)$

$y_n \to y$

Example : $f_n, f : [0,1] \to \mathbb{R}$, $f_n(x) = x^{1/n}$



$$f(x) = \begin{cases} 0 & x = 0 \\ 1 & \text{otherwise} \end{cases}$$

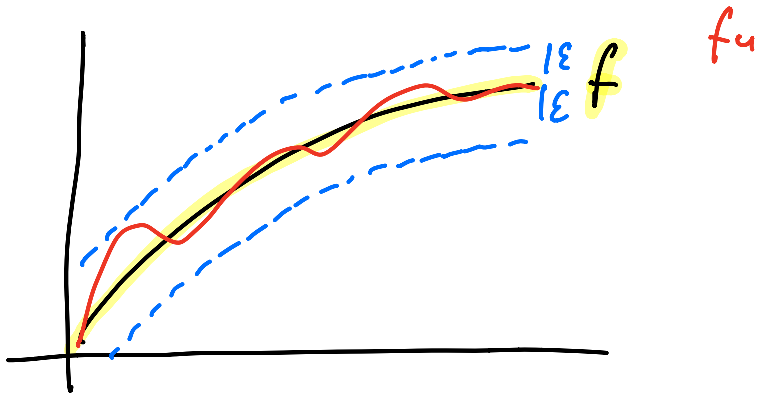
⚠ $f_n \to f$ pointwise, all f_n continuous, this does not imply that f is continuous.

Def $(f_n)_n$ converges to f uniformly if

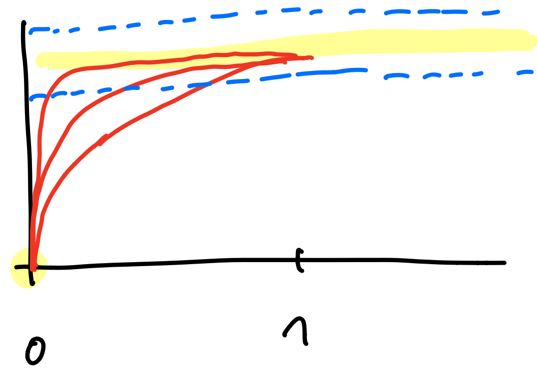
$$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n > N \ \forall x \in D : \ |f_n(x) - f(x)| < \varepsilon$$

Intuition:

uniform convergence: given ε, there exist N such that all f_n with $n > N$ are contained in ε-tube.



Close to 0 there will always be points x close to 0 such that the red fcts $f_n(x)$ are not yet in ε-tube.
$\Rightarrow$ Not uniformly conv.



Alternative definition: $f_n \to f$ uniformly iff $\|f_n - f\|_\infty \to 0$.

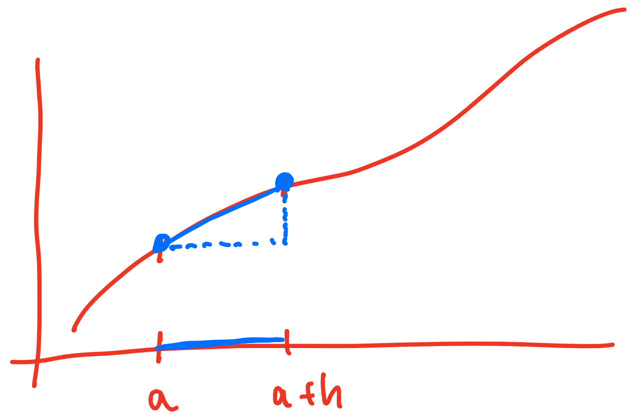
> **Theorem** (Uniform convergence preserves continuity)
> $f_n, f : D \to \mathbb{R}$, $D \subset \mathbb{R}$, all f_n are continuous, $f_n \to f$ uniformly. Then f is continuous.

Derivatives (1-dim case)

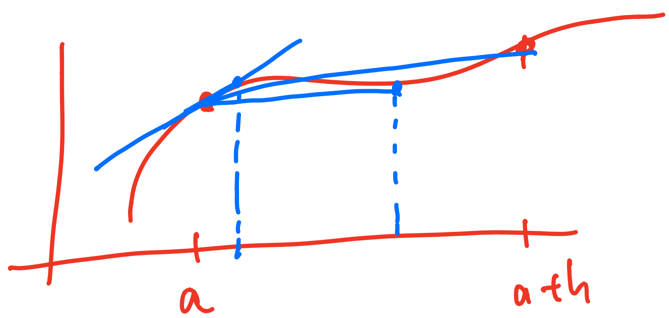
Def $U \subset \mathbb{R}$ an interval, $f : U \to \mathbb{R}$. The function f is called differentiable at $a \in U$ if

$$f'(a) := \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \quad \text{exists.}$$

We often write $f' = \frac{df}{dx}$



$a_n := a + h_n$, $h_n \to 0$
$h_n > 0$



The function f is called differentiable if it is differentiable for all $a \in U$. It is continuously differentiable if it is diff. and the function $f' : U \to \mathbb{R}$, $a \mapsto f'(a)$ is continuous.

Higher derivatives:

We can repeat the process of taking derivatives:

$$f' = \frac{df}{dx} \quad ; \quad f'' = \frac{df'}{dx}$$

Notation: $f^{(n)}$ denotes the n-th derivative (if exists).

Important theorems

Theorem (Differentiable implies continuous)

Let f be differentiable at a. Then there exists a constant c_a such that on a small ball around a we have

$$|f(x) - f(a)| \leq c_a \cdot |x-a|$$

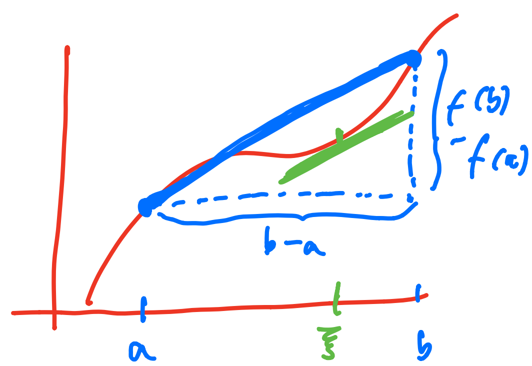
In particular, f is continuous at a.

Theorem (Intermediate value theorem for derivatives)

$f \in \mathcal{C}^1([a,b])$. (i.e. functions on $]a,b[$ that are once cont. differentiable).

Then there exist $\xi \in [a,b]$ such that

$$\frac{f(b) - f(a)}{b-a} = f'(\xi).$$



missing in video

Theorem (Exchanging lim and derivative)

$f_n : [a,b] \to \mathbb{R}$, $f_n \in \mathcal{C}^1[a,b]$. If the limit

$f(x) := \lim_{n\to\infty} f_n(x)$ exists for all $x \in [a,b]$ and the derivatives

f_n' converge uniformly, then f is cont. differentiable and we have

$$(f')(x) = \left(\lim f_n\right)'(x) \overset{!}{=} \left(\lim (f_n')\right)(x)$$

first take limit of f_n, we obtain f, and then we compute its derivative

first compute all f_n', then take limit of these derivatives

⚠ uniform conv. is really important, otherwise would be wrong!

Riemann integrals

Consider a function $f : [a,b] \to \mathbb{R}$, assume that f is bounded

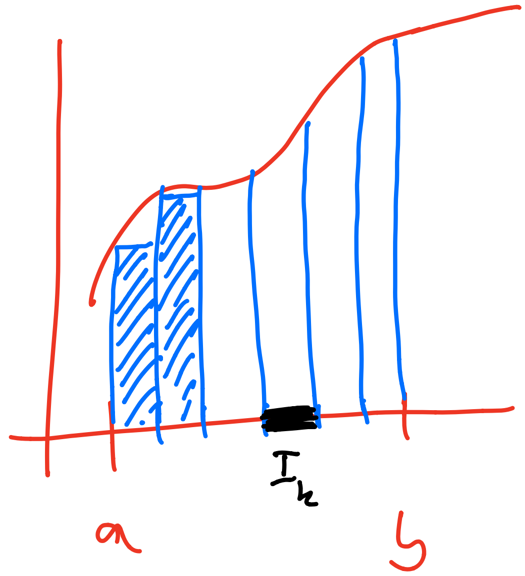
$(\exists\, l, u \in \mathbb{R}\ \forall x \in [a,b] : \ l \le f(x) \le u)$.

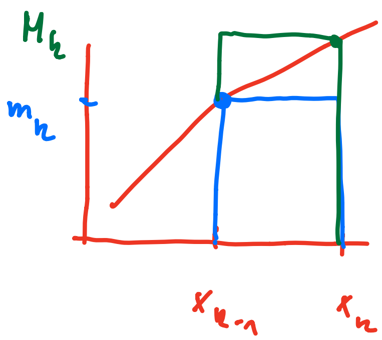
Consider $x_0, x_1, \dots, x_n$ with

$a = x_0 < x_1 < x_2 \ \dots \ < x_n = b.$

These points introduce a partition of $[a,b]$ into n intervals

$$I_k := [x_{k-1}, x_k].$$





Define $m_k := \inf\,(f\,(I_k))$

$M_k := \sup\,(f\,(I_k))$

(exists because f is bounded).

Define the lower sum

length of $I_k = x_k - x_{k-1}$

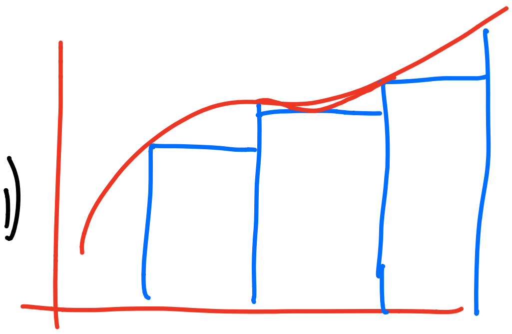
$$s(f, \{x_0, x_1, \dots, x_n\}) = \sum_{k=1}^{n} \underbrace{|I_k|}_{} \cdot m_k$$

and the upper sum

$$S(f, \{x_0, x_1, \dots x_n\}) = \sum_{k=1}^{n} |I_k|\, M_k$$

Now define

$$J_* := \sup_{\text{partitions}} \left(s\left(f, \text{partition}\right)\right)$$

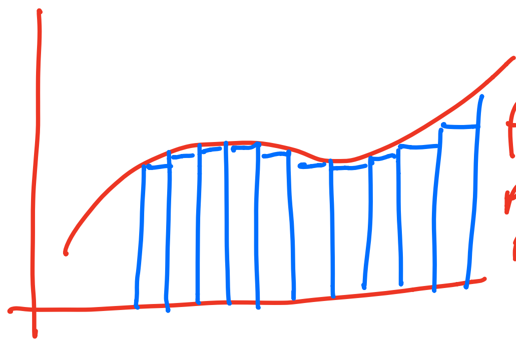


coarse partition from below

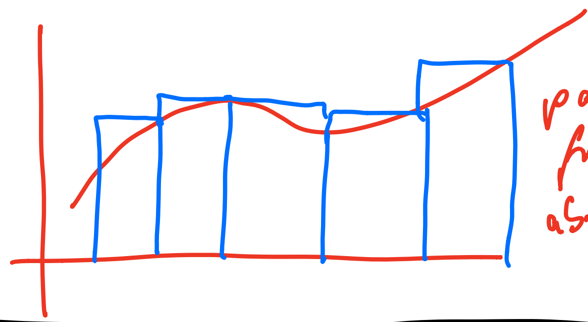
$$J^* := \inf_{\text{partitions}} \left(S\left(f, \text{partition}\right)\right)$$

We call f <u>Riemann-integrable</u> if $J_* = J^*$. Then we denote

$$J_* = J^* =: \int_a^b f(t)\, dt .$$



finer partition from below



partition from above

Theorem:

- $f : [a,b] \to \mathbb{R}$ monotone $\Rightarrow$ integrable

  (i.e. $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$)

- $f : [a,b] \to \mathbb{R}$ continuous $\Rightarrow$ integrable

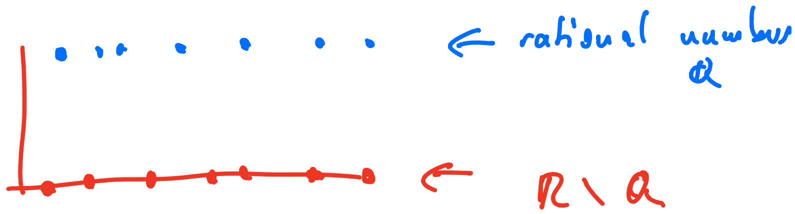
  (even true if f is continuous everywhere except at finitely many point)

Shortcomings:

- Many functions are <u>not</u> integrable.

  For example: $f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & \text{otherwise} \end{cases} = \mathbb{1}_{\mathbb{Q}}$

  Because:



  For any interval $I_k = [x_k, x_{k+1}]$,

  $M_k = +1$

  $m_k = 0$

  Thus $\underset{\underset{|b-a| \cdot 0}{\shortparallel}}{J_*} < \underset{\underset{|b-a| \cdot 1}{\shortparallel}}{J^*}$

- One cannot prove theorems about exchanging "integral" with "lim":

  $$\lim_{n \to \infty} \int f_n \, dt \overset{?}{=} \int \lim f_n \, dt$$

- Hard to extend to "other spaces".

Fundamental theorem of calculus

corrected a typo

Theorem I: $f:[a,b]\to\mathbb{R}$ (Riemann)-integrable and continuous at $\xi\in[a,b]$. Let $c\in[a,b]$. Then the function

$$F(x) := \int_c^x f(t)\,dt$$

is differentiable at ξ and $F'(\xi)=f(\xi)$.

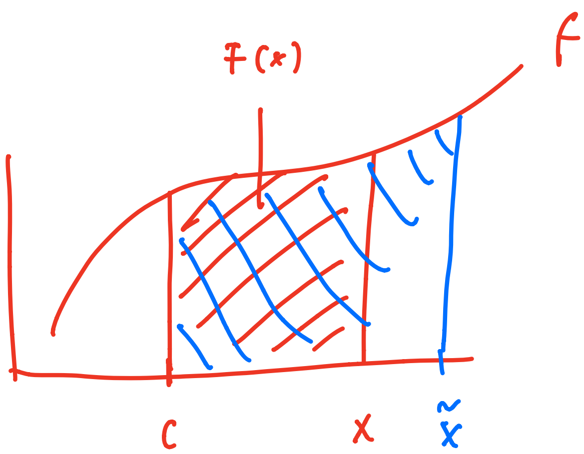
F cont. differentiable

If $f\in\mathcal{C}([a,b])$, then $F\in\mathcal{C}^1([a,b])$ and $F'(x)=f(x)$ for all $x\in[a,b]$.

f cont.

Theorem II: $F:[a,b]\to\mathbb{R}$ continuously differentiable, then

$$\int_a^b F'(t)\,dt = F(b)-F(a).$$



Informal, algebraic version:

The integral operator $I:\mathcal{C}[a,b]\to\mathcal{C}^1_{"c"}([a,b])$

with $\mathcal{C}^1_{"c"}[a,b] := \{f\in\mathcal{C}^1[a,b] : f(c)=0\}$

is an isomorphism (linear, bijective) and its inverse is the differential operator.

Proof I: Need to prove that F is diff. at ξ.

Consider $A(h) := \dfrac{F(\xi+h) - F(\xi)}{h}$

$$= \frac{1}{h}\left(\int_c^{\xi+h} f(t)\,dt - \int_c^{\xi} f(t)\,dt \right)$$

$$= \frac{1}{h} \int_{\xi}^{\xi+h} f(t)\,dt$$

Want to prove: converges to $f(\xi)$ as $h \to 0$

Want to prove:

$$\underbrace{A(h) - f(\xi)}_{} \xrightarrow{!} 0$$

$$= \frac{1}{h}\int_{\xi}^{\xi+h} f(t)\,dt - f(\xi)$$

$$= \frac{1}{h}\int_{\xi}^{\xi+h} \underbrace{f(\xi)}_{\text{does not depend on } t}\,dt = \frac{1}{h}\underbrace{(\xi+h-\xi)}_{\text{length of interval over which we integrate}} \cdot \underbrace{f(\xi)}_{\text{constant in the integral}}$$

$$= \frac{1}{h}\int_{\xi}^{\xi+h} f(t)\,dt - \frac{1}{h}\int_{\xi}^{\xi+h} f(\xi)\,dt$$

$$= \frac{1}{h}\int_{\xi}^{\xi+h} \underbrace{f(t) - f(\xi)}_{\text{Intuition: small due to continuity of } f \text{ at } \xi}\,dt$$

Formally: given $\varepsilon > 0$ we can find $h > 0$ such that
$f(t) - f(\xi) < \varepsilon \quad \forall\, t \in [\xi, \xi+h]$.

Then:

$$\frac{1}{h}\int_{\xi}^{\xi+h} f(t)-f(\xi)\,dt \leq \frac{1}{h}\int_{\xi}^{\xi+h} |f(t)-f(\xi)|\,dt$$

$$\leq \frac{1}{h}\int_{\xi}^{\xi+h} \varepsilon\,dt = \frac{1}{h}\cdot\varepsilon\int_{\xi}^{\xi+h} 1\,dt = \frac{1}{h}\cdot\varepsilon\cdot h = \varepsilon.$$

■ Theorem I

<u>Proof II</u> :

Know that F' continuous. Thus by Theorem I the function

$$G(x) := \int_a^x F'(t)\,dt \quad \text{is differentiable and}$$

(i) $G(a) = 0$ (by def. of G)

(ii) $G'(x) = F'(x)$ on $[a,b]$ (by Theorem I).

Consider $H(x) := F(x) - G(x)$.

By (ii) we know that $H'(x) = F'(x) - G'(x) = 0$ for all x

Hence, H is a constant function.

We know that $H(a) = F(a) - \underbrace{G(a)}_{=0 \ (i)} = F(a)$, thus

(iii) $H(x) \equiv F(a)$ → means: "constant"

Consider $x = b$.

$$F(a) \overset{iii}{=} H(b) \overset{def}{=} F(b) - G(b) \overset{def}{=}$$

$$= F(b) - \int_a^b F'(t)\,dt$$

$$\Rightarrow \int_a^b F'(t)\,dt = F(b) - F(a).$$

$\blacksquare$ Th. II

Power series

Def A series of the form $p(x) := \sum_{n=0}^{\infty} a_n x^n$ is called a power series.

Theorem (Radius of convergence)

For every power series $p(x) = \sum_{n=0}^{\infty} a_n x^n$ there exists a constant r, $0 \le r \le \infty$, called the radius of convergence such that

- The series converges (absolutely) for all x with $|x| < r$ (meaning that $\sum_{n=0}^{\infty} a_n |x|^n$ converges, meaning that the sequence of partial sums $p_N(x) := \sum_{n=0}^{N} a_n |x|^n$ converges "in the usual sense" as $N \to \infty$)

  ⚠ It is unclear what happens for $|x| = r$

- If $|x| < r$, the series even converges uniformly.

The radius of convergence only depends on the $(a_n)_n$ and can be computed by various formulas:

$$\left.\begin{array}{l} \bullet\ r = \dfrac{1}{L} \text{ where } L = \limsup\limits_{n\to\infty} (|a_n|)^{1/n} \\[2ex] \bullet\ r = \lim\limits_{n\to\infty} \left| \dfrac{a_n}{a_{n+1}} \right| \end{array}\right\} \text{ if exists}$$

Examples:

- $p(x) = \sum_{n=0}^{\infty} \underbrace{n^c}_{a_n} \cdot x^n$ for some constant c

$$r = \lim \left| \frac{a_n}{a_{n+1}} \right| = \lim \frac{n^c}{(n+1)^c} = \lim \left(\frac{n}{n+1} \right)^c = 1$$

Case $c = -1$: $\sum \frac{1}{n} x^n$ has conv. radius $r=1$

- For $x = +1$ the series diverges because
$$\sum \frac{1}{n} x^n = \sum \frac{1}{n} 1^n = \sum \frac{1}{n} \to \infty$$
- For $x = -1$ it converges.
- For $x > 1$ it diverges.

no general statement for $|x| = r$

Case $c = 0$: $\sum n^c x^n = \sum x^n$ diverges for $|x| = r$ (both $x = -1$ and $x = +1$).

- Exponential series:
$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad \text{has } r = \infty$$
because $$\left| \frac{a_n}{a_{n+1}} \right| = \frac{1/n!}{1/(n+1)!} = \frac{(n+1)!}{n!} = n+1 \to \infty$$

- $\sum_{n=0}^{\infty} n!\, x^n$ has $r = 0$: $\left| \frac{a_n}{a_{n+1}} \right| = \frac{n!}{(n+1)!} = \frac{1}{n+1} \to 0.$

From power series to Taylor series

Observation: Given power series $\boxed{f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n.}$

Let's take its derivative:

$$f'(x) = \left(a_0 + a_1(x-a) + a_2(x-a)^2 + a_3(x-a)^3 + \dots \right)'$$

$$= a_1 + 2a_2(x-a) + 3a_3(x-a)^2 + \dots$$

$$= \sum_{\underline{n=1}}^{\infty} \underline{n \cdot a_n (x-a)^{n-1}}$$

$$f''(x) = \dots$$

$$\vdots$$

$$f^{(k)}(x) = \sum_{\underline{n=k}}^{\infty} a_n \big(n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)\big) (x-a)^{n-k}$$

In particular, we have

$$f^{(k)}(a) = a_k k! \quad \text{or, stated otherwise} \quad \boxed{a_k = \frac{f^{(k)}(a)}{k!}}$$

<u>Theorem</u>: Let $f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n$ with $r > 0$. Then for x with $|x-a| < r$ we have

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n$$

Intuition: start with a power series that converges. Then we have the neat formula that expresses the coeff. in terms of derivatives.

<u>Question</u> Does it work the other way round? That is, given any function (possibly with nice assumptions), can we simply build the series $\sum \frac{f^{(n)}}{n!}(x-a)^n$ and "hope" that it converges to the function?

$$\stackrel{?}{=} f(x) \quad ???$$

Taylor series

<u>Theorem</u>: $J \subset \mathbb{R}$ open interval, $f : J \to \mathbb{R}$, $f \in \mathcal{C}^{n+1}([a,b])$, $a, x \in J$. Define

$$T_n(x,a) := \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k$$

Taylor series up to degree n

$$R_n(x,a) := \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)\,dt$$

Remainder term

Then $f(x) = T_n(x,a) + R_n(x,a)$

<u>Proof</u> follows from Fundamental theorem, by induction on n.

<u>Base case $n=0$:</u> need to prove

$$f(x) = f(a) + \int_a^x f'(t)\,dt \quad \hat{=} \text{ Fundam. Theorem}$$

<u>Inductive step $n \to n+1$:</u>

- Consider $F(t) = \dfrac{(x-t)^{n+1}}{(n+1)!} f^{(n+1)}(t)$
- Take its derivative
- Integrate and exploit fundamental theorem

Theorem (Taylor with Lagrange remainder)

$f \in \mathcal{C}^{n+1}(J)$, $a, x \in J$. Then there exists some $\xi \in J$ such that

$$R_n(x,a) = \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(\xi)$$

Proof Let $J = [a,b]$.

- Consider two functions $F, G \in \mathcal{C}^{n+1}([a,b])$. Assume that $\boxed{F(a) = G(a) = 0, \text{ and } G' \neq 0 \text{ on } [a,b].}$ $(*)$

  Now:
  $$\frac{F(b) - \underbrace{F(a)}_{=0}}{G(b) - \underbrace{G(a)}_{=0}} = \frac{F(b)}{G(b)} = \frac{F'(\xi)}{G'(\xi)} \quad \text{for some } \xi \in [a,b]$$

  Assume that F' and G' also satisfy $(*)$. We can iterate ...

  We would obtain
  $$\# \quad \frac{F(b)}{G(b)} = \frac{F^{(n+1)}(\xi)}{G^{(n+1)}(\xi)} \quad \text{for some } \xi \in [a,b]$$

- Now chose $F(x) := f(x) - T_n(x,a) = R_n(x,a)$

  $G(x) := (x-a)^{n+1}$

  - For all k in $0 \leq k \leq n$ we have by construction that

    $f^{(k)}(a) = T_n^{(k)}(a)$, so in particular

    $F^{(k)}(a) = 0$, and we have $G^{(k)}(a) = 0$.

- For $n+1$ we now have

$$F^{(n+1)}(x) = f^{(n+1)}(x), \quad G^{(n+1)}(x) = (n+1)!$$

By (#) we obtain

$$F(x) = R_n(x,a) \overset{\#}{=} G(x) \cdot \frac{F^{(n+1)}(\xi)}{G^{(n+1)}(\xi)} =$$

$$= \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(\xi).$$

Theorem $f \in \mathcal{C}^\infty(J)$, $x, a \in J$. Define

$$T(x) := \lim_{n\to\infty} T_n(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Then we have $f(x) = T(x)$ if $R_n(x,a) \xrightarrow{n\to\infty} 0$.

For example, this is the case if there exist constants $\alpha, C > 0$ such that

$$|f^{(n)}(t)| \leq \alpha \cdot C^n \quad \forall t \in J, \ \forall n \in \mathbb{N}.$$

Follows directly from the Lagrangian remainder.

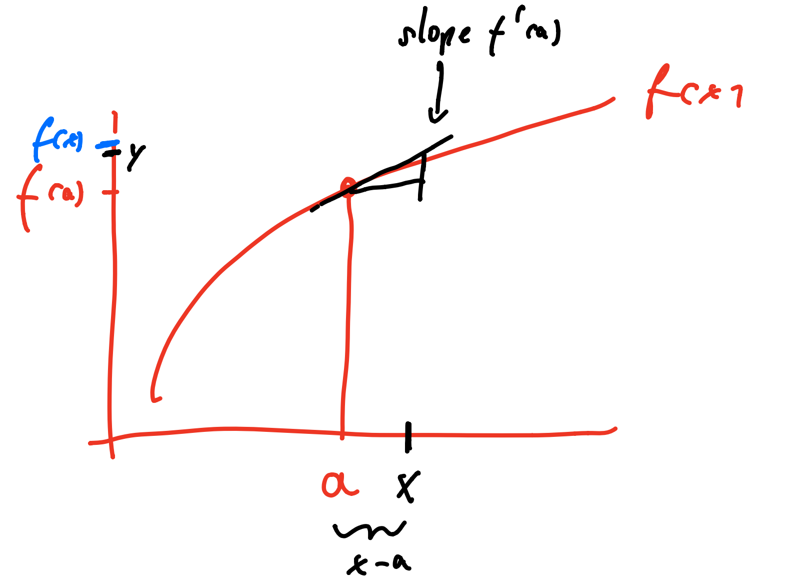
Examples :

- Exponential series:

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

power series with $r = \infty$,
exp always coincides with its Taylor series.

Intuition for Taylor series



$$f(x) \approx \underbrace{f(a) + f'(a)(x-a)}_{y} + \dots$$

- $f(x) = \log(1+x)$, Taylor series about $a = 0$

  Can prove: Convergence radius of Taylor series is $r = 1$

  For x outside of $]-1, 1[$ Taylor series does not make sense at all.

- $$f(x) = \begin{cases} \exp(-1/x^2) & \text{if } x \neq 0 \\ 0 & x = 0 \end{cases}$$

  Has the funny property that $\forall n \in \mathbb{N}: f^{(n)}(0) = 0$

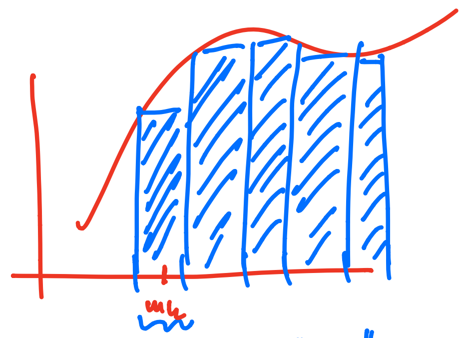
  Consider the Taylor series derived about $a = 0$.

All terms will be 0, so $\forall n$: $T_n(x) = 0$, $r = \infty$

but of course f is not $\equiv 0$, so we get

$\forall x \neq 0$, $T_n(x) \neq f(x)$.

σ - Algebra



Intervals: "vol" $[x_k, x_{k+1}]$

$= x_{k+1} - x_k$

$$\int_a^b f \, dt \approx \sum_k \underline{\text{vol}(I_k)} \, f(m_k)$$

Def X non-empty set. A <u>σ-algebra on X</u> is a non-empty collection $\mathcal{F}$ of subsets of X such that

(i) $\mathcal{F}$ is closed under taking complements:

$$A \in \mathcal{F} \Rightarrow X \setminus A \in \mathcal{F}$$

(ii) $\mathcal{F}$ is closed under countable unions:

$$A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$$

Def A <u>measurable space</u> consist of a set X and a σ-algebra $\mathcal{F}$ over X. Notation: $(X, \mathcal{F})$. The sets in $\mathcal{F}$ are called <u>measurable</u>.

Important σ-algebras

- Trivial σ-algebras: Given X, we can define two (pretty useless) σ-algebras:
  - $\mathcal{F}_1 = \{\emptyset, X\}$
  - $\mathcal{F}_2 = \mathcal{P}(X) = \{A \subset X\}$

- Given X, let $\mathcal{G}$ be any collection of subsets. The <u>σ-algebra generated by $\mathcal{G}$</u> is the smallest σ-algebra $\mathcal{F}$ such that $\mathcal{G} \subset \mathcal{F}$. Notation: $\sigma(\mathcal{G})$

  (Remark: Existence can be proved easily by an explicit construction:
  $$\sigma(\mathcal{G}) = \bigcap \{\Sigma : \Sigma \text{ is a } \sigma\text{-Algebra that contains } \mathcal{G}\})$$

- Consider a metric space (X, d), and let $\mathcal{G}$ be the collection of open subsets of X. Then the <u>Borel-σ-algebra</u> is defined as $\sigma(\mathcal{G})$.

Measures

Def Given a measurable space $(X, \mathcal{F})$, a measure is a map $\mu: \mathcal{F} \to [0, \infty]$ such that

(i) $\mu(\emptyset) = 0$.

(ii) For a countable collection of disjoint subsets $(S_i)_{i \in \mathbb{N}}$ with $S_i \in \mathcal{F}$ we have

$$\mu\Big(\bigcup_{i \in \mathbb{N}} S_i\Big) = \sum_{i \in \mathbb{N}} \mu(S_i).$$

⚠ measure is a function on $\mathcal{F}$, not on X!

Examples:

Discrete measure: Given a finite (or countable) space $X = \{x_1, x_2, x_3, \dots\}$, define the σ-algebra $\mathcal{F}$ to be the set of all subsets of X. Consider a sequence $(m_i)_{i \in \mathbb{N}}$ such that $\sum_{i=1}^{\infty} m_i$ is finite.

(Example: $m_i = \frac{1}{i^2}$)

Want to define $\mu: \mathcal{F} \to \mathbb{R}$. Proceed as follows:

$$\mu(\{x_i\}) := m_i$$ define μ on all "elementary" sets

For all other sets $A \in \mathcal{F}$ you can now deduce the measure (due to countability) by

$$\mu(A) = \sum_{x_i \in A} \mu(\{x_i\}) = \sum_{x_i \in A} m_i$$

<u>Example 2</u>: A funny measure on $\mathbb{R}$:

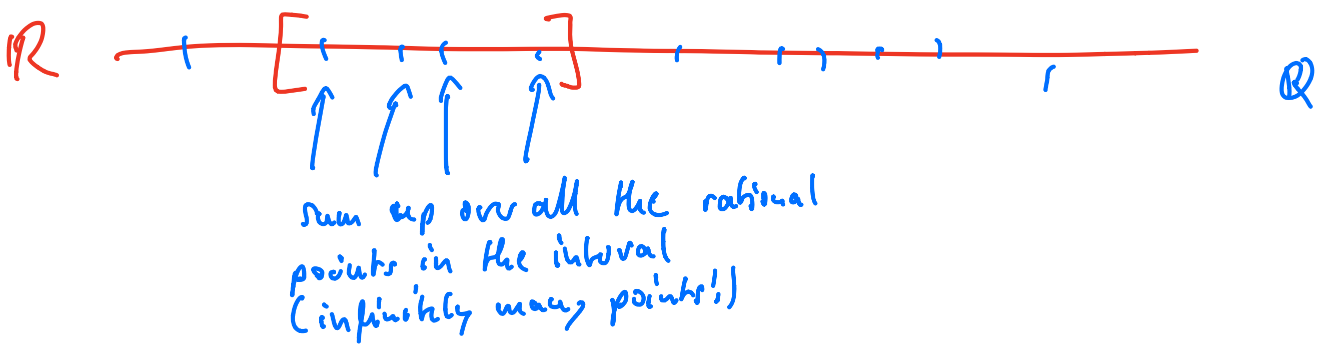
Let $\mathcal{F}$ be the Borel-σ-Algebra on $\mathbb{R}$. Want to define a measure that just assigns mass to rational numbers.

Let $\underbrace{q_1, q_2, q_3, \dots}_{\mathbb{Q}}$ be all rational numbers.

Consider $(m_i)_{i \in \mathbb{N}}$ as before: $m_i = \frac{1}{i^2}$.

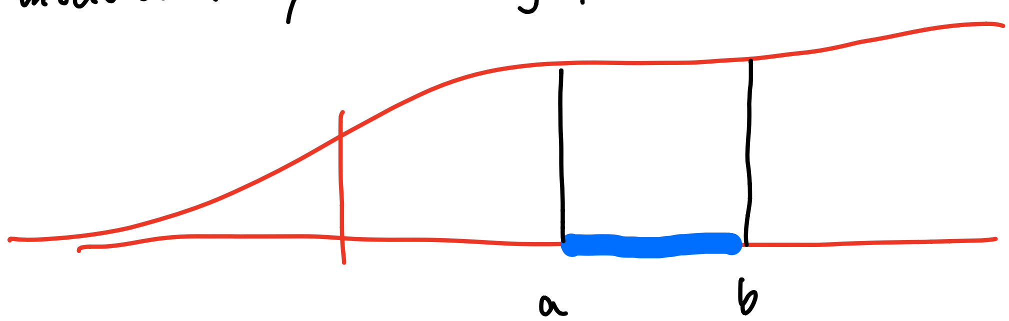
$$\mu(\{q_i\}) = m_i$$

$$\mu([a,b]) = \sum_{q_i \in [a,b]} \mu(\{q_i\}) = \sum_{q_i \in [a,b]} m_i$$



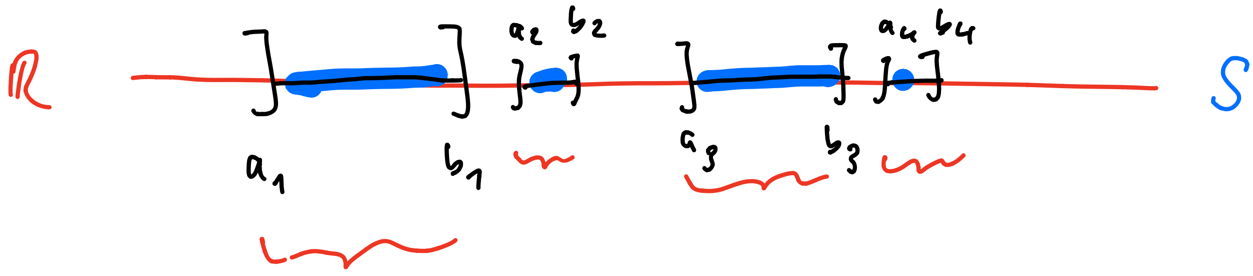
<u>Example</u>: A more useful class of measures on $\mathbb{R}$

$X = \mathbb{R}$, $\mathcal{F}$ Borel-σ-algebra. Let $F: \mathbb{R} \to \mathbb{R}$ be monotonically increasing, continuous.



Define a measure (!) μ_F on $(\mathbb{R}, \mathcal{F})$ by setting

$$\mu_F(S) = \inf\left\{ \sum_{j=1}^{\infty} F(b_j) - F(a_j) \;\middle|\; S \subset \bigcup_{j=1}^{\infty}]a_j, b_j] \right\}$$



$\rightsquigarrow F(b_1) - F(a_1)$

- Cover S by intervals
- To each interval we assign "elementary volume" $F(b) - F(a)$
- Take "best" covering

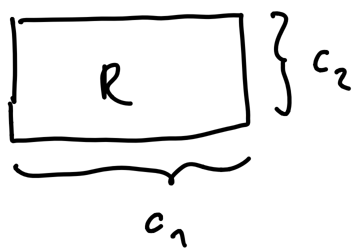
⚠ Need to prove: this is a measure!

<u>Def</u> A measurable space $(X, \mathcal{F})$ endowed with a measure μ is called a <u>measure space</u> $(X, \mathcal{F}, \mu)$.

A subset $N \in \mathcal{F}$ is called a <u>null set</u> if $\mu(N) = 0$.
We say that a property holds ~~at~~ <u>almost everywhere</u> if it holds for all $x \in X$ except for x in a null set N.
(in probability theory, we say "<u>almost surely</u>").

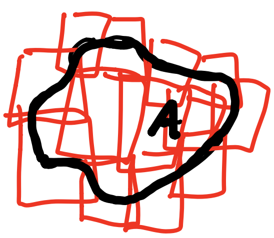
The Lebesgue measure on $\mathbb{R}^n$

Want to construct a measure on $\mathbb{R}^n$. Want that rectangles of the form $[a_1, b_1[\times [a_2, b_2[\times \ldots \times [a_n, b_n[$ have the "natural volume" given by $\prod_{i=1}^{n} (b_i - a_i)$



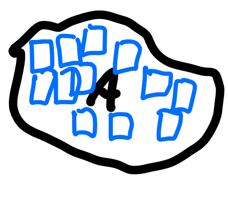
$\rightsquigarrow \mathrm{vol}(R) \overset{!}{=} c_1 \cdot c_2$

First approaches (Jordan, Riemann) attempted the following:



"Outer approximation":

$A \subset \bigcup_{i=1}^{n} \text{rectangles}_i$



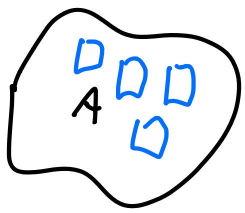
"Inner approximation:"

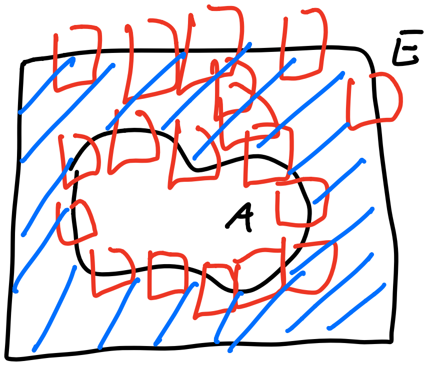
$\bigcup_{i=1}^{n} \text{rect.}_i \subset A$

A would be called "measurable" if outer and inner approximation "converge".

Now: generalization of this approach

- Allow for countable coverings
- Replace inner approximation by an outer approx. of the complement:





outer approx. of $\underline{E \setminus A}$

$$\mu(E) = \underbrace{\mu(E \setminus A)} + \mu(A)$$

$$\mu(A) = \mu(E) - \underbrace{\mu(E \setminus A)}$$

- Need σ-algebra as underlying structure.

Outer Lebesgue measure

Set the "natural volume" of rectangles:

$$R = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n] \subset \mathbb{R}^n$$

$$|R| := \prod_{i=1}^{n} (b_i - a_i)$$

Definition of outer Lebesgue measure:

Let $A \subset \mathbb{R}^n$ be arbitrary. We define

$$\lambda(A) := \inf \left\{ \sum_{i=1}^{\infty} |R_i| \;\middle|\; A \subset \bigcup_{i=1}^{\infty} R_i \,,\ R_i \text{ rectangle} \right\}$$

We cover A by a countable union of rectangles, then take inf.

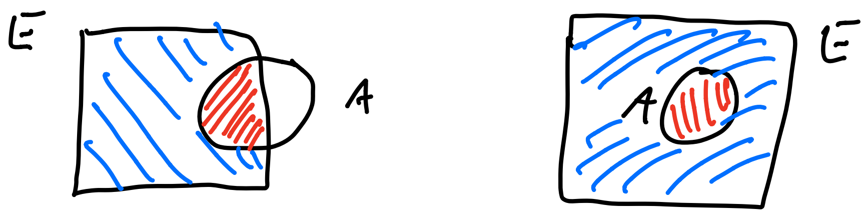
Observe: $\lambda(A) \in \mathbb{R} \cup \{\infty\}$.

Want to make this into a measure. Problem: if we use $\mathcal{P}(\mathbb{R}^n)$ as σ-algebra, we run into contradictions.

Need to restrict ourselves to a smaller σ-algebra...

Definition: We say that a set $A \subset \mathbb{R}^n$ is measurable if for all $E \subset \mathbb{R}^n$

$$\lambda(E) = \lambda(E \cap A) + \lambda(E \setminus A)$$



Denote by $\mathcal{L}$ all measurable subsets of $\mathbb{R}^n$.

Theorem The set $\mathcal{L}$ forms a σ-algebra on $\mathbb{R}^n$. The outer measure λ (defined above) is in fact a measure on $(\mathbb{R}^n, \mathcal{L})$. On rectangles it coincides with the "natural volume".

Examples:

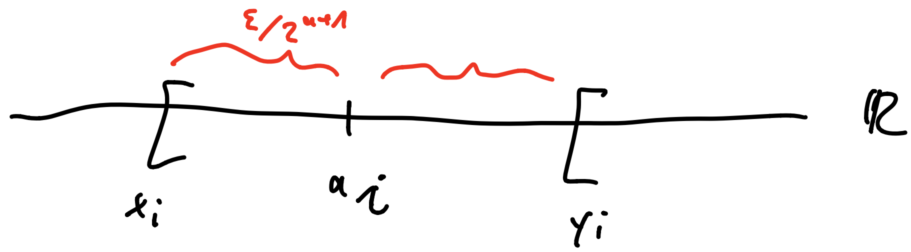
- $\lambda(\{x\}) = 0$
- $\lambda(\mathbb{R}) = \infty$

- $A \subset \mathbb{R}$ countable. The $\lambda(A) = 0$. In particular, $\mathbb{Q}$ is measurable and has $\lambda(\mathbb{Q}) = 0$.

typo corrected

<u>Proof sketch</u>: For $\varepsilon > 0$, define for all $a_i \in A$ the interval $[x_i, y_i[$ such that

$$x_i = a_i - \frac{\varepsilon}{2^{i+1}}, \quad y_i = a_i + \frac{\varepsilon}{2^{i+1}}$$



$$A \subset \bigcup_{i=1}^{\infty} [x_i, y_i[$$

$$\Rightarrow \lambda(A) \leq \sum_{i=1}^{\infty} \lambda([x_i, y_i[)$$

$$= \sum_{i=1}^{\infty} \frac{\varepsilon}{2^{i+1}} = \varepsilon$$

Taking the inf. over all coverings shows that $\lambda(A) = 0$.

<u>Comparing $\mathcal{L}$ (σ-alg. of Lebesgue measurable sets) with the Borel-σ-algebra $\mathcal{B}$</u>

(1) $\mathcal{B} \subset \mathcal{L}$:

- open intervals are measurable, thus in $\mathcal{L}$
- any open set A in $\mathbb{R}^n$ can be written as a countable union of open intervals: $A \subset \bigcup_{i=1}^{\infty} I_i$, I_i open intervals.

(ii) For every Lebesgue-measurable set L there exist a set $B \in \mathcal{B}$ and $N \in \mathcal{L}$ with $\lambda(N) = 0$ such that

$L = B \cup N$.

Summary: $\mathcal{L} \approx \mathcal{B}$ (up to sets of measure 0).

A non-measurable set

Consider $[0,1[$. Define an equivalence relation on $[0,1[$ as follows:

$$x \sim y \;:\Longleftrightarrow\; x - y \in \mathbb{Q}$$

$\frac{\pi}{4}$, $\frac{\pi}{4} + \frac{1}{2}$, $\frac{\pi}{4} + \frac{799}{800}$ would be equivalent

Consider the equivalence classes

$$\frac{\pi}{4} + \mathbb{Q} = \left\{ \frac{\pi}{4} + q \;\middle|\; q \in \mathbb{Q} \right\}$$

$$\frac{e}{3} + \mathbb{Q}$$

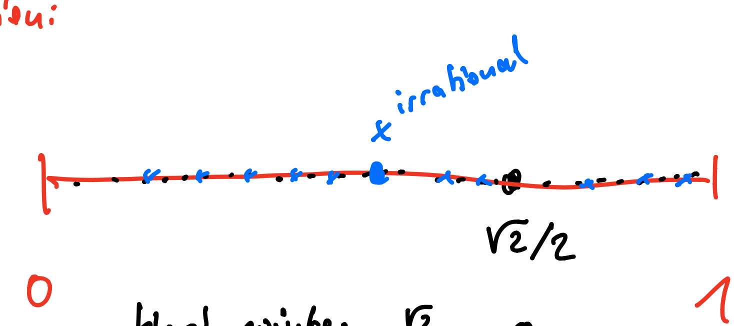
$$\frac{\sqrt{2}}{2} + \mathbb{Q}$$

$$\vdots$$

We pick a representative of each of the classes, and denote by N the set of all such representatives.

$\boxed{\text{Want to prove: } N \text{ is \underline{not} Lebesgue-measurable.}}$

Intuition:



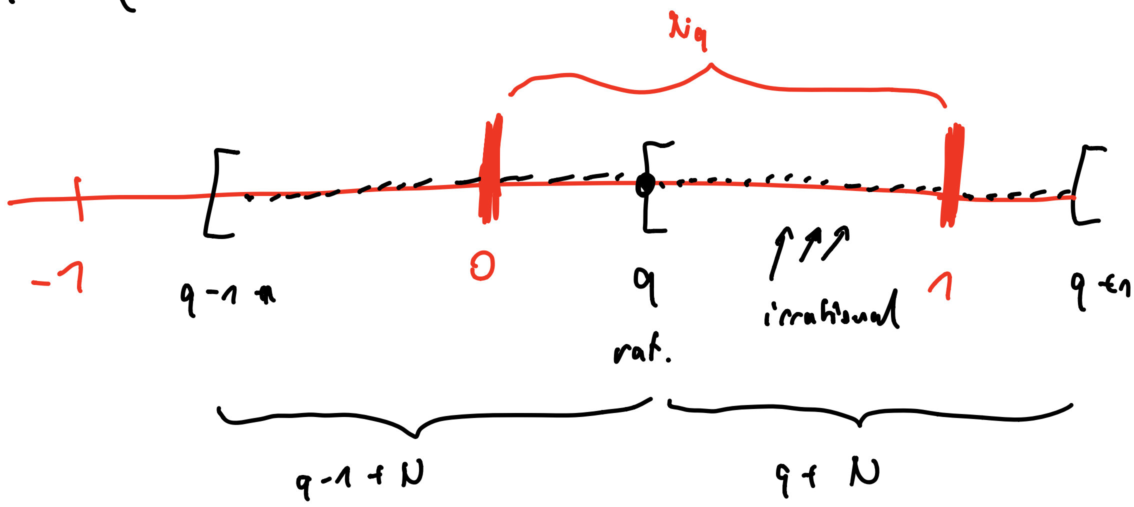
black points: $\frac{\sqrt{2}}{2} + \mathbb{Q}$

blue crosses: $x + \mathbb{Q}$

Proof by contradiction:

Assume N is measurable. We now construct the following sets: For $q \in [0,1[$

$$N_q := \big((q+N) \cup (q-1+N)\big) \cap [0,1[$$



- if N is measurable, then $q+N$ is measurable $\forall q \in [0,1[$ and $\lambda(N_q) = \lambda(N)$

- $$[0,1[\;=\; \bigcup_{q \in [0,1] \cap \mathbb{Q}} N_q$$

- $N_q \cap N_p \neq \emptyset \Rightarrow N_p = N_q$

  Consequently, $\bigcup N_q$ is disjoint.

- σ-additivity:

$$\underbrace{\lambda([0,1[)}_{1} = \lambda\Big(\bigcup_q N_q\Big) = \sum_{q \in [0,1] \cap \mathbb{Q}} \underbrace{\lambda(N_q)}_{\lambda(N)}$$

- Could be that $\lambda(N_q) = 0$. But then

$$\sum_q \lambda(N_q) = 0 \quad \lightning$$

- Could be that $\lambda(N_q) > 0$. But then

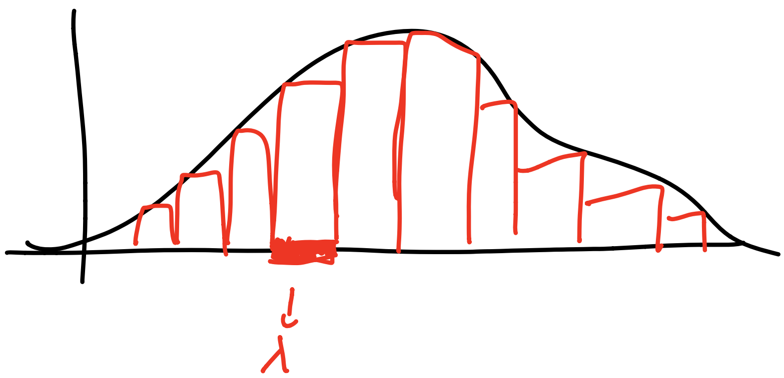
$$\sum_q \lambda(N_q) = \infty$$

$$\lightning$$

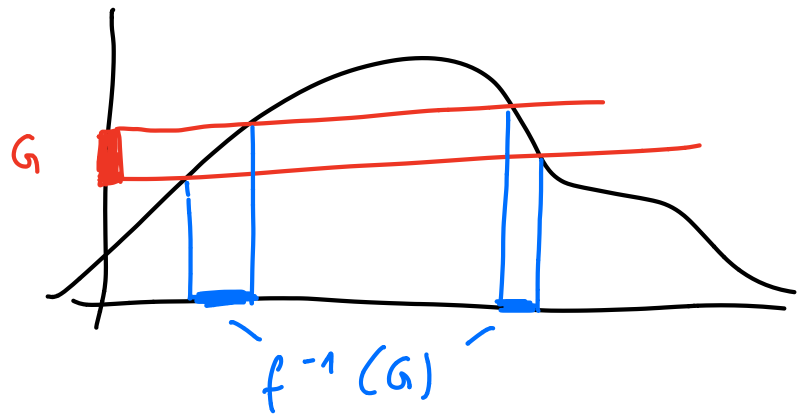
The Lebesgue integral on $\mathbb{R}^n$

Intuition:

Riemann:



Lebesgue:



<u>Def</u> A function $f : (X, \mathcal{F}) \to (Y, \mathcal{G})$ between two measurable spaces is called <u>measurable</u> if pre-images of measurable sets are measurable:

$$\forall G \in \mathcal{G} : \underbrace{f^{-1}(G)}_{=:\{x \in X \mid f(x) \in G\}} \in \mathcal{F}$$

Def $\phi : \mathbb{R}^n \to \mathbb{R}$ is called a *simple function* if there exist measurable sets $S_i \subset \mathbb{R}^n$, $a_i \in \mathbb{R}$ such that

$$\phi = \sum_{i=1}^{n} a_i \underbrace{\mathbb{1}_{\{S_i\}}}_{S_i = \phi^{-1}(a_i)}$$

For such a function we can define its Lebesgue integral as

$$\int \phi \, d\lambda := \sum_{i=1}^{n} a_i \, \lambda(S_i)$$

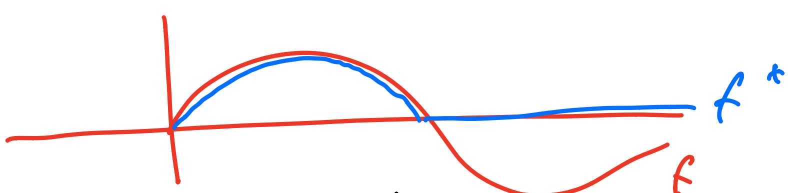
For a function $f^+ : \mathbb{R}^n \to [0, \infty[$ we define its Lebesgue integral

$$\int f^+ d\lambda = \sup \left\{ \int \phi \, d\lambda \;\middle|\; \phi \leq f, \ \phi \text{ simple} \right\}$$

(might be ∞)

For a general function $f : \mathbb{R}^n \to \mathbb{R}$ we split the function into positive and neg. part: $f = f^+ - f^-$

where $f^+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{otherwise} \end{cases}$



Note: f^+, f^- are measurable if f is measurable.

If both f^+ and f^- satisfy $\int f^+ d\lambda < \infty$, $\int f^- d\lambda < \infty$, then we call f integrable and define

$$\int f \, d\lambda = \int f^+ d\lambda - \int f^- d\lambda .$$

Much more powerful notion than Riemann integral.

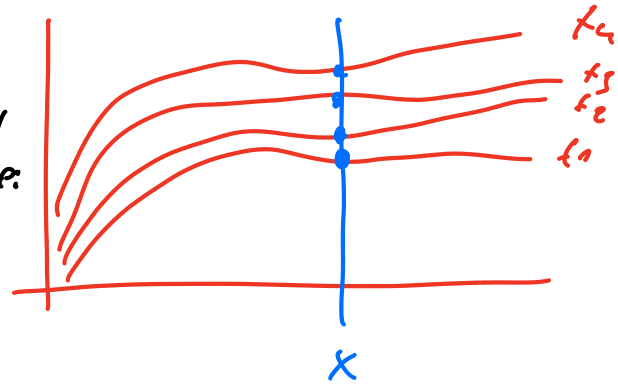
Example: $\int 1\!\!1_{\mathbb{Q}} \, d\lambda = 1 \cdot \lambda(\mathbb{Q}) = 0$

Two important theorems

Theorem (monotone convergence):

Consider a sequence of functions $f_k : \mathbb{R}^n \to [0, \infty[$ that is pointwise non-decreasing:

$\forall x \in \mathbb{R}^n : f_{n+1}(x) \geq f_n(x)$.



Assume that all f_k are measurable, and that the pointwise limit exists:

$\forall x : \lim f_k(x) =: f(x)$

Then:

$$\int f(x) \, dx = \lim_{k \to \infty} \int f_k(x) \, dx$$

$$\shortparallel$$

$$\int \lim_{k \to \infty} f_k(x) \, dx$$

Theorem (dominated convergence):

$f_n : B \to \mathbb{R}$, $|f_n(x)| \leq g(x)$ on B, $g(x)$ is integrable. Assume that the pointwise limit exists: $\forall x \in B$:

$f(x) := \lim\limits_{n \to \infty} f_n(x)$. Then:

$$\int f(x)\,dx = \lim_{n\to\infty} \int f_n(x)\,dx$$

Partial derivatives on $\mathbb{R}^n$

Consider $f : \mathbb{R}^n \to \mathbb{R}$

$\mathbb{R}^n \ni x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\quad f(x) = x_1^2 + x_2^2 \cdot x_1$, $\quad f : \mathbb{R}^2 \to \mathbb{R}$

<u>Def</u> f is called partially differentiable with resp. to variable x_j at point $\xi \in \mathbb{R}^n$ if the function

$$x_j \mapsto g(x_j) := f(\xi_1, \xi_2, \dots, \xi_{j-1}, x_j, \xi_{j+1}, \dots, \xi_n)$$

$g : \mathbb{R} \to \mathbb{R}$

is differentiable at $\xi_j \in \mathbb{R}$.

Notation:
$$\frac{\partial f}{\partial x_j}(\xi) = \lim_{h \to 0} \frac{f(\xi + e_j \cdot h) - f(\xi)}{h}$$

e_j: j-th unit vector $= \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow j$, $\quad h \in \mathbb{R}$

If all partial derivatives exist, then the vector of all partial derivatives is called the <u>gradient</u>:

$$\operatorname{grad}(f)(\xi) = \nabla f(\xi) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\xi) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\xi) \end{pmatrix} \in \mathbb{R}^n$$

If $f: \mathbb{R}^n \to \mathbb{R}^m$, we decompose f into its m component functions $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$. We define the Jacobian matrix

grad f_1

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

⚠ Even if all partial derivatives exist at ξ, we do not know whether f is continuous at ξ!

Example: $f: \mathbb{R}^2 \to \mathbb{R}$,

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

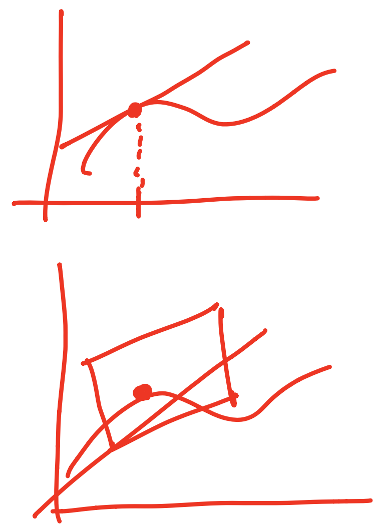
For $(x,y) \neq (0,0)$

$$\operatorname{grad} f(x,y) = \left(y \cdot \frac{y^2 - x^2}{(x^2+y^2)^2} \,,\; x \cdot \frac{x^2 - y^2}{(x^2+y^2)^2} \right)$$

$$\operatorname{grad} f(0,0) = 0 \quad \text{because} \quad \begin{aligned} f(x,0) &= 0 \quad \forall x \\ f(0,y) &= 0 \quad \forall y \end{aligned}$$

but f is not continuous at 0.

Total derivative



$f: \mathbb{R}^n \to \mathbb{R}^m$, $\xi \in U$.

f is differentiable at ξ if there exists a linear mapping $L: \mathbb{R}^n \to \mathbb{R}^m$ such that for $h \in \mathbb{R}^n$

$$f(\xi + h) - f(\xi) = L(h) + r(h)$$

with $\lim_{h \to 0} \frac{r(h)}{|h|} \to 0.$

Intuition: f is "locally linear"

Theorem $f: \mathbb{R}^n \to \mathbb{R}^e$ differentiable at ξ.

- Then f is continuous at ξ.
- The linear functional L coincides with the gradient:

$$f(\xi + h) - f(\xi) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(\xi) \cdot h_j + r(h)$$

$$= \langle \operatorname{grad} f(\xi), h \rangle + r(h)$$

If $f: \mathbb{R}^n \to \mathbb{R}^m$, it is differentiable iff all coordinate functions $f_1, \dots, f_m$ are differentiable. Then all partial derivatives exist and

$$L(h) = \left(\text{Jacobi matrix} \right) \cdot h$$

<u>Theorem</u>: If all partial derivatives exist and are all continuous, then f is differentiable.

⚠ If partial derivatives exist, but are not continuous, then f doesn't need to be differentiable.

Directional derivatives

Def Assume $f: \mathbb{R}^n \to \mathbb{R}$ is cont. differentiable, $v \in \mathbb{R}^n$ with $\|v\| = 1$. The directional derivative of f at ξ in direction of v is defined as

$$D_v f(\xi) = \lim_{t \to 0} \frac{f(\xi + \overbrace{t}^{\in \mathbb{R}} \cdot \overbrace{v}^{\in \mathbb{R}^n,\ \text{direction}}) - f(\xi)}{t}$$

Theorem: $f: \mathbb{R}^n \to \mathbb{R}$ differentiable in ξ. Then all the directional derivatives exist, and we can compute them by

$$D_v f(\xi) = (\operatorname{grad} f)^t \cdot \underbrace{v}_{\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}} = \sum_{i=1}^{n} \overbrace{v_i}^{\in \mathbb{R}} \cdot \overbrace{\frac{\partial f}{\partial x_i}}^{\text{partial der.}}(\xi)$$

The largest value of all directional derivatives is attained in direction

$$v = \frac{\operatorname{grad} f(\xi)}{\|\operatorname{grad} f(\xi)\|}$$

Higher order derivatives

Consider $f: \mathbb{R}^n \to \mathbb{R}$, assume it is differentiable, so all partial derivatives $\frac{\partial f}{\partial x_j}: \mathbb{R}^n \to \mathbb{R}$. If this function is differentiable, we can take its derivative:

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

These are called second order partial derivatives.

⚠ In general, we cannot change the order of derivatives:

$$\frac{\partial f^2}{\partial x_i \partial x_j} \neq \frac{\partial f^2}{\partial x_j x_i}$$

Example: $f(x,x) = \frac{x \cdot y^3}{x^2+y^2}$

$$\operatorname{grad} f(x,y) = \left(\frac{y^3(y^2-x^2)}{(x^2+y^2)^2}, \frac{xy^2(3x^2+y^2)}{(x^2+y^2)^2} \right)$$

Have: • $\frac{\partial f}{\partial x}(0,y) = y$ for all y

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = 1$$

- $\frac{\partial f}{\partial y}(x,0) = 0 \quad \forall$ all x

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = 0$$

<u>Def</u> We say that $f: \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, if all partial derivative $\frac{\partial f}{\partial x_i}$ exist and are continuous.

We say that f is twice continuously differentiable if f is is continuously differentiable and all partial derivatives $\frac{\partial f}{\partial x_i}$ are again continuously differentiable.

Analogously: <u>k times cont. differentiable</u>

Notation: $\mathcal{C}^k(\mathbb{R}^n, \mathbb{R}^m) = \{f: \mathbb{R}^n \to \mathbb{R}^m \mid k$ times cont. diff. $\}$

$\mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R}^m) = \{f: \mathbb{R}^n \to \mathbb{R}^m \mid \infty$ often cont. diff. $\}$

<u>Theorem</u> (Schwartz) Assume that f is twice continuously differentiable. Then we can exchange the order in which we take partial derivatives:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

Analogously: k times cont. diff. $\Rightarrow$ can exchange order of first k partial derivatives.

⚠ Caution: dimensions

$f : \mathbb{R}^n \to \mathbb{R}$ function

$\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ first derivative : n partial deriv. $\frac{\partial f}{\partial x_i}$

$Hf : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ second derivative : n^2 "partial derivatives" $\frac{\partial f}{\partial x_i \partial x_j}$

Def Hessian matrix

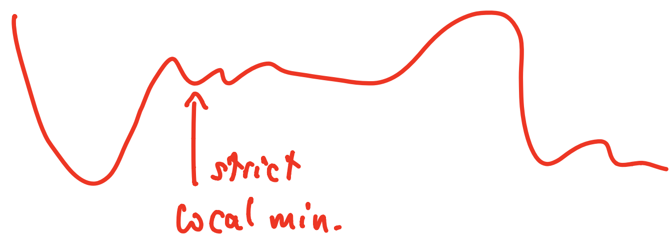
$f : \mathbb{R}^n \to \mathbb{R}$, then we define the Hessian of f at point x by

$$(Hf)_{ij}(x) := \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \qquad i,j = 1, \dots, n$$

Minima / maxima

Def $f: \mathbb{R}^n \to \mathbb{R}$ differentiable. If $\nabla f(x) = 0$ then we call x a <u>critical point</u>.

f has a local minimum at x_0 if there exists $\varepsilon > 0$ such that $\forall x \in B_\varepsilon(x_0) : f(x) \geq f(x_0)$



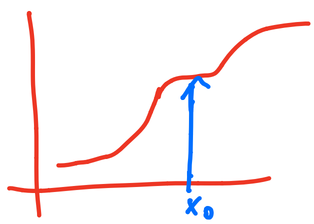
f has a <u>strict local minimum</u> at x_0 ...

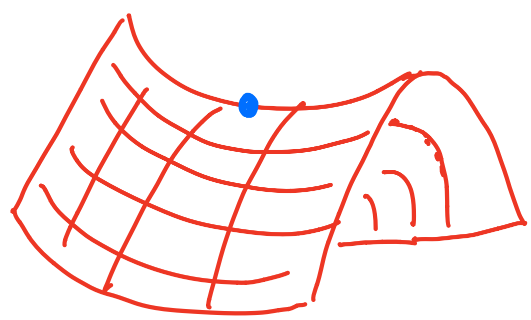
$\forall x \in B_\varepsilon(x_0) : f(x) > f(x_0)$

f has a <u>local maximum</u> (resp. strict local max) ...

$\forall x \in B_\varepsilon(x_0) : f(x) \overset{(<)}{\leq} f(x_0)$.

If f is diff. and x_0 is a critical point that is neither a local min. / local max., we call it <u>saddle point</u>.

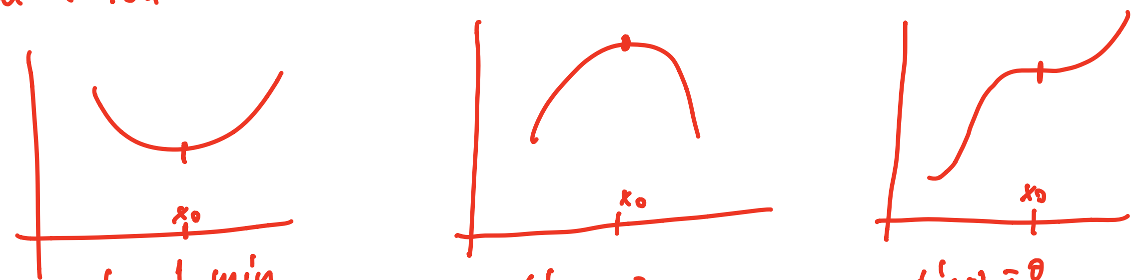




f has a global minimum at x_0 if $\forall x: f(x) \geq f(x_0)$

How can we see which type we have?

Intuition in $\mathbb{R}$:



local min.
$f'(x_0) = 0$
$f''(x_0) > 0$

$f'(x)=0$
$f''(x) < 0$

$f'(x) = 0$
$f''(x) = 0$

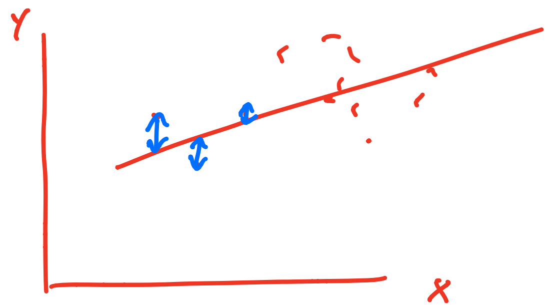
<u>Theorem</u> $f: \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^2(\mathbb{R}^n)$. Assume that x_0 is a critical point, i.e. $\nabla f(x_0) = 0$. Then:

(i) If x_0 is a local minimum (maximum), then the Hessian $Hf(x_0)$ is positive semi-definite (neg. semi-def.)

(ii) If $Hf(x_0)$ is positive definite (neg. definite), then x_0 is a strict local min (max). If $Hf(x_0)$ is indefinite, then x_0 is a saddle point.

Matrix/vector derivatives

Example: Linear least squares



$f: \mathbb{R}^n \to \mathbb{R}$

input data

$f(w) = \| y - A w \|^2$

predicted output values

true output values

weight vector
(parameters we want to find)

how good is prediction
of model with para. w

Want to minimize $f(w)$. Need to look at $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$

Compute gradient by foot:

- Write function explicitly:

$$f\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \sum_{j=1}^{n} \Big(y_j - \overbrace{\sum_{k=1}^{m} a_{jk} w_k}^{(Aw)_j} \Big)^2$$

$$\frac{\partial f}{\partial w_i} = \sum_{j=1}^{n} \left(- a_{ji} \right) \cdot 2 \cdot \Big(y_j - \underbrace{\sum_{k=1}^{m} a_{jk} w_k}_{(Aw)_j} \Big)$$

$$\underbrace{-2 \cdot \sum_{j=1}^{n} a_{ji} \cdot (y - Aw)_j}_{\left(A^t (y - Aw)\right)_i}$$

$$\nabla f(w) = -2\, A^t (y - Aw)$$

Intuition: "Syntax" close to 1-dim case:

$$f(w) = (y - a \cdot w)^2$$

$$f'(w) = -a\,(y - aw) \cdot 2 = -2a\,(y - aw)$$

Matrix-vector calculus: Lookup table ("cookbook") for gradients of many important functions:

$\boxed{f : \mathbb{R}^n \to \mathbb{R}}$

- $f(x) = a^t x \qquad (a \in \mathbb{R}^n)$

  $= \langle a, x \rangle$

  $$\frac{\partial f}{\partial x} = a \qquad \in \mathbb{R}^n$$

- $f(x) = x^t A x \quad \Rightarrow \quad \dfrac{\partial f}{\partial x} = (A + A^t)\, x \qquad \in \mathbb{R}^n$

$\boxed{f : \mathbb{R}^{n \times m} \to \mathbb{R}}$

- $f(X) = \underbrace{a^t}_{1 \times n} \underbrace{X}_{n \times m} \underbrace{b}_{m \times 1}$ for $a, b \in \mathbb{R}^n$

  $$\frac{\partial f}{\partial X} = \underbrace{\underbrace{a}_{n \times 1} \cdot \underbrace{b^t}_{1 \times m}}_{n \times m} \in \mathbb{R}^{n \times m}$$

- $f(X) = \underbrace{a^t}_{1\times m} \underbrace{X^t}_{m\times n} C \underbrace{X}_{n\times m} \underbrace{b}_{m\times 1}$ for $a \in \mathbb{R}^m$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{n\times n}$

$X \in n\times m$

$$\frac{\partial f}{\partial X} = C^t X a b^t + C X b a^t$$

- $f : \mathbb{R}^{n\times n} \to \mathbb{R}$

$$f(X) = tr(X) \Rightarrow \frac{\partial f}{\partial X} = I \in \mathbb{R}^{n\times n}$$

- $f(X) = tr(A \cdot X) \Rightarrow \frac{\partial f}{\partial X} = A$

$$f(X) = tr(X^t A X) \Rightarrow \frac{\partial f}{\partial X} = (A + A^t) X$$

- $f(X) = det(X)$ Determinant

$$\frac{\partial f}{\partial X} = det(X) (X^t)^{-1}$$

$$\frac{\partial\, det}{\partial a_{sr}} = det(A) \cdot (A^{-1})_{rs}$$

- $\boxed{f : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}}$ Inverse

$$f(A) = A^{-1}, \quad f_{ij} := (A^{-1})_{ij}$$

$$\frac{\partial f_{ij}}{\partial a_{uv}} = -(a_{iu})^{-1} (a_{vj})^{-1}$$