MAT237 - LEC5201 - 2019–2020
2019 Fall Term Notes

Jean-Baptiste Campesato

Contents

PRELIMINARIES

Cartesian product

Def: An n-tuple is an ordered list of n elements $(x_1, \dots, x_n)$

Rem: couple ≡ 2-tuple   triple ≡ 3-tuple

Fundamental property: $(x_1, \dots, x_n) = (y_1, \dots, y_n) \Leftrightarrow \forall i,\ x_i = y_i$

Rem: ① $\{1,2,3\} = \{3,2,1\}$ (Sets)

but $(1,2,3) \neq (3,2,1)$ (Tuple)

② $\{1,2,2,3\} = \{1,2,3\}$

but $(1,2,2,3) \neq (1,2,3)$

Def: Given 2 sets A and B: $A \times B = \{(a,b) : a \in A,\ b \in B\}$

Ex: $A = \{\pi, e\}$, $B = \{1, \sqrt{2}, \pi\}$

$$A \times B = \{(\pi,1), (\pi,\sqrt{2}), (\pi,\pi), (e,1), (e,\sqrt{2}), (e,\pi)\}$$

Rem: if A and B are finite then $\#(A \times B) = \#A \cdot \#B$

Def: $A_1 \times A_2 \times \dots \times A_n = \{(a_1, \dots, a_n) : a_i \in A_i\}$

Rem: We will often identify the following sets:

$$\underset{((a,b),c)}{\overset{}{(A\times B)\times C}} \ni \qquad A\times(B\times C) \ni (a,(b,c)) \qquad A\times B\times C \ni (a,b,c)$$

even if they are not formally the same set.

Functions

→ informal definition

Def: A function (or map, or mapping) is the data of two sets A and B together with a "process" that associates to each element $x \in A$ a unique element $f(x) \in B$

Notation: $f : A \longrightarrow B$ domain
name codomain

Notation: let $f: A \to B$ be a function

① the image of $E \subset A$ by f is $f(E) := \{ f(x) : x \in E \}$

② the preimage of $F \subset B$ by f is $f^{-1}(F) := \{ x \in A : f(x) \in F \}$

Def: the graph of $f : A \longrightarrow B$ is $\Gamma_f := \{ (x,y) \in A \times B : y = f(x) \}$

Rem: a function is entirely determined by its graph

Def: $f: A \longrightarrow B$ is injective (or 1-to-1) if $\forall x_1, x_2 \in A, x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$

or equivalently (contrapositive) $\forall x_1, x_2 \in A,\ f(x_1) = f(x_2) \Rightarrow x_1 = x_2$

Def: $f: A \longrightarrow B$ is surjective (or onto) if $\forall y \in B, \exists x \in A,\ y = f(x)$

Def: $f: A \to B$ is bijective if it is injective and surjective

<u>ie</u> $\forall y \in B, \exists! x \in A,\ y = f(x)$

Prop: $f : A \longrightarrow B$ is bijective iff $\exists g: B \longrightarrow A$ such that

$$\begin{cases} g \circ f = id_A \\ f \circ g = id_B \end{cases}$$

Then we say that g is the inverse of f, denoted f^{-1}

Ex: Slides

Geometry of $\mathbb{R}^n$

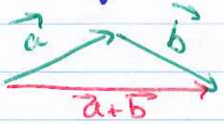
Def: $\mathbb{R}^n := \underbrace{\mathbb{R}\times\cdots\times\mathbb{R}}_{n \text{ times}} = \{(x_1,\dots,x_n) : x_i \in \mathbb{R}\}$

Rem: ① the x_i are bound variables, however, we will often use:
(x,y) for $n=2$
(x,y,z) for $n=3$
$(x_1,\dots,x_n)$ for $n>3$

② In the online notes, an element of $\mathbb{R}^n$ is written in bold, you can also use an arrow to avoid any confusion
$\vec{v}=(x_1,\dots,x_n)$

For $a=(a_1,\dots,a_n)$, $b=(b_1,\dots,b_n)\in\mathbb{R}^n$ and $\lambda\in\mathbb{R}$, we define

Addition: $a+b := (a_1+b_1, a_2+b_2, \dots, a_n+b_n) \in \mathbb{R}^n$



Scalar multiplication: $\lambda a := (\lambda a_1,\dots,\lambda a_n)\in\mathbb{R}^n$

Notation: ① $\vec{e_1}=(1,0,\dots,0)$, $\vec{e_2}=(0,1,0,\dots,0)$, …, $\vec{e_n}=(0,\dots,0,1)$ in $\mathbb{R}^n$

② $\vec{i}=(1,0,0)$, $\vec{j}=(0,1,0)$, $\vec{k}=(0,0,1)$ in $\mathbb{R}^3$

Def: (dot product) $\underset{\mathbb{R}^n}{\underset{\rotatebox[origin=c]{90}{$\in$}}{a}} \cdot \underset{\mathbb{R}^n}{\underset{\rotatebox[origin=c]{90}{$\in$}}{b}} := a_1b_1 + a_2b_2 + \cdots + a_nb_n \in \mathbb{R}$: it takes 2 vectors and gives 1 scalar

Prop: for $a,b,c\in\mathbb{R}^n$, $\lambda\in\mathbb{R}$

① $a\cdot b = b\cdot a$ (commutativity)
② $(\lambda a+b)\cdot c = \lambda(a\cdot c)+b\cdot c$ (bilinearity)
③ $a\neq 0 \Rightarrow a\cdot a>0$ (positive definite)

} the dot product is an inner-product

④ $a\cdot a=0 \Rightarrow a=0$
⑤ $0\cdot a=0$

Ex: $(1,2)\cdot(-1,3) = 5$

$(1,0,3)\cdot(-1,1,-1) = -4$

$(1,-1,1,-1)\cdot(1,0,2,-1) = 4$

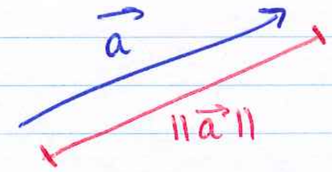
but $(1,1,1)\cdot(1,2)$ is not defined

or $|a|$ in the online notes

Def: (Euclidean norm)
For $a \in \mathbb{R}^n$, we denote $\|a\| := \sqrt{a\cdot a} = \sqrt{a_1^2 + \dots + a_n^2}$

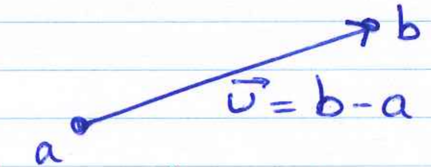
Geometric interpretation

① $\|\vec{a}\|$ is the length of $\vec{a}$
(or magnitude)



② $\|b-a\|$ is the distance between a and b

$\| \quad \sqrt{(b_1-a_1)^2 + \dots + (b_n-a_n)^2}$



Rem: An element of $\mathbb{R}^n$ may represent a vector (velocity, force) or a point (position)

Prop: for $a,b,c \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$

① $\|a\| \geq 0$

② $\|a\| = 0 \Rightarrow a = 0$ (positive definite)

③ $\|\lambda a\| = |\lambda| \cdot \|a\|$ (positive homogeneity)

④ $\|a+b\| \leq \|a\| + \|b\|$ (triangle inequality)

} $\|\cdot\|$ is a norm
2,3,4 ⇒ 1

⑤ $|a\cdot b| \leq \|a\| \cdot \|b\|$ (Cauchy-Schwarz inequality)

⑥ $a\cdot e_j = a_j$, $e_j \cdot e_j = 1$, $e_i \cdot e_j = 0$ for $i \neq j$

⑦ $a\cdot b = \frac{1}{4}\left(\|a+b\|^2 - \|a-b\|^2\right)$ (Polarization identity)

Proof of 5 :

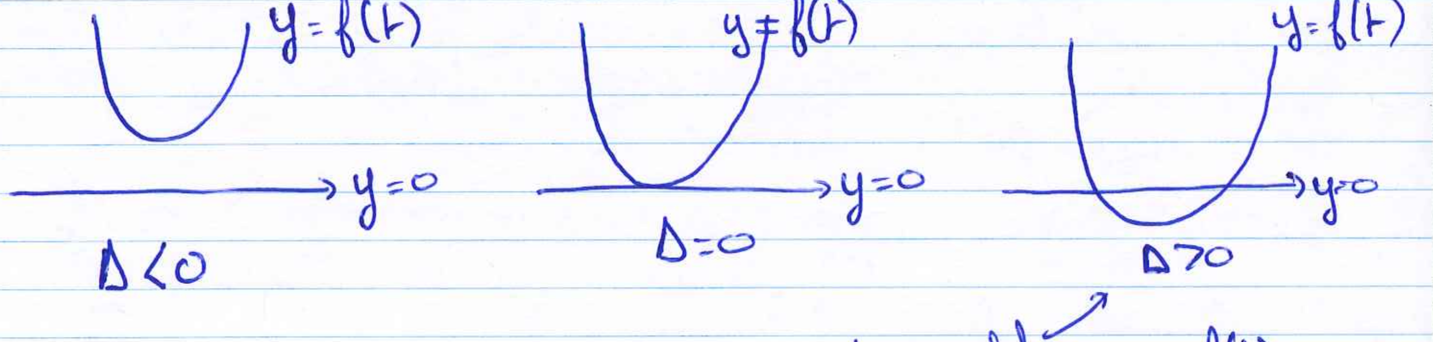
For $t \in \mathbb{R}$ we set $f(t) = \|a + tb\|^2$

then $f(t) = \|b\|^2 t^2 + 2(a \cdot b)\, t + \|a\|^2$

First case : $\|b\| = 0$ and then $b = 0$ and the result is obvious

Second case : $\|b\| \neq 0$ and then f is a quadratic polynomial with positive leading coefficient.

We have the following possibilities :



not possible since $f(t) \geq 0$

Hence $\Delta \leq 0$, but $\Delta = 4(a \cdot b)^2 - 4\|a\|^2\|b\|^2$

so $(a \cdot b)^2 \leq \|a\|^2 \|b\|^2$

and $|(a \cdot b)| \leq \|a\| \, \|b\|$ □

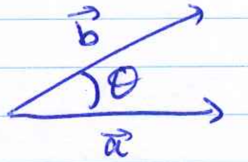
Proof of 4 :

$$\|a+b\|^2 = \|a\|^2 + 2(a \cdot b) + \|b\|^2$$
$$\leq \|a\|^2 + 2|a \cdot b| + \|b\|^2$$
$$\leq \|a\|^2 + 2\|a\| \cdot \|b\| + \|b\|^2$$
$$= (\|a\| + \|b\|)^2$$

thus $\|a+b\| \leq \|a\| + \|b\|$ □

Geometric interpretation of the dot product

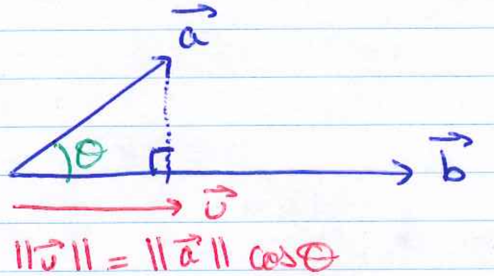
$$\vec{a}\cdot\vec{b} = \|\vec{a}\|\cdot\|\vec{b}\|\cdot\cos\theta$$



Def: We say that $a, b \in \mathbb{R}^n$ are orthogonal when $a\cdot b = 0$

Consequences: (of the geometric definition)

①



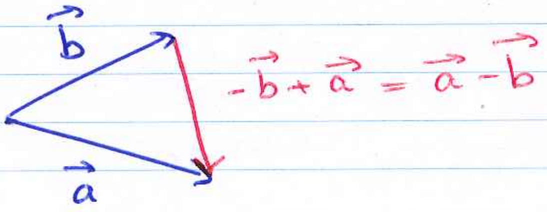
so $\vec{u} = \|\vec{a}\|\cos\theta \cdot \dfrac{\vec{b}}{\|\vec{b}\|}$

$$= \frac{\|\vec{a}\|\cdot\|\vec{b}\|\cdot\cos\theta}{\|\vec{b}\|^2}\,\vec{b}$$

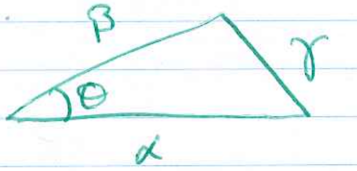
$$= \frac{(\vec{a}\cdot\vec{b})}{\|\vec{b}\|^2}\,\vec{b}$$

Conclusion: $\left(\dfrac{\vec{a}\cdot\vec{b}}{\|\vec{b}\|^2}\right)\vec{b}$ is the orthogonal projection of $\vec{a}$ on the line spanned by $\vec{b}$

② Law of cosines: $\|a-b\|^2 = \|a\|^2 + \|b\|^2 - 2(a\cdot b)$



i.e. in a triangle:



$$\gamma^2 = \alpha^2 + \beta^2 - 2\alpha\beta\cos\theta$$

Homework. Find the angles of the triangle whose vertices are $A(-1, 0)$, $B\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$, $C\left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right)$

Hint: compute $\vec{AB}\cdot\vec{AC}$ using both the geometric and algebraic definitions

Rem: if $a\cdot b = 0$ then $\|a+b\|^2 = \|a\|^2 + \|b\|^2$ (Pythagorean thm)

Cross product (⚠ only in $\mathbb{R}^3$, $n=3$)

Def: (Cross product) for $a, b \in \mathbb{R}^3$

$$a \times b = (a_2b_3 - a_3b_2,\ a_3b_1 - a_1b_3,\ a_1b_2 - a_2b_1) \in \mathbb{R}^3$$

(takes 2 vectors, gives 1 vector)

Mnemonic devices:

①

$$\begin{array}{cc} \cancel{a_1} & \cancel{b_1} \\ a_2 & b_2 \\ a_3 & b_3 \\ a_1 & b_1 \\ a_2 & b_2 \end{array}$$

$a_2b_3 - a_3b_2$

$a_3b_1 - a_1b_3$

$a_1b_2 - a_2b_1$

② Compute the following determinant w.r.t the last column:

$$\begin{vmatrix} a_1 & b_1 & \vec{i} \\ a_2 & b_2 & \vec{j} \\ a_3 & b_3 & \vec{k} \end{vmatrix} = \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}\vec{i} \underset{\uparrow ⚠}{-} \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix}\vec{j} + \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}\vec{k}$$

Prop: for $a, b, c \in \mathbb{R}^3$, $\lambda \in \mathbb{R}$

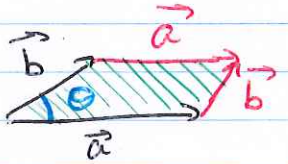
① $b \times a = -(a \times b)$

② $(\lambda a + b) \times c = \lambda(a \times c) + b \times c$

③ $a \times a = 0$

④ $\|a \times b\|^2 + (a \cdot b)^2 = \|a\|^2 \cdot \|b\|^2$

⑤ $\|a \times b\| = \|a\| \cdot \|b\| \cdot |\sin\theta|$ → area of the parallelogram defined by $\vec{a}$ and $\vec{b}$



⑥ $a \cdot (a \times b) = 0$, $b \cdot (a \times b) = 0$

⑦ $\vec{i} \times \vec{j} = \vec{k}$, $\vec{j} \times \vec{k} = \vec{i}$, $\vec{k} \times \vec{i} = \vec{j}$

⚠ $(a \times b) \times c \neq a \times (b \times c)$: example $(\vec{i} \times \vec{i}) \times j = \vec{0}$ $\quad \vec{i} \times (\vec{i} \times \vec{j}) = -\vec{j}$

But: ⑧ $a \times (b \times c) = (a \cdot c)b - (a \cdot b)c$

$(a \times b) \times c = (a \cdot c)b - (b \cdot c)a$

⑨ $a \times (b \times c) + b \times (c \times a) + c \times (a \times b) = 0$ (Jacobi identity)

⑩ $(a\times b)\cdot c = \det\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$

Homework: let $\vec{a}=(2,-3,1)$, $\vec{b}=(3,-5,2)$, $\vec{c}=(4,-5,1)$

(i) Compute $\vec{a}\times\vec{b}$. Are $\vec{a}$ and $\vec{b}$ collinear?
(ii) Compute $(\vec{a}\times\vec{b})\cdot\vec{c}$. Are $\vec{a},\vec{b},\vec{c}$ coplanar?

Geometric interpretation of the cross-product.

- if $\vec{a}$ and $\vec{b}$ are collinear then $\vec{a}\times\vec{b}=\vec{0}$
- otherwise $\vec{a}\times\vec{b}$ is the unique vector $\vec{c}$ such that
  (i) $\|\vec{c}\| = \|\vec{a}\|\cdot\|\vec{b}\|\,|\sin\theta|$
  (ii) $\vec{c}$ is orthogonal to $\vec{a}$ and $\vec{b}$
  (iii) $\det\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} > 0$

↳ Right-hand rule: thumb $=\vec{a}$, index $=\vec{b}$, middle $=\vec{a}\times\vec{b}$
↳ ⚠ not left!

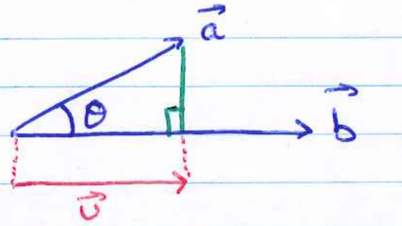
Δ • if $\vec{a}$ and $\vec{b}$ are collinear then $\|\vec{a}\times\vec{b}\| = \|\vec{a}\|\cdot\|\vec{b}\|\,|\sin\theta| = 0$ by ⑤

• otherwise: (i) $\|\vec{a}\times\vec{b}\| = \|\vec{a}\|\cdot\|\vec{b}\|\cdot|\sin\theta|$ by ⑤
(ii) $\vec{a}\cdot(\vec{a}\times\vec{b})=0$, $\vec{b}\cdot(\vec{a}\times\vec{b})=0$ by ⑥
(iii) $\det\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} = (\vec{a}\times\vec{b})\cdot(\vec{a}\times\vec{b})$ by ⑩ if $\vec{c}=\vec{a}\times\vec{b}$
$= \|\vec{a}\times\vec{b}\|^2 > 0$ since $\vec{a}\times\vec{b}\neq\vec{0}$ by (i)

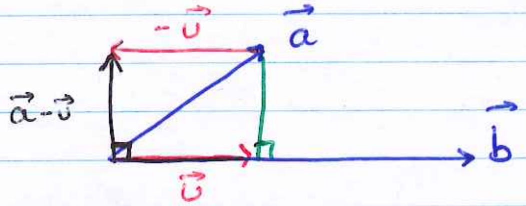
then $\vec{a}\times\vec{b}$ is uniquely determined since we know its direction and length Δ

Extra definition: (via triple product)
$\vec{a}\times\vec{b}$ is the unique vector of $\mathbb{R}^3$ s.t.
$\forall \vec{c}\in\mathbb{R}^3$, $\det\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} = (\vec{a}\times\vec{b})\cdot\vec{c}$

Extra: since two of you asked me for a geometric proof of the Cauchy-Schwarz inequality, here it is:



$\vec{v} = \frac{(\vec{a}\cdot\vec{b})}{\|\vec{b}\|^2}\vec{b}$ is the orthogonal projection of $\vec{a}$ on the line spanned by $\vec{b}$



we see that $\vec{a}-\vec{v}$ is orthogonal to $\vec{v}$, ie $(\vec{a}-\vec{v})\cdot\vec{v} = 0$

(check that $(\vec{a}-\vec{v})\cdot\vec{v} = 0$ algebraically)

hence $\|\vec{a}\|^2 = \|(\vec{a}-\vec{v})+\vec{v}\|^2$

$= \|\vec{a}-\vec{v}\|^2 + \|\vec{v}\|^2$ since $(\vec{a}-\vec{v})\cdot\vec{v} = 0$

$\geqslant \|\vec{v}\|^2$ since $\|\vec{a}-\vec{v}\| \geqslant 0$

$= \frac{(\vec{a}\cdot\vec{b})^2}{\|\vec{b}\|^4}\cdot\|\vec{b}\|^2$ by definition of $\vec{v}$

$$= \frac{(\vec{a}\cdot\vec{b})^2}{\|\vec{b}\|^2}$$

Hence $(\vec{a}\cdot\vec{b})^2 \leqslant \|\vec{a}\|^2\|\vec{b}\|^2$

and $|\vec{a}\cdot\vec{b}| \leqslant \|\vec{a}\|\,\|\vec{b}\|$

QED

□

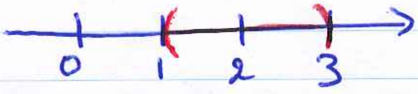
SOME TOPOLOGICAL NOTIONS

Terminology: Balls and Spheres

Def: For $a \in \mathbb{R}^m$ and $r \in \mathbb{R}_{>0}$, the open ball centred at a with radius r is:

$$B(a,r) := \{x \in \mathbb{R}^m : \|x-a\| < r\} \subset \mathbb{R}^m$$

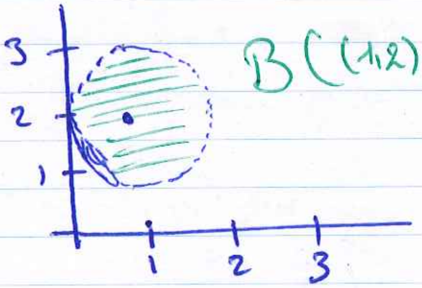
Ex:



$B(2,1) = (1,3) \subset \mathbb{R}$

$\overline{B}(2,1) = [1,3] \subset \mathbb{R}$

$S(2,1) = \{1,3\} \subset \mathbb{R}$



$B((1,2),1) \subset \mathbb{R}^2$

Def: the closed ball centred at a with radius r is

$$\overline{B}(a,r) := \{x \in \mathbb{R}^m : \|x-a\| \leq r\} \subset \mathbb{R}^m$$

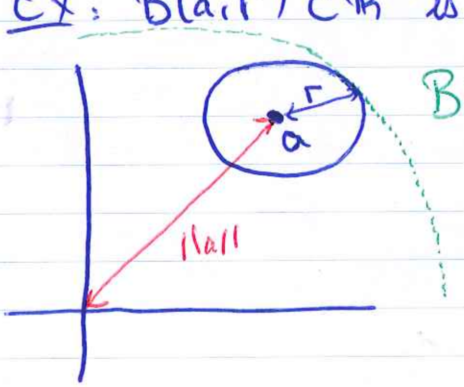
Def: the sphere centred at a with radius r is

$$S(a,r) := \{x \in \mathbb{R}^m : \|x-a\| = r\}$$

Def: A subset $S \subset \mathbb{R}^m$ is bounded if there exists $r \in \mathbb{R}_{>0}$ st. $S \subset B(0,r)$

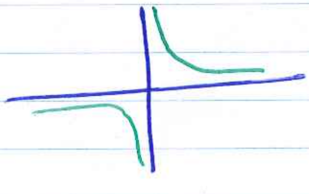
ie $\exists r > 0, \forall x \in S, \|x\| < r$

Ex: $B(a,r) \subset \mathbb{R}^m$ is bounded



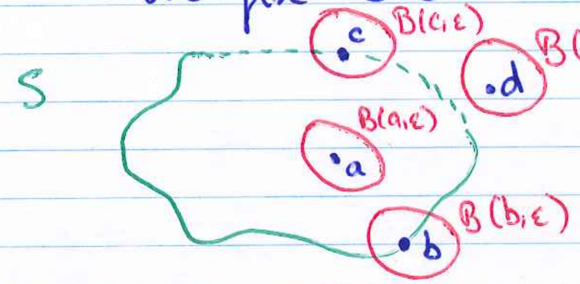
$B(a,r) \subset B(0, \|a\|+r)$

Ex: $\{(x,y) : xy = 1\}$ is not bounded



Terminology: Interior, closure, boundary

We fix $S \subset \mathbb{R}^m$



$a \in S$	$a \in \mathring{S}$	$a \in \overline{S}$	$a \notin \partial S$
$b \in S$	$b \notin \mathring{S}$	$b \in \overline{S}$	$b \in \partial S$
$c \notin S$	$c \notin \mathring{S}$	$c \in \overline{S}$	$c \in \partial S$
$d \notin S$	$d \notin \mathring{S}$	$d \notin \overline{S}$	$d \notin \partial S$

Def: We say that $x \in \mathbb{R}^m$ is an interior point of S if there exists $\varepsilon > 0$ s.t. $B(x,\varepsilon) \subset S$

Notation: the interior of S is $\mathring{S} := \{x \in \mathbb{R}^m : \exists \varepsilon > 0, B(x,\varepsilon) \subset S\}$
or S^{int}

Def: we say that $x \in \mathbb{R}^m$ is a closure point (or adherent point) of S if $\forall \varepsilon > 0, B(x,\varepsilon) \cap S \neq \emptyset$

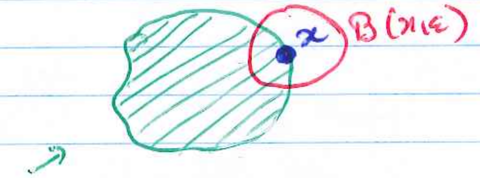
Notation: the closure of S is $\overline{S} := \{x \in \mathbb{R}^m : \forall \varepsilon > 0, B(x,\varepsilon) \cap S \neq \emptyset\}$

Theorem: $\mathring{S} \subset S \subset \overline{S}$

Δ $\mathring{S} \subset S$: let $x \in \mathring{S}$ then $\exists \varepsilon > 0$ s.t. $B(x,\varepsilon) \subset S$
hence $x \in B(x,\varepsilon) \subset S$
so $x \in \mathring{S} \Rightarrow x \in S$

$S \subset \overline{S}$: let $x \in S$.
For any $\varepsilon > 0$, $x \in S \cap B(x,\varepsilon) \neq \emptyset$, so $x \in \overline{S}$ □

Def: the boundary of S is $\partial S := \overline{S} \setminus \mathring{S}$



Prop: $\overline{S} = S \cup \partial S$ and $\partial S \cap \mathring{S} = \emptyset$

Theorem: $x \in \partial S \Leftrightarrow \forall \varepsilon > 0, B(x,\varepsilon) \cap S \neq \emptyset$ and $B(x,\varepsilon) \cap S^c \neq \emptyset$

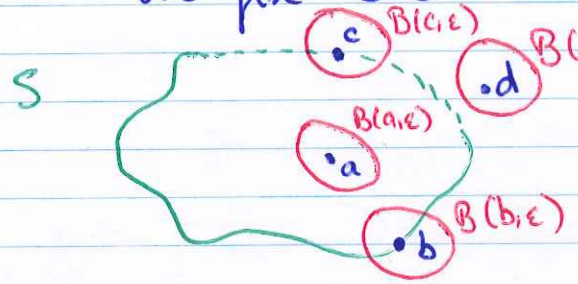
Δ $x \in \overline{S} \Leftrightarrow \forall \varepsilon > 0, B(x,\varepsilon) \cap S \neq \emptyset$
$x \notin \mathring{S} \Leftrightarrow$ no $(\exists \varepsilon > 0, B(x,\varepsilon) \subset S) \Leftrightarrow \forall \varepsilon > 0, B(x,\varepsilon) \cap S^c \neq \emptyset$ □

Cor: $\partial S = \partial(S^c)$

Δ Notice that $(S^c)^c = S$ □

Terminology: Interior, closure, boundary

We fix $S \subset \mathbb{R}^m$



$a \in S$	$a \in \mathring{S}$	$a \in \overline{S}$	$a \notin \partial S$
$b \in S$	$b \notin \mathring{S}$	$b \in \overline{S}$	$b \in \partial S$
$c \notin S$	$c \notin \mathring{S}$	$c \in \overline{S}$	$c \in \partial S$
$d \notin S$	$d \notin \mathring{S}$	$d \notin \overline{S}$	$d \notin \partial S$

Def: We say that $x \in \mathbb{R}^m$ is an interior point of S if there exists $\varepsilon > 0$ s.t. $B(x,\varepsilon) \subset S$

Notation: the interior of S is $\mathring{S}$ or S^{int} $:= \{x \in \mathbb{R}^m : \exists \varepsilon > 0, B(x,\varepsilon) \subset S\}$

Def: we say that $x \in \mathbb{R}^m$ is a closure point (or adherent point) of S if $\forall \varepsilon > 0, B(x,\varepsilon) \cap S \neq \emptyset$

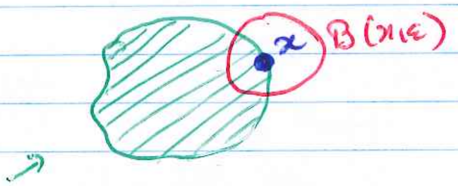
Notation: the closure of S is $\overline{S} := \{x \in \mathbb{R}^m : \forall \varepsilon > 0, B(x,\varepsilon) \cap S \neq \emptyset\}$

Theorem: $\mathring{S} \subset S \subset \overline{S}$

Δ $\mathring{S} \subset S$: let $x \in \mathring{S}$ then $\exists \varepsilon > 0$ s.t. $B(x,\varepsilon) \subset S$
hence $x \in B(x,\varepsilon) \subset S$
so $x \in \mathring{S} \Rightarrow x \in S$

$S \subset \overline{S}$: let $x \in S$.
For any $\varepsilon > 0$, $x \in S \cap B(x,\varepsilon) \neq \emptyset$, so $x \in \overline{S}$ □

Def: the boundary of S is $\partial S := \overline{S} \setminus \mathring{S}$



Prop: $\overline{S} = S \cup \partial S$ and $\partial S \cap \mathring{S} = \emptyset$

Theorem: $x \in \partial S \Leftrightarrow \forall \varepsilon > 0, B(x,\varepsilon) \cap S \neq \emptyset$ and $B(x,\varepsilon) \cap S^c \neq \emptyset$

Δ $x \in \overline{S} \Leftrightarrow \forall \varepsilon > 0, B(x,\varepsilon) \cap S \neq \emptyset$
$x \notin \mathring{S} \Leftrightarrow$ no $(\exists \varepsilon > 0, B(x,\varepsilon) \subset S) \Leftrightarrow \forall \varepsilon > 0, B(x,\varepsilon) \cap S^c \neq \emptyset$ □

Cor: $\partial S = \partial(S^c)$

Δ Notice that $(S^c)^c = S$ □

Terminology: open and closed sets

Def: $S \subset \mathbb{R}^m$ is open if $\mathring{S} = S$

Theorem: S is open $\Leftrightarrow S \cap \partial S = \emptyset$

Δ $\Rightarrow$ Assume that S is open, then $S \cap \partial S = \mathring{S} \cap \partial S = \emptyset$

$\Leftarrow$: We are going to prove the contrapositive: S is not open $\Rightarrow S \cap \partial S \neq \emptyset$
Assume that S is not open then $\mathring{S} \subsetneq S$ and there exists $x \in S \setminus \mathring{S}$
but $S \setminus \mathring{S} \subset \bar{S} \setminus \mathring{S} = \partial S$
so $x \in S \cap \partial S \neq \emptyset$ □

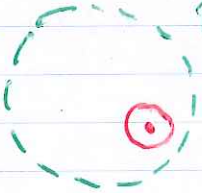
↳ that's why we will like to have open sets as domains in calculus.

Theorem: S is open $\Leftrightarrow \forall x \in S, \exists \varepsilon > 0, B(x,\varepsilon) \subset S$

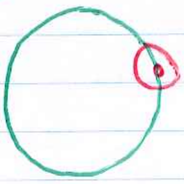
Δ $\Rightarrow$: let $x \in S$, then $x \in \mathring{S}$ since $\mathring{S} = S$.
so $\exists \varepsilon > 0$, $B(x,\varepsilon) \subset S$

$\Leftarrow$: we already know that $\mathring{S} \subset S$. Let's prove that $S \subset \mathring{S}$.
Let $x \in S$, then $\exists \varepsilon > 0$ s.t $B(x,\varepsilon) \subset S$
hence $x \in \mathring{S}$ and therefore $S \subset \mathring{S}$ □

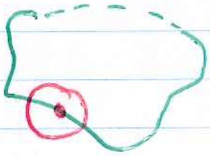
Ex



$B(a,r)$ is open



$\bar{B}(a,r)$ is not open

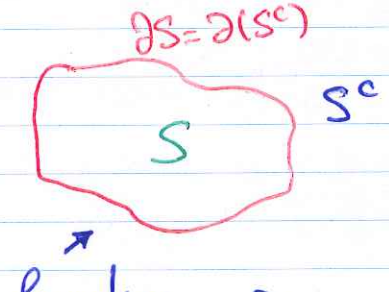


S is not open

Def: $S \subset \mathbb{R}^m$ is closed if $\bar{S} = S$

Theorem: S is closed $\Leftrightarrow \partial S \subset S$

Δ $\Rightarrow \partial S \subset \bar{S} = S$
$\Leftarrow$ $\bar{S} = S \cup \partial S = S$ □



Theorem: S is closed $\Leftrightarrow S^c$ is open

Δ Recall that $\partial(S^c) = \partial S$ and look at the boundary □

Question: Find a subset of $\mathbb{R}$ which is

① Open but not closed

② Closed but not open

③ Neither closed nor open

④ Both open and closed

Question: Find all the subsets of $\mathbb{R}^n$ that are both open and closed

Question: Prove that $\mathring{S}$ is open and that $\overline{S}$ is closed

Question: Prove $(\mathring{S})^c = \overline{S^c}$

$$(\overline{S})^c = \mathring{\overline{S^c}}$$

Do the questions at the end of section 1.1 !

Advanced questions:

① Let $(O_i)_{i\in I}$ be a possibly infinite family of open subsets of $\mathbb{R}^n$

prove that $O := \bigcup_{i\in I} O_i = \{x \in \mathbb{R}^n : \exists i \in I, x \in O_i\}$

is open

② Prove that if $U, V \subset \mathbb{R}^n$ are open then $U \cap V$ is open

③ Find an (infinite) family of open sets whose intersection is not open

④ Using that "S closed $\Leftrightarrow$ S^c open" obtain results about closed sets

Limits of multivariable functions

In class activity: start with the definition of $\ell = \lim_{x\to a} f(x)$ for $f: I \to \mathbb{R}$ a one variable function defined on an interval I containing a.
Generalize the above definition and check that we need to restrict to limit point

Def: let $S \subset \mathbb{R}^n$. We say that $a \in \mathbb{R}^n$ is a limit point of S if

$$\forall \delta > 0, \exists x \in S, \; 0 < \|x-a\| < \delta$$

or geometrically: $\forall \delta > 0, \; (B(a,\delta) \cap S) \setminus \{a\} \neq \emptyset$

Theorem: a is a limit point of $S \Leftrightarrow a \in \overline{S \setminus \{a\}}$

Δ Notice that $(B(a,\delta) \cap S) \setminus \{a\} = B(a,\delta) \cap (S \setminus \{a\})$ □

Ex: ① 0 is a limit point of $\{\frac{1}{n} : n \in \mathbb{N}_{>0}\} \subset \mathbb{R}$

② 0 is a limit point of $[0,3)$ or of $(0,3)$

③ 0 is NOT a limit point of $\{0\} \cup [1,2)$

Intuition: a limit point is a closure point which is not isolated

Remark: if $a \in \mathring{S}$ then a is a limit point of S

Def: let $S \subset \mathbb{R}^n$, a be a limit point of S, $f: S \to \mathbb{R}^k$ and $L \in \mathbb{R}^k$
We say that L is the limit of f at a, denoted $\lim_{x \to a} f(x) = L$ if

$$\forall \varepsilon > 0, \exists \delta > 0, \forall x \in S, \; 0 < \|x-a\| < \delta \Rightarrow \|f(x) - L\| < \varepsilon$$

Proposition: let $f,g: S \to \mathbb{R}$ (⚠ $k=1$) and a be a limit point of S then

$$\lim_{x\to a} f(x) = L, \; \lim g(x) = M \Rightarrow \begin{cases} \lim (f+g) = M+L \\ \lim fg = ML \end{cases}$$

Proposition: $f,g,h: S \to \mathbb{R}$ (⚠ $k=1$) and a limit point of S

$$\begin{cases} f \leq g \leq h \\ \lim_a f = \lim_a h = L \end{cases} \Rightarrow \lim_{x\to a} g = L$$

Theorem: let $f=(f_1,\dots,f_k): S \longrightarrow \mathbb{R}^k$, a be a limit point of S and $L=(L_1,\dots,L_k) \in \mathbb{R}^k$.

$$\lim_{x\to a} f(x) = L \iff \forall i=1,\dots,k, \quad \lim_{x\to a} f_i(x) = L_i$$

Comment: hence it's enough to understand the real-valued case

Δ First notice that $\lim_{a} f = L \Leftrightarrow \lim_{x\to a} \|f(x)-L\| = 0$

$\Rightarrow$: $|f_i(x)-L_i| \leqslant \left(|f_1(x)-L_1|^2+\dots+|f_k(x)-L_k|^2\right)^{1/2} = \|f(x)-L\|$

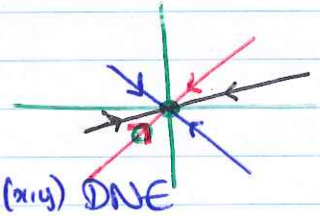
so $\lim_{x\to a} f(x) = L \Rightarrow \lim_{x\to a} f_i(x) = L_i$

$\Leftarrow$: $\|f(x)-L\| = \left(|f_1(x)-L_1|^2+\dots+|f_k(x)-L_k|^2\right)^{1/2}$

so $\left(\forall i, \ |f_i - L_i| \longrightarrow 0\right) \Rightarrow \lim_{x\to a} f(x) = L$ □

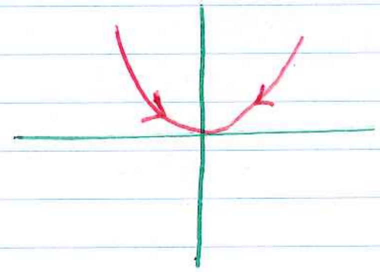
Rem: in the one variable case it's enough to check that the limits from the right and from the left coincide.
In $\mathbb{R}^n$ the situation is more subtle since we have more "freedom" to approach $a \in \mathbb{R}^n$

Ex①: let $f(x,y) = \dfrac{xy}{x^2+y^2}$



then $f(x,cx) = \dfrac{c}{1+c^2}$ so $\lim_{(x,y)\to(0,0)} f(x,y)$ DNE

② let $f(x,y) = \dfrac{x^2y}{x^4+y^2}$



then $f(x,cx) = \dfrac{cx^3}{x^4+c^2x^2} \xrightarrow[x\to 0]{} 0$

but $f(x,x^2) = \dfrac{x^4}{x^4+x^4} = \dfrac{1}{2} \longrightarrow \dfrac{1}{2} \neq 0$

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

Ccl: it's not enough to look along lines (or even parabolas, what of )

Ex: $f(x,y) = \dfrac{x^2y^2}{x^2+y^2}$ at $(0,0)$

$$|f(x,y)| = \frac{|xy|}{x^2+y^2} \cdot |xy| \leqslant \frac{1}{2}|xy| \xrightarrow[(x,y)\to(0,0)]{} 0$$

so $\lim\limits_{(x,y)\to(0,0)} f(x,y) = 0$

Here, I used the following <u>very</u> useful inequality

$$\boxed{\forall (x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}, \quad \frac{|xy|}{x^2+y^2} \leqslant \frac{1}{2}}$$

Δ let $(x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$ then

$$0 \leqslant (|x|-|y|)^2 = x^2+y^2-2|x||y|$$

$$\Rightarrow 2|xy| \leqslant x^2+y^2$$

$$\Rightarrow \frac{|xy|}{x^2+y^2} \leqslant \frac{1}{2}$$

$\square$

Ex: $\dfrac{|x|y^2}{x^2+y^4} \leqslant \dfrac{1}{2}$

Continuity of multivariable functions

Def: let $S \subset \mathbb{R}^m$, $f : S \to \mathbb{R}^k$ and $a \in S$.
We say that f is continuous at a if:

$$\forall \varepsilon > 0, \exists \delta > 0, \forall x \in S, \ \|x-a\| < \delta \Rightarrow \|f(x) - f(a)\| < \varepsilon$$

Remarks: ① We don't require a to be a limit point of S: if a is isolated then f is continuous at a.

② However, if a is a limit point of S, then
f is continuous at $a \Leftrightarrow \lim_{x \to a} f(x) = f(a)$

Theorem: $f = (f_1, \dots, f_k) : S \to \mathbb{R}^k$ is continuous at a if and only if each component $f_i : S \to \mathbb{R}$ is continuous at a.

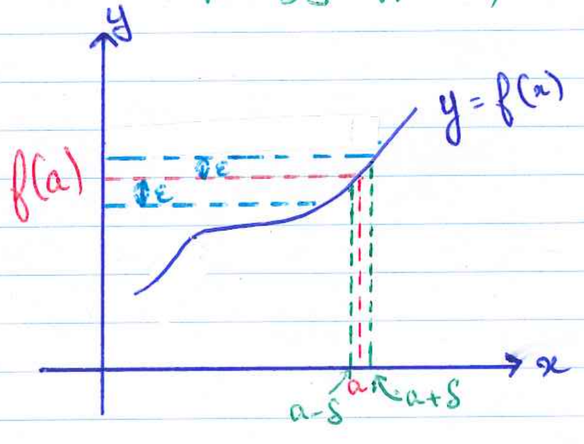
Again, it is enough to understand well the real valued case ($k=1$)

Remark: in the real valued case ($k=1$) the usual "limit laws" remain true so we can build continuous functions using the elementary functions.

Homework: read theorem 5 of section 1.2 (online notes)

⚠ f is continuous at a iff: $\forall \varepsilon > 0, \exists \delta > 0, \ f(B(a,\delta) \cap S) \subset B(f(a), \varepsilon)$

(it's where topology appears)



$\forall \varepsilon > 0, \exists \delta > 0,$

$f((a-\delta, a+\delta)) \subset (f(a)-\varepsilon, f(a)+\varepsilon)$

Theorem: let $S \subset \mathbb{R}^m$ and $f : S \to \mathbb{R}^k$

TFAE ① f is continuous

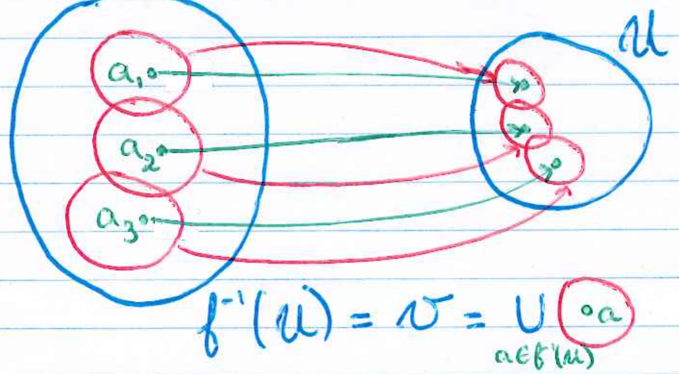
② $\forall \mathcal{U} \subset \mathbb{R}^k$ open set, $\exists \mathcal{V} \subset \mathbb{R}^m$ open set, s.t. $f^{-1}(\mathcal{U}) = \mathcal{V} \cap S$

③ $\forall C \subset \mathbb{R}^k$ closed set, $\exists D \subset \mathbb{R}^m$ closed set, s.t. $f^{-1}(C) = D \cap S$

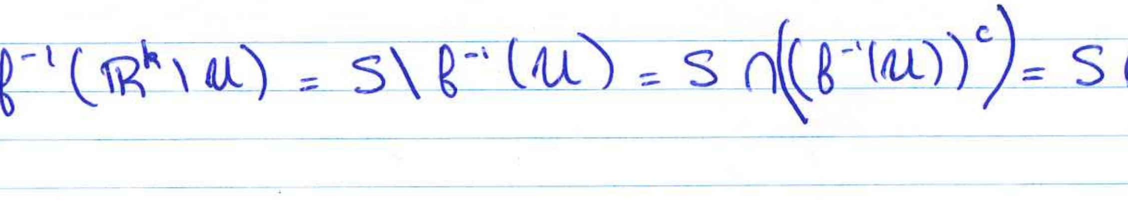
if the domain is $\mathbb{R}^m$: f is continuous $\Leftrightarrow$ the inverse image of an open is open
$\Leftrightarrow$ ——————— a closed is closed

Δ 1⇒2: let $a \in f^{-1}(\mathcal{U})$ then $f(a) \in \mathcal{U}$ open so $\exists \varepsilon > 0$ s.t. $B(f(a), \varepsilon) \subset \mathcal{U}$
then, by continuity of f, $\exists \delta_a > 0$ s.t. $f(B(a, \delta_a) \cap S) \subset B(f(a), \varepsilon) \subset \mathcal{U}$

we can take $\mathcal{V} = \bigcup_{a \in f^{-1}(\mathcal{U})} B(a, \delta_a)$



2⇒1: let $a \in S$, let $\varepsilon > 0$, then $B(f(a), \varepsilon)$ is open as an open ball.
by assumption $\exists \mathcal{V} \subset \mathbb{R}^m$ open s.t. $f^{-1}(B(f(a), \varepsilon)) = S \cap \mathcal{V}$
since $a \in \mathcal{V}$ open, $\exists \delta > 0$, $B(a, \delta) \subset \mathcal{V}$
then $f(B(a, \delta) \cap S) \subset f(\mathcal{V} \cap S) \subset B(f(a), \varepsilon)$



2⇔3: $f^{-1}(\mathbb{R}^k \setminus \mathcal{U}) = S \setminus f^{-1}(\mathcal{U}) = S \cap ((f^{-1}(\mathcal{U}))^c) = S \cap (\mathcal{V}^c)$

□

Ex. ① $S = \{(x,y) \in \mathbb{R}^2 : |x-y| = 1\}$ is closed

indeed $S = f^{-1}(\{1\})$ where $f(x,y) = |x-y|$ is continuous and $\{1\} \subset \mathbb{R}$ is closed

② $T = \{(x,y) \in \mathbb{R}^2 : |x-y| > 1\}$ is open

indeed $T = f^{-1}((1,+\infty))$ where $(1,+\infty)$ is open

Homework: Q from S.1.2 of the lecture notes

Sequences in $\mathbb{R}^m$

Def: A sequence in $\mathbb{R}^m$ is a function $\begin{array}{ccc}\{k\in\mathbb{N} : k \geq k_0\} & \longrightarrow & \mathbb{R}^m \\ k & \longmapsto & a_k\end{array}$

We use the notation $(a_k)_{k\geq k_0}$

The online notes use $\{a_k\}_{k\geq k_0}$, but I prefer (a_k) since the order matters

Def: We say that a sequence $(a_k)_{k\geq k_0}$ in $\mathbb{R}^m$ converges to $L\in\mathbb{R}^m$ if

$$\forall \varepsilon > 0,\ \exists K\in\mathbb{N},\ \forall k\in\mathbb{N}_{\geq k_0},\quad k\geq K \Rightarrow \|a_k - L\| < \varepsilon$$

denoted by $\lim\limits_{k\to+\infty} a_k = L$

Remark. $\lim\limits_{k\to+\infty} a_k = L \Leftrightarrow \lim\limits_{k\to+\infty} \overbrace{\|a_k - L\|}^{\text{real valued sequence}} = 0$

The proof is the same as the one for functions

Theorem: let $(a_k)_{k\geq k_0}$ be a sequence in $\mathbb{R}^m$. Denote $a_k = (a_{k1}, \dots, a_{km})$.

Then $\lim\limits_{k\to+\infty} a_k = L \Leftrightarrow \forall i = 1,\dots,m,\ \lim\limits_{k\to+\infty} a_{ki} = L_i$

Again, it's enough to understand well the real valued case to compute limits

Ex: $\left(\frac{1}{k}, \frac{2k^2}{k^2+1}\right) \xrightarrow[k\to+\infty]{} (0,2)$ since $\begin{cases}\lim\limits_{k\to+\infty} \frac{1}{k} = 0 \\ \lim\limits_{k\to+\infty} \frac{2k^2}{k^2+1} = 2\end{cases}$

Theorem: let $(a_k)_{k\geq k_0}$ be a convergent sequence in $\mathbb{R}^m$ and $S\subset\mathbb{R}^m$

If $\forall k\geq k_0,\ a_k\in S$ then $\lim\limits_{k\to+\infty} a_k \in \overline{S}$

Δ Denote $L = \lim\limits_{k\to+\infty} a_k$, let $\varepsilon > 0$, then $\exists K\in\mathbb{N}$ s.t. $\|a_K - L\| < \varepsilon$, ie $a_K\in B(L,\varepsilon)$.

hence $B(L,\varepsilon)\cap S \neq \emptyset$. Furthermore $L\in\overline{S}$ $\square$

Ex: A sequence "can't escape" from a closed set.

Ex: $\{x\in\mathbb{R}^2 : \|x\|\notin\mathbb{Q}\}$ is not closed : $a_m = \left(\frac{1}{m}, \frac{1}{m}\right)$

Ex: $\{(x,y) : |x|<1, |y|<1\}$ not closed $\left(1-\frac{1}{m}, 1\right)$

Def: We say that a sequence $(a_k)_{k \geq k_0}$ is bounded if

$$\exists M > 0,\ \forall k \in \mathbb{N}_{\geq k_0},\ \|a_k\| < M$$

Def: A subsection of a sequence $(a_k)_{k \geq k_0}$ is a sequence $(a_{\varphi(j)})_{j \in \mathbb{N}}$ where $\varphi : \mathbb{N} \to \{k \in \mathbb{N} : k \geq k_0\}$ is increasing.

⚠ in the notes they write a_{k_j} for $a_{\varphi(j)}$, but it may be confusing.

Intuitively, we omit some terms

$$\begin{array}{cccccccccc} \cancel{a_{k_0}} & a_{k_0+1} & \cancel{a_{k_0+2}} & \cancel{a_{k_0+3}} & a_{k_0+4} & a_{k_0+5} & \cancel{a_{k_0+6}} & a_{k_0+7} & \cancel{a_{k_0+8}} & \cancel{a_{k_0+9}} \ \dots \\ & \| & & & \| & \| & & \| & & \\ & a_{\varphi(0)} & & & a_{\varphi(1)} & a_{\varphi(2)} & & a_{\varphi(3)} & & \end{array}$$

$\varphi(0) = k_0+1,\ \varphi(1) = k_0+4, \dots$

Ex: Let $a_m = (-1)^m$ for $m \in \mathbb{N}$ be a sequence in $\mathbb{R}$, then $a_{2m+1} = -1$ is a subsequence of (a_n)

$$\cancel{1}\ \enclose{circle}{-1}\ \cancel{1}\ \enclose{circle}{-1}\ \cancel{1}\ \enclose{circle}{-1}\ \dots$$

→ the contrapositive may be useful.

Lemma: if $\lim\limits_{k \to +\infty} a_k = L$ then $\lim\limits_{j \to +\infty} a_{\varphi(j)} = L$ (for any subsequence)

"Any subsequence of a convergent sequence converges to the same limit".

Theorem: A bounded sequence $(a_k)_{k \geq k_0}$ in $\mathbb{R}^m$ admits a convergent subsequence.

Δ
- the first component a_{k1} of a_k is bounded so by the real case $\exists \varphi_1$ s.t. $a_{\varphi_1(j),1}$ is convergent
- we repeat the process to the second component of $(a_{\varphi_1(j)})$, then we get φ_2 s.t. the first two components of $(a_{\varphi_1(\varphi_2(j))})$ are CV
- And so on

□

Ex: $a_m = (-1)^m$ is bounded, not CV, but $a_{2m} = 1$ CV

Compactness

Def: A subset $S \subset \mathbb{R}^m$ is compact if any sequence with elements in S admits a subsequence which is convergent in S

↳ the limit of the subsequence is in S

⚠ That's not the usual definition, but it's equivalent for $\mathbb{R}^m$

Theorem (Bolzano-Weierstrass) S compact $\Leftrightarrow$ S closed + bounded

Δ $\Leftarrow$ Assume that S is closed and bounded and let (a_k) be a sequence with values in S.
Then (a_k) is bounded and admits a CV subsequence $(a_{\varphi(j)})$ with limit L. ↳ by the last theorem

$L \in \overline{S} = S$ → since S is closed

↳ by a previous theorem

$\Rightarrow$ by contrapositive

1st case: S is not bounded

$\forall k \in \mathbb{N}, \exists a_k \in S, \|a_k\| > k$

then any subsequence of (a_k) satisfies $\|a_{\varphi(j)}\| \xrightarrow[j \to +\infty]{} +\infty$

2nd case: S is not closed, ie $S \subsetneq \overline{S}$

$\exists L \in \overline{S} \setminus S$

$\forall k \in \mathbb{N}, \exists a_k \in B(L, 1/k) \cap S$

then $\|a_k - L\| < \frac{1}{k} \xrightarrow[k \to \infty]{} 0$ so (a_k) converges to $L \notin S$
and any subsequence converges to $L \notin S$ □

Theorem: The continuous image of a compact set is compact.
ie: if $f : \mathbb{R}^m \to \mathbb{R}^k$ is continuous and $S \subset \mathbb{R}^m$ compact then $f(S) \subset \mathbb{R}^k$ compact

Δ let (a_k) be a sequence in $f(S)$, then $a_k = f(b_k)$ for $b_k \in S$

Next, (b_k) admits a CV subsequence $(b_{\varphi(j)})$ with limit $L \in S$

Homework: $\lim_{j \to +\infty} a_{\varphi(j)} = f(L) \in f(S)$ □

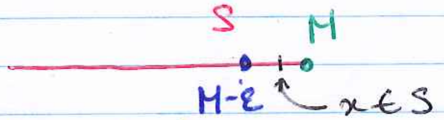
Homework: questions in section 1.3 of the online notes

→ A $\overset{\text{nonempty}}{\text{compact}}$ set of $\mathbb{R}$ admits a supremum / infimum (which is in S)!

Proposition: If $\underset{\neq \emptyset}{S} \subset \mathbb{R}$ is compact then $\sup S \in S$
(Δ n=1) and $\inf S \in S$

Δ ① Since S is bounded, $M = \sup S$ exists by the LUB principle

② Let $\varepsilon > 0$. Then $M - \varepsilon$ is not an upper bound of S so there exists $x \in S$ s.t. $M - \varepsilon < x$



so $S \cap (M-\varepsilon, M+\varepsilon) \neq \emptyset$

Hence $\sup S = M \in \bar{S} = S$ since S is closed □

Corollary (EVT) Let $K \subset \mathbb{R}^m$ be a compact set and $f : K \longrightarrow \mathbb{R}$ be a continuous function.

Then f has a min and a max
ie: $\exists c, d \in K, \forall x \in K, \; f(c) \leq f(x) \leq f(d)$

Δ By a previous theorem $f(K)$ is compact.

Since $f(K) \subset \mathbb{R}$, by the previous proposition $\exists m, M \in f(K)$ such that $\forall x \in K, \; m \leq f(x) \leq M$.

Since $m \in f(K)$, $\exists c \in K$ st $m = f(c)$, similarly for M □

The EVT from MAT137 is a particular case of the above since $[a,b]$ a segment line is compact.

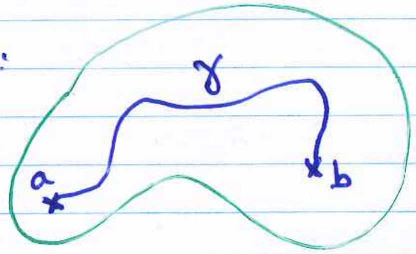
Homework: questions from 1.4 of the online notes

The IVT:

Def. A subset $S \subset \mathbb{R}^n$ is path-connected if $\forall a, b \in S$

$$\exists \gamma : [0,1] \to \mathbb{R}^n \text{ continuous s.t. } \begin{cases} \gamma(0) = a \\ \gamma(1) = b \\ \forall t \in [0,1],\ \gamma(t) \in S \end{cases}$$

Ex:



is path-connected

Ex:

is not path-connected

Lemma: the path-connected subsets of $\mathbb{R}$ are the intervals

Δ • An interval is path-connected: let I be an interval and $a, b \in I$ and set $\gamma(t) = (1-t)a + tb$ then γ is continuous and $\begin{cases} \gamma(0) = a \\ \gamma(1) = b \\ \forall t \in [0,1] \\ \gamma(t) \in I \end{cases}$

• A path-connected ~~set~~ subset of $\mathbb{R}$ is an interval: let $S \subset \mathbb{R}$ be path-connected. let $a, c, b \in \mathbb{R}$ s.t. $a < c < b$ and $a, b \in S$.
let γ be a path from a to b.
By the IVT (MAT137), $\exists t_0$ s.t. $\gamma(t_0) = c$ so $c \in S$

$\square$

Theorem: the continuous image of a path-connected set is path-connected
i.e: if $f : \mathbb{R}^n \to \mathbb{R}^k$ is continuous and $S \subset \mathbb{R}^n$ is path-connected then $f(S) \subset \mathbb{R}^k$ is too

Δ let $a, b \in f(S)$. then $a = f(\alpha)$ and $b = f(\beta)$ with $\alpha, \beta \in S$
since S is path-connected, $\exists \tilde{\gamma}$ a path from α to β

then $\gamma = f \circ \tilde{\gamma}$ is a path from a to b

$\square$

Cor (IVT): $f: S \to \mathbb{R}$ continuous and $S \subset \mathbb{R}^m$ path-connected.

if $\exists a, b \in S, \exists t \in \mathbb{R}$ s.t. $f(a) < t < f(b)$

then $\exists c \in S$ s.t. $f(c) = t$

Δ We know that $f(S)$ is path-connected and $\subset \mathbb{R}$ so it is an interval.
Hence $t \in f(S)$ $\square$

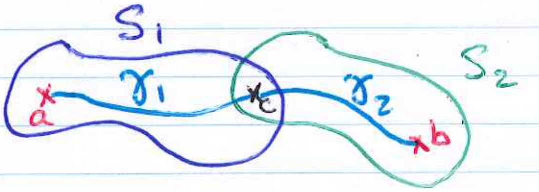
Proposition: Let $S_1, S_2 \subset \mathbb{R}^m$ be path-connected

If $S_1 \cap S_2 \neq \phi$ then $S_1 \cup S_2$ is path-connected

↳ ⚠ 

each is path-connected but not their union

Δ



$S = S_1 \cup S_2$

Let $a \in S_1$, $b \in S_2$ we want a path γ on S from a to b
Let $c \in S_1 \cap S_2$
Let $\gamma_1: [0,1] \to \mathbb{R}^m$ be a path from a to c in S_1
and $\gamma_2: [0,1] \to \mathbb{R}^m$ ―――――――― c to b in S_2

$$\text{then } \gamma(t) = \begin{cases} \gamma_1(2t) & \text{if } t \in [0, 1/2] \\ \gamma_2(2t-1) & \text{if } t \in [1/2, 1] \end{cases}$$

is a path from a to b in S $\square$

Homework: questions from 1-5 in the online notes

DIFFERENTIABILITY

The real-valued case

Def: Let $U \subset \mathbb{R}^n$ open, $f : U \to \mathbb{R}$, $x \in U$, $v \in \mathbb{R}^n$

The directional derivative of f at x along v is

$$\partial_v f(x) := \lim_{t \to 0} \frac{f(x+tv) - f(x)}{t}$$ whenever it exists

Remark: since U is open, $x + tv \in U$ for t "small enough".

Remark: if $\alpha \in \mathbb{R}$ then $\partial_{\alpha v} f(x) = \alpha \partial_v f(x)$.

Hence, if we know $\partial_v f(x)$ for some v, we know the directional derivatives for all the vectors with same direction.

Intuitively: by the above remark, we may assume that $\|v\| = 1$ and then $\partial_v f(x)$ is the instantaneous rate of change of f through x along the direction of v.

Ex: let $f(x,y) = x\cos(y) + ye^x$ and $v = (\cos\theta, \sin\theta)$

Then
$$\partial_v f(0,0) = \lim_{t \to 0} \frac{f((0,0) + t(\cos\theta, \sin\theta)) - f(0,0)}{t}$$

$$= \lim_{t \to 0} \frac{f(t\cos\theta, t\sin\theta)}{t}$$

$$= \lim_{t \to 0} \cos\theta \cos(t\sin\theta) + \sin\theta\, e^{t\cos\theta}$$

$$= \cos(\theta) + \sin(\theta)$$

The highest rate of change through $(0,0)$ is along $\theta = \frac{\pi}{4}$ and the lowest at $\theta = \frac{5\pi}{4}$

Homework: plot the graph on MathSageCell.

Prop: $U \subset \mathbb{R}^n$ open, $f, g : U \to \mathbb{R}$, $x \in U$, $v \in \mathbb{R}^n$, $c \in \mathbb{R}$

If $\partial_v f(x)$ and $\partial_v g(x)$ exist then $\partial_v (f+g)(x)$, $\partial_v (cf)(x)$, $\partial_v (fg)(x)$ exist and

① $\partial_v (f+g)(x) = \partial_v f(x) + \partial_v g(x)$

② $\partial_v (cf)(x) = c\, \partial_v f(x)$

③ $\partial_v (fg)(x) = g(x) \cdot \partial_v f(x) + f(x) \cdot \partial_v g(x)$ "Leibniz rule"

Δ ① & ②: obvious

③: $$\lim_{t \to 0} \frac{f(x+tv)\, g(x+tv) - f(x) g(x)}{t}$$

$$= \lim_{t \to 0} \frac{f(x+tv)\, g(x+tv) - f(x)\, g(x+tv) + f(x) g(x+tv) - f(x) g(x)}{t}$$

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

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

Remark: $g(x+tv) = \frac{g(x+tv) - g(x)}{t} \cdot t + g(x) \xrightarrow[t \to 0]{} 0 + g(x)$

$\square$

Prop: $U \subset \mathbb{R}^n$ open, $x \in U$, $f : U \to \mathbb{R}$, $h : I \to \mathbb{R}$, $v \in \mathbb{R}^n$

If $\partial_v f(x)$ exists and h differentiable at $f(x)$ then $\partial_v (h \circ f)(x)$ exists and $\partial_v (h \circ f)(x) = h'(f(x)) \cdot \partial_v f(x)$

Δ Later, but you can adapt the proof of the chain rule from MAT137 $\square$

Def: Let $U \subset \mathbb{R}^m$ open, $x \in U$, $i = 1, \dots, m$

We define the *i-th partial derivative of f at x* by

$$\frac{\partial f}{\partial x_i}(x) := \partial_{e_i} f(x) = \lim_{t \to 0} \frac{f(x+te_i) - f(x)}{t}$$

whenever it exists.

In practice: $f(x+te_i) = f(x_1, x_2, \dots, x_{i-1}, x_i+t, x_{i+1}, \dots, x_m)$

so $\frac{\partial f}{\partial x_i}(x) = g'(x_i)$ where $g(x_i) = f(x_1, \dots, x_i, \dots, x_m)$

where all the other variables are frozen

Ex: $f(x,y) = x^2 e^{xy}$

$\frac{\partial f}{\partial x}(x,y) = 2x e^{xy} + x^2 y e^{xy}$ $\qquad \frac{\partial f}{\partial x}(1,1) = 3e$

$\frac{\partial f}{\partial y}(x,y) = x^3 e^{xy}$ $\qquad \frac{\partial f}{\partial y}(1,1) = e$

Ex: $f(x,y) = |x|(1+y)$

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

$\frac{\partial f}{\partial y}(0,0) = 0$

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

$\frac{\partial f}{\partial x}(0,0) = 0$, $\frac{\partial f}{\partial y}(0,0) = 0$ exist

but $\partial_{(1,1)} f(0,0) = \lim_{t \to 0} \frac{f(t,t)}{t} = \lim_{t \to 0} \frac{1}{2t}$ DNE

⚠ So: the partial derivatives exist $\not\Rightarrow$ all the directional derivatives exist

Def: $\mathcal{U} \subset \mathbb{R}^m$ open , $f : \mathcal{U} \to \mathbb{R}$, $x \in \mathcal{U}$, $\frac{\partial f}{\partial x_i}(x)$ exists $\forall i$

The gradient of f at x is

$$\nabla f(x) := \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), -, \frac{\partial f}{\partial x_m}(x) \right) \in \mathbb{R}^m$$

Ex: $f(x,y) = x\cos(y) + y e^x$

$$\nabla f(x,y) = \left(\cos(y) + y e^x , -x\sin(y) + e^x \right)$$

$$\nabla f(0,0) = (1,1)$$

Remark: We have already seen that $\overset{\nabla f(0,0)}{\overset{\shortparallel}{(1,1)}}$ $(\theta = \frac{\pi}{4})$ was the direction where f has a the highest rate of change through $(0,0)$.

We will see later that it is a general phenomenon

Def: $\mathcal{U} \subset \mathbb{R}^m$ open , $f : \mathcal{U} \to \mathbb{R}$, $x \in \mathcal{U}$

We say that f is differentiable at x if there exists a linear function $d_x f : \mathbb{R}^m \to \mathbb{R}$ s.t.

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

$d_x f(\lambda u + \mu v) = \lambda d_x f(u) + \mu d_x f(v)$

i.e. $f(x+h) = f(x) + d_x f(h) + E(h)$ where $\frac{E(h)}{\|h\|} \to 0$

"linear approximation of f at x"

Ex: ① $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, $f(x+h) = f(x) + 2xh + h^2$

so f is differentiable on $\mathbb{R}$ and $d_x f(h) = 2xh$

② $g : \mathbb{R}^m \to \mathbb{R}$, $g(x) = \|x\|^2$, $g(x+h) = \|x+h\|^2 = \|x\|^2 + 2(x \cdot h) + \|h\|^2$

so g is differentiable on $\mathbb{R}^m$ and $d_x g(h) = 2(x \cdot h)$

Theorem: If f is differentiable at x then its differential d_xf is unique

Δ Assume that f has 2 differentials at x: $\ell_1, \ell_2 : \mathbb{R}^n \longrightarrow \mathbb{R}$

Let $\underset{\neq \vec{0}}{h \in} \mathbb{R}^n$. Since U is open, $x+th \in U$ when $t^{>0}$ is small.

Then: $\ell_1(h) - \ell_2(h) = \dfrac{\ell_1(th) - \ell_2(th)}{t}$

$$= - \frac{f(x+th) - f(x) - \ell_1(th)}{t\|h\|}\|h\| + \frac{f(x+th) - f(x) - \ell_2(th)}{t\|h\|}\|h\|$$

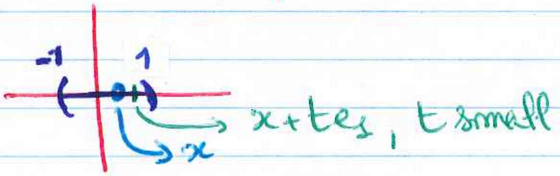
$$\xrightarrow[t \to 0^+]{} 0$$

Hence $\ell_1(h) = \ell_2(h)$ □

⚠ That's why we want the domain to be open. Otherwise the differential could not be unique.

Ex: $U = (-1,1) \times \{0\}$

We can determine $d_xf(e_1)$:



But we have no condition for $d_xf(e_2)$ since $x+te_2 \notin U$ $\forall t \neq 0$ so we can take whatever we want.

Prop: $U \subset \mathbb{R}^n$ open, $f: U \to \mathbb{R}$, $x \in U$

If f is differentiable at x then f is continuous at x.

Δ $f(v) = f(x + (v-x)) = f(x) + d_xf(v-x) + E(v-x) \xrightarrow[v \to x]{} f(x) + 0 + 0$ □

Prop: $f, g : U \longrightarrow \mathbb{R}$ differentiable at $x \in U$, $U \in \mathbb{R}^n$ open, then

① $d_x(f+g) = d_xf + d_xg$
② $d_x(\lambda f) = \lambda\, d_xf$, $\lambda \in \mathbb{R}$
③ $d_x(fg) = g(x)\, d_xf + f(x)\, d_xg$
④ $d_x(1/f) = -1/f(x)^2 \cdot d_xf$ if $f(x) \neq 0$

Theorem: $U \subset \mathbb{R}^n$ open, $f : U \longrightarrow \mathbb{R}$, $x \in U$

If f is differentiable at x, then

① All the directional derivatives of f at x exist

② $\forall v \in \mathbb{R}^n,\ \partial_v f(x) = d_x f(v)$

③ $\forall h \in \mathbb{R}^n,\ d_x f(h) = \nabla f(x) \cdot h = \frac{\partial f(x)}{\partial x_1} h_1 + \dots + \frac{\partial f}{\partial x_n}(x) \cdot h_n$

$\downarrow$

(that's why physicists write $df = \sum_i \frac{\partial f}{\partial x_i} dx_i$)

④ $\forall v \in \mathbb{R}^n,\ \partial_v f(x) = \nabla f(x) \cdot v$

Δ ①+②:

$$\frac{f(x+tv)-f(x)}{t} = \frac{f(x+tv)-f(x)-d_xf(tv)+d_xf(tv)}{t}$$

$$= \frac{f(x+tv)-f(x)-d_xf(tv)}{t\,\|v\|}\|v\| + d_xf(v)$$

$$\xrightarrow[t\to 0]{} 0 + d_xf(v)$$

So $\partial_v f(x)$ exists and $\partial_v f(x) = d_x f(v)$

③ $d_x f(e_i) = \partial_{e_i} f(x) = \frac{\partial f}{\partial x_i}(x)$

Hence $d_x f(h_1, \dots, h_n) = d_x f(\sum_i h_i e_i) = \sum_i h_i\, d_x f(e_i) = \sum_i h_i \frac{\partial f}{\partial x}(e_i) = \nabla f(x) \cdot h$

④ $\partial_v f(x) = d_x f(v) = \nabla f(x) \cdot v$ □

Remark: if f is differentiable at x then $d_x f(h) = \nabla f(x) \cdot h$, there is no other possibility

Ex: $f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$ is not differentiable at $(0,0)$ since $\partial_{(1,1)} f(0,0)$ DNE

⚠ The converse of the above theorem is false:

$f(x,y) = \begin{cases} \frac{x^3}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$: all the directional derivatives exist at $(0,0)$ but f is not differentiable at $(0,0)$

- $\partial_v f(0,0) = \lim_{t\to 0} \frac{t^3 v_1^3}{t^3 v_1^2 + t^3 v_2^2} = \frac{v_1^3}{v_1^2+v_2^2}$
- Assume by contradiction that f is differentiable at 0, then

$$d_0 f(1,1) = d_0 f(1,0) + d_0 f(0,1) = \partial_{(1,0)} f(0,0) + \partial_{(0,1)} f(0,0) = 1 + 0 = 1$$
$$= \partial_{(1,1)} f(0,0) = 1/2$$

Remark: If $f: \mathcal{U} \to \mathbb{R}$ is differentiable at x and $\nabla f(x) \neq \vec{0}$ then the direction of $\nabla f(x)$ is the direction of fastest increase at x and the magnitude of $\nabla f(x)$ is the instantaneous rate of change in that direction.

Δ Let $v \in \mathbb{R}^m$ be a unit vector then

$$\partial_v f(x) = d_x f(v) = \nabla f(x) \cdot v = |v| \cdot |\nabla f(x)| \cdot \cos\theta = |\nabla f(x)| \cdot \cos\theta$$

the max is when $\cos\theta = 1$, ie $v = \frac{\nabla f(x)}{|\nabla f(x)|}$ and then $\partial_v f(x) = |\nabla f(x)|$ □

Compare with the examples about $f(x,y) = x\cos(y) + ye^x$ above.

Theorem: $\mathcal{U} \subset \mathbb{R}^m$ open, $f: \mathcal{U} \to \mathbb{R}$, $x \in \mathcal{U}$ at least in an open ball around x

If the partial derivatives of f exist on $\mathcal{U}$ and are continuous at x then f is differentiable at x

Δ ~~Idea~~: we apply the MVT to $t \mapsto f(t, x_2, \dots, x_m)$ on $[x_1, x_1+h_1]$

then $\exists \theta_1$ st. $f(x_1+h_1, x_2, \dots, x_m) - f(x_1, \dots, x_m) = h_1 \frac{\partial f}{\partial x_1}(x_1+\theta_1 h_1, x_2, \dots, x_m)$

By MVT to $t \mapsto f(x_1+h_1, t, x_3, \dots, x_m)$ on $[x_2, x_2+h_2]$

$\exists \theta_2$ st. $f(x_1+h_1, x_2+h_2, x_3, \dots, x_m) - f(x_1+h_1, x_2, \dots, x_m) = h_2 \frac{\partial f}{\partial x_2}(x_1+h_1, x_2+\theta_2 h_2, \dots$

And so on for $x_3, \dots, x_m$.

$$\text{Then: } f(x+h) - f(x) - \sum_i h_i \frac{\partial f}{\partial x_i}(x) = \sum_i h_i \Big(\overbrace{\frac{\partial f}{\partial x_i}(a_i) - \frac{\partial f}{\partial x_i}(x)}^{\xrightarrow[h\to 0]{} 0 \text{ by continuity of } \frac{\partial f}{\partial x_i}} \Big)$$

where $a_i = (x_1+h_1, \dots, x_i+h_i\theta_i, x_{i+1}, \dots, x_m) \xrightarrow[h\to 0]{} x$ □

The converse is false: $f(x) = \begin{cases} x^2\sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{otherwise} \end{cases}$

is differentiable on $\mathbb{R}$ but f' is not continuous at 0

Summary: $f : \mathcal{U} \longrightarrow \mathbb{R}$, $\mathcal{U} \subset \mathbb{R}^m$ open

Name	Nature	Notation
Directional derivative at $x \in \mathcal{U}$ along $v \in \mathbb{R}^m$	Real number	$\partial_v f(x)$
Partial derivative at x	Real number	$\frac{\partial f}{\partial x_i}(x)$
Gradient at x	Vector of $\mathbb{R}^m$	$\nabla f(x)$
Differential at x	Linear function $\mathbb{R}^m \to \mathbb{R}$	$d_x f(h)$

$$\begin{tikzcd}[column sep=large]
 & & \text{Continuous at } x \arrow[dl, Rightarrow, shift left, "\times" description, "ⓑ"'] \\
\text{Partial derivatives exist on } \mathcal{U} \text{ and are continuous at } x \arrow[r, Rightarrow, shift left] & \text{Differentiable at } x \arrow[l, Rightarrow, shift left, "\times" description, "ⓐ"] \arrow[ur, Rightarrow, shift left] \arrow[dr, Rightarrow, shift left] & \\
 & & \text{Directional derivatives exist at } x \text{ and} \arrow[ul, Rightarrow, shift left, "\times" description, "ⓒ"]
\end{tikzcd}$$

① $\partial_v f(x) = d_x f(v)$

② $d_x f(h) = \nabla f(x) \cdot h$

$= \sum_{i=1}^{m} \frac{\partial f}{\partial x_i}(x)\, h_i$

③ $\partial_v f(x) = \nabla f(x) \cdot v$

useful to prove that a function is not differentiable. See ⓒ

Counter-examples:

ⓐ $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$

ⓑ $f(x) = |x|$

ⓒ $f(x,y) = \begin{cases} \frac{x^3}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$ on $\mathbb{R}^2$

Partial derivatives exist $\not\Rightarrow$ Directional derivatives exist

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

DIFFERENTIABILITY

Linear maps and matrices (Recollection)

Def: A map $\varphi : \mathbb{R}^m \longrightarrow \mathbb{R}^k$ is linear if

- $\forall u, v \in \mathbb{R}^m, \ \varphi(u+v) = \varphi(u) + \varphi(v)$
- $\forall u \in \mathbb{R}^m, \forall \lambda \in \mathbb{R}, \ \varphi(\lambda u) = \lambda \varphi(u)$

Notation: $e_i^m = \underbrace{(0, \dots, 0, \underset{\hookrightarrow \text{ ith component}}{1}, 0, \dots, 0)}_{m \text{ components}}$

So that $(e_i^m)_{i=1 \dots m}$ is the standard basis of $\mathbb{R}^m$

Remark: A linear map $\varphi : \mathbb{R}^m \longrightarrow \mathbb{R}^k$ is entirely determined by the values $\varphi(e_i^m)$:

Let $\underset{(u_1, \dots, u_m)}{\overset{\shortparallel}{u}} \in \mathbb{R}^m$, then $u = \sum_{i=1}^m u_i e_i^m$ and $\varphi(u) = \sum_{i=1}^m u_i \varphi(e_i^m)$

Def: We denote by $(a_{ij})_{i=1 \dots k}$ the components of $\varphi(e_j^m)$, where $\varphi : \mathbb{R}^m \to \mathbb{R}^k$ is linear

ie: $\varphi(e_j^m) = \sum_{i=1}^k a_{ij} e_i^k$

The matrix of φ (in the standard bases) is

$$\mathrm{Mat}(\varphi) := \begin{array}{c} \begin{array}{cccc} \varphi(e_1^m) & \varphi(e_2^m) & & \varphi(e_m^m) \\ \downarrow & \downarrow & & \downarrow \end{array} \\ \underbrace{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{km} \end{pmatrix}}_{n \text{ columns}} \end{array} \left.\vphantom{\begin{pmatrix} a \\ a \\ a \\ a \\ a \end{pmatrix}}\right\} k \text{ rows} \in M_{k,m}(\mathbb{R})$$

Remark: φ is entirely determined by the above matrix.

Def: $M = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{km} \end{pmatrix} \in M_{km}(\mathbb{R})$, $N = \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & & \vdots \\ b_{k1} & \cdots & b_{km} \end{pmatrix} \in M_{km}(\mathbb{R})$

then $M + N := \begin{pmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{k1} & \cdots & c_{km} \end{pmatrix} \in M_{km}(\mathbb{R})$

where $c_{ij} = a_{ij} + b_{ij}$

Def: $M = \begin{pmatrix} a_{11} & \cdots & a_{1l} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kl} \end{pmatrix} \in M_{kl}(\mathbb{R})$, $N = \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{l1} & \cdots & b_{lm} \end{pmatrix} \in M_{lm}(\mathbb{R})$

$$MN := \begin{pmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{k1} & \cdots & c_{km} \end{pmatrix} \in M_{km}(\mathbb{R})$$

where $c_{ij} = \sum_{\kappa=1}^{l} a_{i\kappa} b_{\kappa j}$

⚠ $M_{kl} \times M_{lm} \longrightarrow M_{km}$

the number of columns of the first matrix must be equal to the number of lines of the second matrix.

Prop: Let $\varphi : \mathbb{R}^m \to \mathbb{R}^l$ and $\psi : \mathbb{R}^l \to \mathbb{R}^k$ be two linear maps then the matrix of the linear map $\psi \circ \varphi : \mathbb{R}^m \to \mathbb{R}^k$ is

$$\text{Mat}(\psi \circ \varphi) = \text{Mat}(\psi)\ \text{Mat}(\varphi)$$

Prop: Let $\varphi : \mathbb{R}^m \to \mathbb{R}^k$ linear, $v = \sum_{i=1}^{m} v_i e_i^m \in \mathbb{R}^m$

then $\varphi(v) = \sum_{i=1}^{k} c_i e_i^k$ where

$$\begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} = \underset{\underset{M_{km}(\mathbb{R})}{\cap}}{\text{Mat}(\varphi)} \underset{\underset{M_{m1}(\mathbb{R})}{\cap}}{\begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}} \quad \in M_{k1}(\mathbb{R})$$

DIFFERENTIABILITY

Vector-valued functions

Def: $U \subset \mathbb{R}^m$ open set, $f: U \longrightarrow \mathbb{R}^k$, $x_0 \in U$.

We say that f is differentiable at x_0 if there exists a linear map $d_{x_0}f: \mathbb{R}^m \longrightarrow \mathbb{R}^k$ s.t.

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

or equivalently $f(x_0+h) = f(x_0) + d_{x_0}f(h) + E(h)$

with $\lim_{h \to 0} \frac{E(h)}{\|h\|} = 0$

$d_{x_0}f: \mathbb{R}^m \longrightarrow \mathbb{R}^k$ is called the differential or total derivative of f at x_0

Def: $U \subset \mathbb{R}^m$ open set, $f = (f_1, \dots, f_k): U \longrightarrow \mathbb{R}^k$, $x_0 \in U$

Assume that all the partial derivatives $\left(\frac{\partial f_i}{\partial x_j}(x_0)\right)_{\substack{i=1,\dots,k \\ j=1,\dots,m}}$ exist. The we define the Jacobian matrix of f at x_0 by

$$\underbrace{Df(x_0)} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \cdots & \frac{\partial f_1}{\partial x_m}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1}(x_0) & \cdots & \frac{\partial f_k}{\partial x_m}(x_0) \end{pmatrix}$$

Notation used in the textbook.
The notation $J_f(x_0)$ is very common

Theorem: $\mathcal{U} \subset \mathbb{R}^m$ open, $f = (f_1, \dots, f_k) : \mathcal{U} \longrightarrow \mathbb{R}^k$, $x_0 \in \mathcal{U}$.

f is differentiable at $x_0 \Leftrightarrow \forall i$, f_i is differentiable at x_0.

Moreover, if the above holds,

$$d_{x_0} f = (d_{x_0} f_1, \dots, d_{x_0} f_k) : \mathbb{R}^m \longrightarrow \mathbb{R}^k$$

Δ Notice that componentwise, if $\ell : \mathbb{R}^m \to \mathbb{R}^k$ is linear,

$$f(x_0 + h) = f(x_0) + \ell(h) + \varepsilon(h)$$

becomes

$$\begin{pmatrix} f_1(x_0+h) \\ \vdots \\ f_k(x+h) \end{pmatrix} = \begin{pmatrix} f_1(x_0) \\ \vdots \\ f_k(x_0) \end{pmatrix} + \begin{pmatrix} \ell_1(h) \\ \vdots \\ \ell_k(h) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(h) \\ \vdots \\ \varepsilon_k(h) \end{pmatrix}$$

and $\frac{1}{\|h\|} \varepsilon(h) = \begin{pmatrix} \frac{\varepsilon_1(h)}{\|h\|} \\ \vdots \\ \frac{\varepsilon_k(h)}{\|h\|} \end{pmatrix}$

$\square$

Hence, it is enough to understand well the real-valued case:

Theorem: $\mathcal{U} \subset \mathbb{R}^m$ open, $f = (f_1, \dots, f_k) : \mathcal{U} \longrightarrow \mathbb{R}^k$, $x_0 \in \mathcal{U}$

If f is differentiable at x_0 then all the directional derivatives $\partial_v f_i(x_0)$ exist ($i = 1, \dots, k$, $v \in \mathbb{R}^m$)

And

$$\mathrm{Mat}(d_{x_0} f) = D(f)(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \cdots & \frac{\partial f_1}{\partial x_m}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1}(x_0) & \cdots & \frac{\partial f_k}{\partial x_m}(x_0) \end{pmatrix}$$

Δ We apply the result from the real-valued case to $f_1, \dots, f_k$ and then use the previous theorem to get that

$$d_{x_0} f(e_i^n) = \left(d_{x_0} f_1(e_i^n), —, d_{x_0} f_k(e_i^n)\right)$$

$$= \left(\frac{\partial f_1}{\partial x_i}(x_0), —, \frac{\partial f_k}{\partial x_i}(x_0)\right)$$

Hence

$$\mathrm{Mat}(d_{x_0} f) = \left(\begin{array}{cccc} d_{x_0} f(e_1^n) & d_{x_0} f(e_2^n) & & d_{x_0} f(e_n^n) \\ \downarrow & \downarrow & & \downarrow \\ \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) & — & \frac{\partial f_1}{\partial x_n}(x_0) \\ | & | & \diagdown & | \\ \frac{\partial f_k}{\partial x_1}(x_0) & \frac{\partial f_k}{\partial x_2}(x_0) & — & \frac{\partial f_k}{\partial x_n}(x_0) \end{array}\right)$$

□

<u>Theorem</u>: $U \subset \mathbb{R}^n$ open, $f = (f_1, -, f_k) : U \longrightarrow \mathbb{R}^k$, $x_0 \in U$

If all the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist on U and are continuous at x_0 then f is differentiable at x_0.

Δ We apply the theorem from the real-valued case to $f_1, -, f_k$ □

Theorem: If f is differentiable at x_0 then f is continuous at x_0

Proof: apply the result from the real-valued case to the components. QED

Ex: $f=(f_1,\dots,f_k):(a,b)\longrightarrow \mathbb{R}^k$

f is differentiable at $t_0\in(a,b)$ iff $f_1'(t_0),\dots,f_k'(t_0)$ exist and then

$$Df(t_0)=\begin{pmatrix} f_1'(t_0)\\ \vdots \\ f_k'(t_0)\end{pmatrix}$$

In this case it is common to use the notation $f'(t_0)$ instead of $Df(t_0)$.

Comment: if f is differentiable at t_0 and $f'(t_0)\neq\vec{0}$ then:

① $f'(t_0)$ is tangent to the parametrized curve at $f(t_0)$

② $\theta(h)=f(t_0)+hf'(t_0)$, $h\in\mathbb{R}$, parametrizes the tangent line ~~at~~ of the parametrized curve at $f(t_0)$

See examples 2 and 3 in Section 2.2.

Ex: $U\subset\mathbb{R}^m$ open, $f:U\longrightarrow\mathbb{R}$, $x_0\in U$

If f is differentiable at x_0 then

$$Df(x_0)=\left(\frac{\partial f}{\partial x_1}(x_0)\ \cdots\ \frac{\partial f}{\partial x_m}(x_0)\right)$$

Notice that $Df(x_0)\cdot\begin{pmatrix}h_1\\ \vdots\\ h_m\end{pmatrix}=\sum_{i=1}^{m}\frac{\partial f}{\partial x_i}(x_0)\cdot h=\nabla f(x_0)\cdot h$

We recover the gradient from the previous section.

But be careful:

$Df(x_0)$ is the $1\times m$ matrix of a linear map

$\nabla f(x_0)$ is a vector of $\mathbb{R}^m$

DIFFERENTIABILITY

The chain rule.

Theorem: $\mathcal{U} \subset \mathbb{R}^n$ open, $f : \mathcal{U} \longrightarrow \mathbb{R}^\ell$, $x_0 \in \mathcal{U}$
$\mathcal{V} \subset \mathbb{R}^\ell$ open, $g : \mathcal{V} \longrightarrow \mathbb{R}^k$.

Assume that $f(\mathcal{U}) \subset \mathcal{V}$ so that $g \circ f : \mathcal{U} \longrightarrow \mathbb{R}^k$ is well-defined

If $\begin{cases} f \text{ is differentiable at } x_0 \\ g \text{ is differentiable at } f(x_0) \end{cases}$ then $\begin{cases} g\circ f \text{ is differentiable at } x_0 \text{ and} \\ d_{x_0}(g\circ f) = (d_{f(x_0)} g) \circ (d_{x_0} f) \end{cases}$

Δ $f(x_0+h) = f(x_0) + d_{x_0}f(h) + \|h\|\,\varepsilon_1(h)$, $\varepsilon_1(h) \xrightarrow[h\to 0]{} 0$

$g(f(x_0)+k) = g(f(x_0)) + d_{f(x_0)}g(k) + \|k\|\,\varepsilon_2(k)$, $\varepsilon_2(k) \xrightarrow[k\to 0]{} 0$

Hence
$$\begin{aligned} g\circ f(x_0+h) &= g(f(x_0+h)) \\ &= g(f(x_0) + d_{x_0}f(h) + \|h\|\,\varepsilon_1(h)) \\ &= g(f(x_0)) + d_{f(x_0)}g\big(d_{x_0}f(h) + \|h\|\,\varepsilon_1(h)\big) \\ &\quad + \|d_{x_0}f(h) + \|h\|\varepsilon_1(h)\| \cdot \varepsilon_2\big(d_{x_0}f(h) + \|h\|\,\varepsilon_1(h)\big) \\ &= g(f(x_0)) + d_{f(x_0)}g(d_{x_0}f(h)) \\ &\quad + \|h\|\Big(\underbrace{d_{f(x_0)}g(\varepsilon_1(h))}_{\xrightarrow[h\to 0]{} 0} + \underbrace{\|d_{x_0}f(\tfrac{h}{\|h\|}) + \varepsilon_1(h)\|}_{\text{bounded}} \cdot \underbrace{\varepsilon_2(\overbrace{d_{x_0}f(h) + \varepsilon_1(h)}^{\xrightarrow[h\to 0]{} 0})}_{\xrightarrow[h\to 0]{} 0}\Big) \end{aligned}$$

bounded:

- $\frac{h}{\|h\|} \in S^1$ compact and $d_{x_0}f$ continuous since linear
- $\varepsilon_1(h) \xrightarrow[h\to 0]{} 0$

since $\varepsilon_2 \xrightarrow[h\to 0]{} 0$

$\square$

Corollary: Under the same assumptions

$$\underbrace{D(g\circ f)(x_0)}_{\text{Jacobian matrix of } g\circ f \text{ at } x_0} = \underbrace{Dg(f(x_0))}_{\text{Jacobian matrix of } g \text{ at } f(x_0)} \cdot \underbrace{Df(x_0)}_{\text{Jacobian matrix of } f \text{ at } x_0}$$

Δ Recall that $Df(x_0) = \mathrm{Mat}(d_{x_0}f)$

and that $\mathrm{Mat}(d_{f(x_0)}g \circ d_{x_0}f) = \mathrm{Mat}(d_{f(x_0)}g)\,\mathrm{Mat}(d_{x_0}f)$ ☐

Remark: If we look at the (i,j)-component, we get:

$$\frac{\partial(g_i\circ f)}{\partial x_j}(x_0) = \sum_{\alpha=1}^{l} \frac{\partial g_i}{\partial y_\alpha}(f(x_0)) \cdot \frac{\partial f_\alpha}{\partial x_j}(x_0)$$

$$\left[\text{Recall that: } f: \begin{array}{ccc} \mathbb{R}^m & \longrightarrow & \mathbb{R}^l \\ x & \longmapsto & f(x) \end{array} \text{ and } g: \begin{array}{ccc} \mathbb{R}^l & \longrightarrow & \mathbb{R}^k \\ y & \longmapsto & g(y) \end{array}\right]$$

Comment: Physicist notation "à la Leibniz":

$$\frac{\partial v_i}{\partial x_j} = \frac{\partial v_i}{\partial y_1} \cdot \frac{\partial y_1}{\partial x_j} + \dots + \frac{\partial v_i}{\partial y_l} \cdot \frac{\partial y_l}{\partial x_j}$$

where $v = g(\overset{y}{\widetilde{f(x)}}) = (g_1(f(x)), \dots, g_k(f(x)))$, $y = f(x)$

$\frac{\partial v_i}{\partial x_j} = \frac{\partial(g_i\circ f)}{\partial x_j}$: we see $v(x) = g\circ f(x)$, as a function of $x \in \mathbb{R}^m$

$\frac{\partial v_i}{\partial y_\alpha} = \frac{\partial g_i}{\partial y_\alpha}$: we see $v(y) = g(y)$ as a function of $y \in \mathbb{R}^l$

$\partial y_\alpha / \partial x_j = \partial f_\alpha / \partial y_j$: $y(x) = f(x)$

Ex: Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ differentiable

Define $\varphi : \mathbb{R}^3 \longrightarrow \mathbb{R}$ by $\varphi(x,y,z) = f(x^2-yz, xyz)$

$$\frac{\partial \varphi}{\partial x}(x,y,z) = 2x \frac{\partial f}{\partial x}(x^2-yz, xyz) + yz \frac{\partial f}{\partial y}(x^2-yz, xyz)$$

$$\frac{\partial \varphi}{\partial y}(x,y,z) = -z \frac{\partial f}{\partial x}(x^2-yz, xyz) + xz \frac{\partial f}{\partial y}(x^2-yz, xyz)$$

$$\frac{\partial \varphi}{\partial z}(x,y,z) = -y \frac{\partial f}{\partial x}(x^2-yz, xyz) + xy \frac{\partial f}{\partial y}(x^2-yz, xyz)$$

$\frac{\partial f}{\partial x}$ is just a notation to say partial derivative w.r.t. first variable

You compute the partial derivative of f w.r.t the first variable and then evaluate it at (x^2-yz, xyz)

Be sure you understand the above computations before continuing...

You worst enemy in calculus is going to be notation:

① There are as many notation as people

eg: $\partial_x, \partial_{e_1}, \partial^{e_1}, \frac{\partial}{\partial x}, f_x, D^{(1,0)}, \ldots$ In MAT237, we'll use $\partial_x f$ or $\frac{\partial f}{\partial x}$.

② Notation can be confusing

eg: $\frac{\partial f}{\partial x}(x^2-yz, xyz)$ means ⓐ compute $\frac{\partial f}{\partial x}$ first partial derivative of f then ⓑ evaluate at (x^2-yz, xyz) ↗ ✓

It does NOT mean: compute $f(x^2-yz, xyz)$ and then differentiate wrt to x → ✗

Be sure that you understand what you are computing

Ex: $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$ differentiable

$S = \{(x,y) \in \mathbb{R}^2 : y \neq 0\}$ open

Define $\varphi(x,y) = f(x, xy, x/y)$

$$\frac{\partial \varphi}{\partial x}(x,y) = \frac{\partial f}{\partial x}(x, xy, x/y) + y\,\frac{\partial f}{\partial y}(x, xy, x/y) + \frac{1}{y}\frac{\partial f}{\partial z}(x, xy, x/y)$$

$$\frac{\partial \varphi}{\partial y}(x,y) = 0 + x\,\frac{\partial f}{\partial y}(x, xy, x/y) - \frac{x}{y^2}\frac{\partial f}{\partial z}(x, xy, x/y)$$

Comment:

$$D\varphi = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ y & x \\ 1/y & -\frac{x}{y^2} \end{pmatrix}$$

We recover the same result !

⚠ It's common to drop the variables during the computations as I did in the comment, in order to lighten the notation.

If you do so: ① Be careful to not forget to add them back at the end

② Keep track of them

Ex: $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$ differentiable

$S = \{(x,y) \in \mathbb{R}^2 : y \neq 0\}$ open

Define $\varphi(x,y) = f(x, xy, x/y)$

$$\frac{\partial \varphi}{\partial x}(x,y) = \frac{\partial f}{\partial x}(x, xy, x/y) + y \frac{\partial f}{\partial y}(x, xy, x/y) + \frac{1}{y}\frac{\partial f}{\partial z}(x, xy, x/y)$$

$$\frac{\partial \varphi}{\partial y}(x,y) = 0 + x \frac{\partial f}{\partial y}(x, xy, x/y) - \frac{x}{y^2}\frac{\partial f}{\partial z}(x, xy, x/y)$$

Comment:

$$D\varphi = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ y & x \\ 1/y & \frac{-x}{y^2} \end{pmatrix}$$

We recover the same result!

⚠ It's common to drop the variables during the computations as I did in the comment, in order to lighten the notation.

If you do so : ① Be careful to not forget to add them back at the end

② Keep track of them

Ex: Polar coordinates

$f : \mathbb{R}^2 \longrightarrow \mathbb{R}$, $S = \{(r,\theta) \in \mathbb{R}^2 : r \geqslant 0\}$

$g : S \longrightarrow \mathbb{R}^2$ defined by $g(r,\theta) = (r\cos\theta, r\sin\theta)$

$\varphi = f \circ g$ so that $\varphi(r,\theta) = f(r\cos\theta, r\sin\theta)$

Ⓐ Chain rule for the Jacobian matrix:

$$
\begin{aligned}
D\varphi(r,\theta) &= Df(g(r,\theta)) \cdot Dg(r,\theta) \\
&= Df(r\cos\theta, r\sin\theta) \cdot Dg(r,\theta) \\
&= \begin{pmatrix} \frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta) & \frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta) \end{pmatrix} \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}
\end{aligned}
$$

$$
\begin{pmatrix} \frac{\partial\varphi}{\partial r}(r,\theta) & \frac{\partial\varphi}{\partial\theta}(r,\theta) \end{pmatrix} = \left(\frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta)\cos\theta + \frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta)\sin\theta, -\frac{\partial f}{\partial x} r\sin\theta + \frac{\partial f}{\partial y} r\cos\theta \right)
$$

Ⓑ Chain rule for the partial derivatives

$$\frac{\partial\varphi}{\partial r}(r,\theta) = \frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta)\cos\theta + \frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta)\sin\theta$$

$$\frac{\partial\varphi}{\partial\theta}(r,\theta) = -\frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta)\, r\sin\theta + \frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta)\, r\cos\theta$$

We read them in the components of the Jacobian matrix

Ⓒ Chain rule for the differentials:

$$
\begin{aligned}
d_{(r,\theta)}\varphi(h,k) &= d_{g(r,\theta)}f \circ d_{(r,\theta)}g(h,k) = d_{g(r,\theta)}f(\cos\theta h - r\sin\theta k, \sin\theta h + r\cos\theta k) \\
&= \frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta)\left(\cos\theta h - r\sin\theta k\right) + \frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta)\left(\sin\theta h + r\cos\theta k\right)
\end{aligned}
$$

Ex : changing the names of the variables

$$f : \begin{array}{ccc} \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x,y) & \longmapsto & f(x,y) \end{array}$$

$\varphi : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by $\varphi(r,s) = f(re^s, rs)$

$$\frac{\partial \varphi}{\partial r}(r,s) = e^s \frac{\partial f}{\partial x}(re^s, rs) + s\frac{\partial f}{\partial y}(re^s, rs)$$

$$\frac{\partial \varphi}{\partial s}(r,s) = re^s \frac{\partial f}{\partial x}(re^s, rs) + r\frac{\partial f}{\partial y}(re^s, rs)$$

Level sets and the gradient

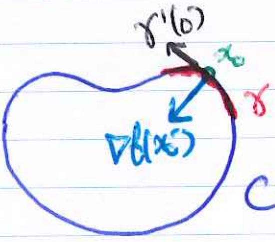
Setup: $\mathcal{U} \subset \mathbb{R}^n$ open, $f : \mathcal{U} \to \mathbb{R}$ differentiable at $x_0 \in \mathcal{U}$

$$C = \{x \in \mathcal{U} : f(x) = f(x_0)\}$$

C is the level set of f at $f(x_0)$

Def: We say that $v \in \mathbb{R}^n$ is $\boxed{\text{tangent}}$ to C at x_0 if there exists $\gamma : I \to \mathbb{R}^n$, $I \subset \mathbb{R}$ open interval, $0 \in I$, such that $\forall t \in I$, $\gamma(t) \in C$, $\gamma(0) = x_0$, $\gamma'(0) = v$

↳ particularly, $\gamma'(0)$ exists



$C = \{x \in \mathcal{U} \text{ st. } f(x) = f(x_0)\} \subset \mathcal{U}$ ⚠

Claim: If v is tangent to C at x_0 then $v \cdot \nabla f(x_0) = 0$

i.e.: $\nabla f(x_0)$ is orthogonal to the level set

Δ Take γ as in the above definition

Define $h : I \to \mathbb{R}$ by $h(t) = f(\gamma(t))$, then

① $\forall t \in I$, $h(t) = f(\gamma(t)) = f(x_0)$ since $\gamma(t) \in C$
$\Rightarrow h'(0) = 0$

② By the chain rule

$$0 \underset{\substack{\uparrow \\ \text{by ①}}}{=} h'(0) = (f \circ \gamma)'(0)$$

$= d_0(f \circ \gamma)(1)$ → Recall that for $g : \overset{\overset{\text{open interval}}{\mathbb{R}}}{\overset{\cup}{I}} \to \mathbb{R}$, $d_{t_0}g(h) = g'(t_0)h$
Hence $g'(t_0) = d_{t_0}g(1)$

$= d_{\gamma(0)}f \circ d_0\gamma(1)$ ↳ chain rule for the differentials

$= d_{\gamma(0)}f(d_0\gamma(1))$

$= d_{x_0}f(v)$ → $\gamma(0) = x_0$, $d_0\gamma(1) = \gamma'(0) \cdot 1 = v$

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

□

Ex: Find the tangent plane of

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

at $a = (1,1,1)$

Δ C is the level set $f(x,y,z) = 2$ for $f(x,y,z) = x^2 - 2xy + 4yz - z^2$

$\nabla f(a) = (0,2,2)$

So the tangent plance of C at a is

$$\{(x,y,z) \in \mathbb{R}^3 : (x-1, y-1, z-1) \cdot (0,2,2) = 0\}$$

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

It has a for equation $y + z = 2$.

$\square$

TODO: Recap slides

Homework: Questions from §2.3

DIFFERENTIABILITY

The Mean Value Theorem

Theorem: (MVT) $\mathcal{U} \subset \mathbb{R}^m$ open, $f : \mathcal{U} \longrightarrow \mathbb{R}$ differentiable on $\mathcal{U}$

Let $a, b \in \mathcal{U}$, assume that $L_{a,b} = \{(1-t)a + tb : t \in [0,1]\} \subset \mathcal{U}$

Then $\exists c \in L_{a,b}$ such that $f(b) - f(a) = \nabla f(c) \cdot (b-a)$

Δ Set $\gamma(t) = (1-t)a + tb$, $t \in [0,1]$

and $\varphi(t) = f(\gamma(t))$, $\varphi : [0,1] \rightarrow \mathbb{R}$ is differentiable by compos^n

By 1st year MVT, $\exists t_0 \in (0,1)$ such that

$$\varphi'(t_0) = \frac{\varphi(1) - \varphi(0)}{1-0} = f(b) - f(a)$$

By the Chain Rule:
$$\begin{aligned}\varphi'(t_0) &= d_{t_0}\varphi(1) = d_{t_0} f \circ \gamma(1) \\ &= d_{\gamma(t_0)} f \circ d_{t_0}\gamma(1) \\ &= d_{\gamma(t_0)} f(b-a) \\ &= \nabla f(\gamma(t_0)) \cdot (b-a)\end{aligned}$$
□

Take $c = \gamma(t_0) \in L_{a,b}$
$= (1-t_0)a + t_0 b$, $t_0 \in (0,1)$

Def: A subset $S \subset \mathbb{R}^m$ is convex if:

$$\forall a, b \in S, \forall t \in [0,1], \ (1-t)a + tb \in S$$

ie: given 2 points in S, the line segment between them is in S

Prop: A convex subset is path-connected

Δ For $a, b \in S$, take $\gamma(t) = (1-t)a + tb$ □

⚠

Theorem: $\mathcal{U} \subset \mathbb{R}^m$ open and convex, $f : \mathcal{U} \longrightarrow \mathbb{R}$ differentiable on $\mathcal{U}$

If there exists $M > 0$ s.t. $\forall x \in \mathcal{U}$, $\|\nabla f(x)\| \leq M$

then $\forall a,b \in \mathcal{U}$, $|f(b) - f(a)| \leq M \cdot \|b-a\|$

Δ Let $a,b \in \mathcal{U}$, since $\mathcal{U}$ is convex $L_{a,b} = \{(1-t)\,a + tb : t \in [0,1]\} \subset \mathcal{U}$

so by the MVT, $\exists c \in L_{a,b}$ s.t.

$$f(b) - f(a) = (b-a) \cdot \nabla f(c)$$

$$\Rightarrow |f(b) - f(a)| = |(b-a) \cdot \nabla f(c)|$$

$$\leq \|b-a\| \cdot \|\nabla f(c)\| \text{ by Cauchy-Schwarz}$$

$$\leq M \cdot \|b-a\| \text{ by assumption}$$

□

⚠

Theorem: $\mathcal{U} \subset \mathbb{R}^m$ open and convex, $f : \mathcal{U} \longrightarrow \mathbb{R}$ differentiable on $\mathcal{U}$

If $\forall x \in \mathcal{U}$, $\nabla f(x) = \vec{0}$ then f is constant on $\mathcal{U}$

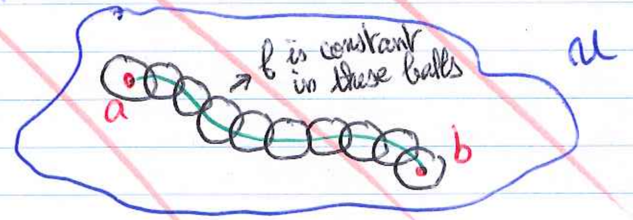
Δ Let $a,b \in \mathcal{U}$, then
$|f(a) - f(b)| \leq 0 \cdot \|b-a\| = 0$, ie: $\forall a,b \in \mathcal{U}$, $f(a) = f(b)$ □

⚠

Theorem: $\mathcal{U} \subset \mathbb{R}^m$ open and path-connected, $f : \mathcal{U} \to \mathbb{R}$ differentiable on $\mathcal{U}$

If $\forall x \in \mathcal{U}$, $\nabla f(x) = \vec{0}$ then f is constant on $\mathcal{U}$

Δ Idea:



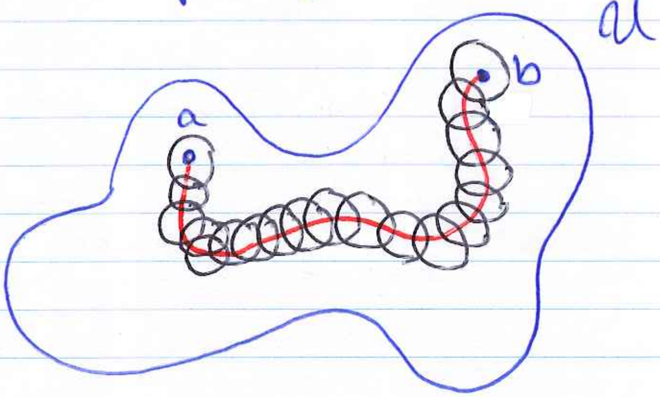
~~take $a,b \in \mathcal{U}$~~

~~take γ a C^∞ path from a to b in $\mathcal{U}$~~

⚠ we don't know if γ is differentiable

$\Rightarrow$ don't try to apply the chain rule □

Δ Idea of proof



take $a, b \in \mathcal{U}$

By path-connectedness,

$\exists \gamma : [0,1] \longrightarrow \mathbb{R}^m$ s.t.

$$\begin{cases} \gamma \text{ is continuous} \\ \gamma(0) = a \\ \gamma(1) = b \\ \forall t \in [0,1], \gamma(t) \in \mathcal{U} \end{cases}$$

⚠ We don't know if γ is differentiable, so we can't mimick the proof of the MVT and compute $(f \circ \gamma)'(t)$

FACT: $\gamma([0,1])$ is compact as the continuous image of a compact set

Hence we may cover $\gamma([0,1])$ by finitely many open balls included in $\mathcal{U}$ which overlap as in the above drawing

Each of the balls are convex, hence f is ~~contin~~ constant on the balls and hence along γ.

Therefore $f(a) = f(b)$ □

Higher order partial derivatives

Def: $\mathcal{U} \subset \mathbb{R}^n$ open, $f : \mathcal{U} \longrightarrow \mathbb{R}$, $a \in \mathcal{U}$

We set $\dfrac{\partial^2 f}{\partial x_j \partial x_i}(a) := \dfrac{\partial}{\partial x_j}\left(\dfrac{\partial f}{\partial x_i}\right)(a)$

whenever it makes sense.

"second-order partial derivative"

Comment: "whenever it makes sense" means that $\dfrac{\partial f}{\partial x_i}$ exists in a small ball around a and admits a directional derivative at a along e_j.

Comment: we first differentiate with respect to x_i and then with respect to x_j (we read from right to left)

More generally, we set $\dfrac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \cdots \partial x_{i_1}}(a) := \dfrac{\partial}{\partial x_{i_k}}\left(\dfrac{\partial}{\partial x_{i_{k-1}}}\left(\cdots\left(\dfrac{\partial f}{\partial x_{i_1}}\right)\right)\right)(a)$

whenever it makes sense.

"partial derivative of order k"

Other notation: $\partial x_{i_k} \partial x_{i_{k-1}} \cdots \partial_{x_{i_1}} f$

Comment: Again we read from right to left: we first differentiate w.r.t x_{i_1} then x_{i_2}, ..., then x_{i_k}

Def: f is of class C^k if all its partial derivatives up to order k exist and are continuous ← don't forget the continuity | $C^0 \equiv$ continuous ⚠

$C^1 \equiv$ "continuously differentiable"

Ex : the order matters :

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

$$\frac{\partial f}{\partial x}(x,y) = \begin{cases} y\dfrac{x^2-y^2}{x^2+y^2} + \dfrac{4x^2y^3}{(x^2+y^2)^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

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

Similarly :

$$\frac{\partial f}{\partial y}(x,y) = \begin{cases} x\dfrac{x^2-y^2}{x^2+y^2} - \dfrac{4x^3y^2}{(x^2+y^2)^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

Then

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{t \to 0} \frac{\frac{\partial f}{\partial y}(t,0) - \frac{\partial f}{\partial y}(0,0)}{t} = \lim_{t \to 0} \frac{t}{t} = 1$$

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = \lim_{t \to 0} \frac{\frac{\partial f}{\partial x}(t,0) - \frac{\partial f}{\partial x}(0,0)}{t} = \lim_{t \to 0} \frac{-t}{t} = -1$$

Hence $\frac{\partial^2 f}{\partial x \partial y}(0,0) \neq \frac{\partial^2 f}{\partial y \partial x}(0,0)$

Theorem : C^k functions are closed by the elementary operations

↳ Before the example

Nevertheless, we have the following result

1740 ↓ ← First correct proof 1873

Theorem: (Clairaut, Schwarz) In MAT237, we use "Clairaut's thm"

$U \subset \mathbb{R}^m$, $f: U \longrightarrow \mathbb{R}$ of class C^2 on U (Δ), $a \in U$

Then $\forall i,j = 1,..,m$, $\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a)$

"If the second partial derivatives are continuous then the order doesn't matter"

Δ WLOG: we assume $U \subset \mathbb{R}^2$, $a = (x_0, y_0) \in U$

WTS: $\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) = \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$

Let $h > 0$, $k > 0$ s.t. $[x_0, x_0+h] \times [y_0, y_0+k] \subset U$

Set $\delta_{h,k} = f(x_0+h, y_0+k) - f(x_0+h, y_0) - f(x_0, y_0+k) + f(x_0, y_0)$

Define $\varphi : [x_0, x_0+h] \to \mathbb{R}$ by $\varphi(x) = f(x, y_0+k) - f(x, y_0)$

then $\delta_{h,k} = \varphi(x_0+h) - \varphi(x_0)$

MVT to φ: $\exists \theta_1 \in (0,1)$ s.t. $\delta_{h,k} = \varphi(x_0+h) - \varphi(x_0) = h\varphi'(x_0+\theta_1 h)$

ie: $\delta_{h,k} = h\left(\frac{\partial f}{\partial x}(x_0+\theta_1 h, y_0+k) - \frac{\partial f}{\partial x}(x_0+\theta_1 h, y_0)\right)$

$\psi : [y_0, y_0+k] \to \mathbb{R}$ defined by $\psi(y) = \frac{\partial f}{\partial x}(x_0+\theta_1 h, y)$

By the MVT to ψ, $\exists \theta_2 \in (0,1)$ s.t.

$$\delta_{h,k} = hk \frac{\partial^2 f}{\partial y \partial x}(x_0+\theta_1 h, y_0+\theta_2 k)$$

Similarly, by repeating the above with y then x, $\exists \theta_3, \theta_4 \in (0,1)$ s.t. $\delta_{h,k} = hk \frac{\partial^2 f}{\partial x \partial y}(x_0+\theta_3 h, y_0+\theta_4 k)$

Therefore :

$$\begin{tikzcd}
\frac{\partial^2 f}{\partial x \partial y}(x_0+\theta_3 h, y_0+\theta_4 k) \arrow[d, "(h,k)\to(0,0)"] & = & \frac{\partial^2 f}{\partial y \partial x}(x_0+\theta_1 h, y_0+\theta_2 k) \arrow[d, "(h,k)\to(0,0)"] \\
\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) & = & \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)
\end{tikzcd}$$

since $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ are continuous

$\square$

By an induction, we get that ⚠

<u>Corollary</u>: $\mathcal{U} \subset \mathbb{R}^m$ open, $f : \mathcal{U} \longrightarrow \mathbb{R}$ of class C^k, $a \in \mathcal{U}$

then $\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \cdots \partial x_{i_1}}(a)$ doesn't depend on the order of the $i_1, \ldots, i_k$

<u>Notation</u>: if f is of class C^k, since the order doesn't matter, the following notation is quite useful:

$$\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{N}^m$$

$$\partial^\alpha f(a) = \frac{\partial^{\alpha_1 + \cdots + \alpha_m} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_m^{\alpha_m}}(a)$$

<u>Homework</u> do the examples and questions of §2.5 of the lecture notes.

Taylor's theorem

The one-variable case (From MAT137)

Def. $I \subset \mathbb{R}$ open interval, $f : I \longrightarrow \mathbb{R}$, $a \in I$.
assume that f is k-th time differentiable at a, then the k-th order Taylor polynomial of f at a is

$$P_{a,k}(x) = f(a) + f'(a)\,x + \frac{f''(a)}{2}x^2 + \dots + \frac{f^{(k)}(a)}{k!}x^k$$
$$= \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!}x^j$$

Prop. $P_{a,k}$ is the unique polynomial of degree at most k s.t.

$$P_{a,k}(0) = f(a),\ P'_{a,k}(0) = f'(a),\ P''_{a,k}(0) = f''(a), \dots P^{(k)}_{a,k}(0) = f^{(k)}(a)$$

Theorem. (Taylor or Taylor-Young)

$I \subset \mathbb{R}$ ~~open~~ interval, $f : I \to \mathbb{R}$ of class C^{k-1} on I, $a \in I$

If $f^{(k)}(a)$ exists then ← I don't assume that $f^{(k)}$ is C^0 at a

Then $f(a+h) = P_{a,k}(h) + E(h)$ where $\dfrac{E(h)}{h^k} \xrightarrow[h \to 0]{} 0$

Δ We set $E(h) = f(a+h) - P_{a,k}(h)$ and $G(h) = h^k$

By L'Hopital's rule applied $(k-1)$ times (check the assumptions for each)

$$\lim_{h\to 0} \frac{E(h)}{G(h)} \overset{L'H}{=} \lim_{h\to 0} \frac{E'(h)}{G'(h)} \overset{L'H}{=} \dots \overset{L'H}{=} \lim_{h\to 0} \frac{E^{(k-1)}(h)}{G^{(k-1)}(h)}$$

(we can go up to there since f is C^{k-1})

by definition of $P_{a,k}$ ← $$= \lim_{h\to 0} \frac{f^{(k-1)}(a+h) - f^{(k-1)}(a) - h f^{(k)}(a)}{k!\,h}$$

since $f^{(k)}(a)$ exists, it is the first term of the lim ← $$= \lim_{h\to 0} \frac{f^{(k-1)}(a+h) - f^{(k-1)}(a)}{h} \cdot \frac{1}{k!} - \frac{f^{(k)}(a)}{k!} = 0 \quad \square$$

particularly f is C^k on I

Theorem (Taylor-Lagrange)

$I \subset \mathbb{R}$ interval, $f: I \to \mathbb{R}$ $(k+1)$-times differentiable on I, $a \in I$

Let $h \in \mathbb{R} \setminus \{0\}$ s.t. $\begin{cases} [a, a+h] \subset I \text{ if } h > 0 \\ \text{or} \\ [a+h, a] \subset I \text{ if } h < 0 \end{cases}$

Then $\begin{cases} \exists \xi \in (a, a+h) \text{ if } h > 0 \\ \text{or} \\ \exists \xi \in (a+h, a) \text{ if } h < 0 \end{cases}$ s.t.

$$f(a+h) = P_{a,k}(h) + \frac{f^{(k+1)}(\xi)}{(k+1)!} h^{k+1} = \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!} h^j + \frac{f^{(k+1)}(\xi)}{(k+1)!} h^{k+1}$$

Δ WLOG, we may assume that $h > 0$

Define $\varphi : [a, a+h] \to \mathbb{R}$ by

$$\varphi(t) = f(a+h) - f(t) - f'(t)(a+h-t) - \dots - \frac{f^{(k)}(t)}{k!}(a+h-t)^k - \frac{A}{(k+1)!}(a+h-t)^{k+1}$$

where we pick $A \in \mathbb{R}$ s.t. $\varphi(a) = 0$, notice that $\varphi(a+h) = 0$

Since φ is C^0 on $[a, a+h]$ and differentiable on $(a, a+h)$,

by Rolle's theorem, $\exists \xi \in (a, a+h)$ s.t. $\varphi'(\xi) = 0$

But, $\forall t \in (a, a+h)$, $\varphi'(t) = -\frac{f^{(k+1)}(t)}{k!}(a+h-t)^k + \frac{A}{k!}(a+h-t)^k$

(when we compute the derivative, the other terms cancel)

Hence $\quad 0 = \varphi'(\xi) = -\frac{f^{(k+1)}(\xi)}{k!}\underset{\neq 0}{\underline{(a+h-\xi)^k}} + \frac{A}{k!}\underset{\neq 0}{\underline{(a+h-\xi)^k}}$

$\Rightarrow A = f^{(k+1)}(\xi)$

Then $0 = \varphi(a) = f(a+h) - P_{a,k}(h) - \frac{f^{(k+1)}(\xi)}{(k+1)!} h^{k+1}$ $\square$

Taylor's theorem in several variables

At order 1

Prop: $\mathcal{U} \underset{\text{open}}{\subset} \mathbb{R}^m$, $f : \mathcal{U} \longrightarrow \mathbb{R}$, $a \in \mathcal{U}$

If f is differentiable at a then

$$f(a+h) = f(a) + \sum_{j=1}^{m} \frac{\partial f}{\partial x_j}(a) h_j + E(h)$$

where $\frac{E(h)}{\|h\|} \xrightarrow[h \to 0]{} 0$

Δ It is just the definition noticing that $d_a f(h) = \sum_{j=1}^{m} \frac{\partial f}{\partial x_j}(a) h_j$ □

At order 2

⚠

Theorem: $\mathcal{U} \subset \mathbb{R}^m$ open, $f : \mathcal{U} \longrightarrow \mathbb{R}$ of class C^2, $a \in \mathcal{U}$, $h \in \mathbb{R}^m$

Assume that $\forall t \in [0,1]$, $a+th \in \mathcal{U}$

Then $\exists \theta \in (0,1)$ s.t. $f(a+h) = f(a) + \sum_{j=1}^{m} \frac{\partial f}{\partial x_j}(a) h_j + \frac{1}{2} \sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j}(a+\theta h) h_i h_j$

Δ Define $\varphi : [0,1] \to \mathbb{R}$ by $\varphi(t) = f(a+th)$

By the chain-rule: $\forall t \in (0,1)$, $\varphi'(t) = \sum_{j=1}^{m} \frac{\partial f}{\partial x_j}(a+th) h_j$

$$\varphi''(t) = \sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j}(a+th) h_i h_j$$

Then by the one variable Taylor-Lagrange $\exists \theta \in (0,1)$

s.t. $\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2}\varphi''(\theta)$

□

Theorem. *Let* $U \subset \mathbb{R}^n$ *open,* $f : U \to \mathbb{R}$ *of class* C^2, $\mathbf{a} \in U$.
Then

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})h_i + \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})h_i h_j + E(\mathbf{h})$$

where $\lim\limits_{\mathbf{h}\to 0} \dfrac{E(\mathbf{h})}{\|\mathbf{h}\|^2} = 0$.

Proof. Let $\mathbf{h} \in \mathbb{R}^n$ be of norm small enough to ensure that $\forall t \in [0,1]$, $\mathbf{a} + t\mathbf{h} \in U$.
By the previous theorem, there exists $\theta \in (0,1)$ such that

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})h_i + \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}+\theta\mathbf{h})h_i h_j$$

Hence

$$\begin{aligned} E(\mathbf{h}) &= f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})h_i - \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})h_i h_j \\ &= \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} \left(\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}+\theta\mathbf{h}) - \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) \right) h_i h_j \end{aligned}$$

So that

$$\frac{E(\mathbf{h})}{\|\mathbf{h}\|^2} = \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n} \left(\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}+\theta\mathbf{h}) - \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) \right) \frac{h_i h_j}{\|\mathbf{h}\|^2}$$

Notice that by continuity of the second order partial derivatives

$$\lim_{\mathbf{h}\to 0} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}+\theta\mathbf{h}) - \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) - \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) = 0$$

and that $\frac{|h_i h_j|}{\|\mathbf{h}\|^2} = \frac{|h_i|}{\|\mathbf{h}\|}\frac{|h_j|}{\|\mathbf{h}\|} \leq 1$.
Hence $\lim\limits_{\mathbf{h}\to 0} \dfrac{E(\mathbf{h})}{\|\mathbf{h}\|^2} = 0$. ■

Definition: $\mathcal{U} \subset \mathbb{R}^m$ open, $f : \mathcal{U} \longrightarrow \mathbb{R}$, $a \in \mathcal{U}$

We define the Hessian matrix of f at a by

$$\mathcal{H}_f(a) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(a) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_m}(a) \\ \vdots & & \\ \frac{\partial^2 f}{\partial x_m \partial x_1}(a) & \cdots & \frac{\partial^2 f}{\partial x_m \partial x_m}(a) \end{pmatrix} \in M_{m,m}(\mathbb{R})$$

whenever it makes sense.

Remark: $$\sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j}(a)\, h_i h_j = h^t \cdot \mathcal{H}_f(a) \cdot h$$

At higher-order

Theorem: $\mathcal{U} \subset \mathbb{R}^m$ open, $f : \mathcal{U} \longrightarrow \mathbb{R}$ of class C^k, $a \in \mathcal{U}$

$$f(a+h) = \sum_{|\alpha| \leq k} \frac{\partial^\alpha f(a)}{\alpha!} h^\alpha + E(h), \quad \frac{E(h)}{\|h\|^k} \xrightarrow[h \to 0]{} 0$$

where $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{N}^m_{\geq 0}$

$|\alpha| = \alpha_1 + \dots + \alpha_m$

$\alpha ! = \alpha_1 ! \alpha_2 ! \cdots \alpha_m !$

$h^\alpha = h_1^{\alpha_1} h_2^{\alpha_2} \cdots h_m^{\alpha_m}$

$\partial^\alpha f(a) = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_m^{\alpha_m}}(a)$

⚠ We admit this one □

How to compute some multivariable Taylor polynomials.

Ex: $\dfrac{e^{x-2y}}{1+x^2-y} = \dfrac{e^{x-2y}}{1-(y-x^2)}$

$$= \left(1+(x-2y)+\frac{(x-2y)^2}{2}+\dots\right)\left(1+(y-x^2)+(y-x^2)^2+\dots\right)$$

$$= 1+y-x^2+(y-x^2)^2+(x-2y)+(x-2y)(y-x^2)+\frac{(x-2y)^2}{2}+\dots$$

$$= \underbrace{1}_{\text{order 0}} + \underbrace{x-y}_{\text{order 1}} \underbrace{-\frac{x^2}{2}-xy+y^2}_{\text{order 2}} + E(x,y)$$

where $\dfrac{E(x,y)}{\|(x,y)\|^2} \xrightarrow[(x,y)\to(0,0)]{} 0$

Hence $P_{(0,0),2}(x,y) = 1+x-y-\dfrac{x^2}{2}-xy+y^2$

Homework: Questions from §2.6 of the online lecture notes
"Basic Skill: 1→4"

Do not attempt the "advanced": I gave alternative proofs in class

Higher order partial derivatives : polar coordinates

Let $f: \mathbb{R}^2 \longrightarrow \mathbb{R}$ of class C^2

Define $S = \{(r,\theta) \in \mathbb{R}^2, \ r > 0\}$ and $\varphi : S \to \mathbb{R}$ by

$$\varphi(r,\theta) = f(r\cos\theta, r\sin\theta)$$

From Oct 22 :

$$\partial_r \varphi = \cos\theta \, \partial_x f + \sin\theta \, \partial_y f$$

$$\partial_\theta \varphi = -r\sin\theta \, \partial_x f + r\cos\theta \, \partial_y f$$

$$\left.\begin{array}{l} \textbf{Comment :} \text{ by } \partial_r\varphi, \text{ I mean } \dfrac{\partial \varphi}{\partial r}(r,\theta) \\ \text{and by } \partial_x f, \text{ I mean } \dfrac{\partial f}{\partial x}(r\cos\theta, r\sin\theta) \end{array}\right\} \begin{array}{l} \text{to lighten} \\ \text{the notation} \end{array}$$

Hence :

$$\partial_r^2 \varphi = \cos^2\theta \, \partial_x^2 f + \cos\theta\sin\theta \, \partial_y\partial_x f + \cos\theta\sin\theta \, \partial_x\partial_y f + \sin^2\theta \, \partial_y^2 f$$

Clairaut's thm $\leftarrow$ $= \cos^2\theta \, \partial_x^2 f + 2\cos\theta\sin\theta \, \partial_x\partial_y f + \sin^2\theta \, \partial_y^2 f$

$$\partial_\theta^2 \varphi = -r \overset{\partial_r\varphi}{\boxed{\cos\theta \, \partial_x f}} + r^2\sin^2\theta \, \partial_x^2 f - r^2 \sin\theta\cos\theta \, \partial_y\partial_x f$$

$$- r\boxed{\sin\theta \, \partial_y f} + r^2\cos^2\theta \, \partial_y^2 f - r^2\sin\theta\cos\theta \, \partial_x\partial_y f$$

Clairaut's thm $= -r\partial_r\varphi + r^2\sin^2\theta \, \partial_x^2 f + r^2\cos^2\theta \, \partial_y^2 f - 2r^2\sin\theta\cos\theta \, \partial_x\partial_y f$

$$\underbrace{\Delta f := \partial_x^2 f + \partial_y^2 f}_{} = \partial_r^2\varphi + \frac{1}{r}\partial_r\varphi + \frac{1}{r^2}\partial_\theta^2\varphi$$

Laplacian operator : heat eqn, wave eqn, ...

$$\partial_\theta \partial_r \varphi = -\sin\theta\, \partial_x f - r\cos\theta\sin\theta\, \partial_x^2 f + r\cos^2\theta\, \partial_y \partial_x f$$

$$+ \cos\theta\, \partial_y f - r\sin^2\theta\, \partial_x \partial_y f + r\cos\theta\sin\theta\, \partial_y^2 f$$

$$= \frac{1}{r}\partial_\theta \varphi - \frac{r}{2}\sin(2\theta)\, \partial_x^2 f + \frac{r}{2}\sin(2\theta)\, \partial_y^2 f + r\cos(2\theta)\, \partial_x \partial_y f$$

Clairaut's thm & $\sin(2\theta) = 2\sin\theta\cos\theta$ & $\cos^2\theta - \sin^2\theta = \cos 2\theta$

Solving the one-dimensional wave equation (Extracurricular)

$$f : \begin{array}{l} \mathbb{R}^2 \longrightarrow \mathbb{R} \\ (x,t) \mapsto f(x,t) \end{array} \quad , \quad c > 0 \quad , \quad f \text{ of class } C^2$$

the eqn : $\dfrac{\partial^2 f}{\partial x^2} - \dfrac{1}{c^2}\dfrac{\partial^2 f}{\partial t^2} = 0$

Define $\varphi : \mathbb{R}^2 \to \mathbb{R}$ by $\varphi(u,v) = f\left(\frac{u+v}{2}, \frac{u-v}{2c}\right)$

$$\frac{\partial \varphi}{\partial u} = \frac{1}{2}\frac{\partial f}{\partial x} + \frac{1}{2c}\frac{\partial f}{\partial t}$$

$$\frac{\partial^2 \varphi}{\partial v \partial u} = \frac{1}{4}\frac{\partial^2 f}{\partial x^2} - \frac{1}{4c}\frac{\partial^2 f}{\partial t \partial x} + \frac{1}{4c}\frac{\partial^2 f}{\partial x \partial t} - \frac{1}{4c^2}\frac{\partial^2 f}{\partial t^2}$$

$$= \frac{1}{4}\left(\frac{\partial^2 f}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2}\right)$$

CCL : $\dfrac{\partial^2 f}{\partial x^2} - \dfrac{1}{c^2}\dfrac{\partial^2 f}{\partial t^2} = 0 \iff \dfrac{\partial^2 \varphi}{\partial u \partial v} = 0$

$\iff \varphi(u,v) = A(u) + B(v)$

$A, B : \mathbb{R} \to \mathbb{R}$ of class C^2

$$\begin{cases} x = \dfrac{u+v}{2} \\ t = \dfrac{u-v}{2c} \end{cases} \iff \begin{cases} u = x + ct \\ v = x - ct \end{cases}$$

$\underline{f(x,y) = A(x+ct) + B(x-ct) \quad , \quad A, B : \mathbb{R} \to \mathbb{R} \text{ of class } C^2}$

Critical points

Def: $\mathcal{U} \subset \mathbb{R}^n$ open, $f : \mathcal{U} \to \overset{\text{differentiable}}{\mathbb{R}}$, $a \in \mathcal{U}$

We say that a is a critical point of f if $\nabla f(a) = \vec{0}$

Def: $\mathcal{U} \subset \mathbb{R}^n$ open, $f : \mathcal{U} \to \mathbb{R}$, $a \in \mathcal{U}$

We say that a is a local min of f if

$$\exists r > 0, \forall x \in \mathcal{U}, \|x - a\| < r \Rightarrow f(a) \leq f(x)$$

ie: $\exists r > 0, \forall x \in B(a,r) \cap \mathcal{U}, \quad f(a) \leq f(x)$

Comment: since $B(a,r) \cap \mathcal{U}$ is open as the intersection of two open sets we may assume that $B(a,r) \subset \mathcal{U}$ up to shrinking r

Def: We say that a is a local max of f if

$$\exists r > 0, \forall x \in \mathcal{U}, \|x - a\| < r \Rightarrow f(a) \geq f(x)$$

ie $\exists r > 0, \forall x \in B(a,r) \cap \mathcal{U}, \quad f(a) \geq f(x)$

Def: local extremum := local min or local max

Theorem (First derivative test)

Let $\mathcal{U} \subset \mathbb{R}^n$ open, $f : \mathcal{U} \to \mathbb{R}$ differentiable, $a \in \mathcal{U}$.
If a is a local extremum then it is a critical point.

ie: a local extremum $\Rightarrow$ a critical point : $\left\{ \begin{array}{l} \text{the local extrema are} \\ \text{among the critical points} \end{array} \right.$

Δ Let $a = (a_1, \dots, a_n) \in \mathcal{U}$ be a local extremum of f
Then a_j is a local extremum of $g(t) = f(a_1, \dots, a_{j-1}, t, a_{j+1}, \dots, a_n)$

hence $g'(a_j) = 0$ by MAT137; but $g'(a_j) = \frac{\partial f}{\partial x_j}(a)$

Therefore $\nabla f(a) = \vec{0}$ $\square$

Study up to order 2

$U \subset \mathbb{R}^m$ open, $f : U \to \mathbb{R}$ of class C^2, $a \in U$ critical point

then $f(a+h) = f(a) + \nabla f(a) \cdot h + \frac{1}{2} h^t Hf(a)\, h + E(h)$ where $\frac{E(h)}{\|h\|^2} \xrightarrow[h \to 0]{} 0$

becomes $f(a+h) - f(a) = \frac{1}{2} h^t Hf(a)\, h + E(h)$ ⊛

and by Clairaut's theorem $Hf(a)$ is a symmetric matrix

Definitions: $A \in M_{m,m}(\mathbb{R})$ symmetric (ie $A^t = A$) is said to be

- positive definite if $\forall h \in M_{m,1}(\mathbb{R})$, $h \neq 0 \Rightarrow h^t A h > 0$
- nonnegative definite if $\forall h \in M_{m,1}(\mathbb{R})$, $h^t A h \geqslant 0$
- negative definite if $h \neq 0 \Rightarrow h^t A h < 0$
- non positive definite if $\forall h$, $h^t A h \leqslant 0$
- non-definite if it is not nonnegative definite neither nonpositive definite
  ie $\exists h, k$ s.t. $h^t A h < 0 < k^t A k$
- degenerate if $\det A = 0$
- non-degenerate if $\det A \neq 0$

Theorem: ① positive definite $\Leftrightarrow$ eigenvalues are > 0
② nonnegative definite $\Leftrightarrow$ eigenvalues are $\geqslant 0$
③ negative definite $\Leftrightarrow$ eigenvalues are < 0
④ nonpositive definite $\Leftrightarrow$ eigenvalues are $\leqslant 0$
⑤ indefinite $\Leftrightarrow$ some eigenvalues are > 0 and some are < 0

Δ ① $\Rightarrow$ let λ be an eigenvalue with eigenvector $h \neq 0$ then

$$0 < h^t A h = h^t \lambda h = \lambda h^t h = \lambda \|h\|^2 \Rightarrow \lambda > 0$$

$\Leftarrow$ recall that, since A is symmetric, we may find an orthogonal basis (∗∗) of $\mathbb{R}^m$ made of eigenvectors of A □

Corollary: ① positive definite $\Leftrightarrow$ non-degenerate + nonnegative definite
② negative definite $\Leftrightarrow$ non-degenerate + non positive definite

Δ The determinant of a symmetric matrix is the product of its eigenvalues (with mult)

indeed by (∗∗) $A = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}$ in some basis and then $\det A = \lambda_1 \cdots \lambda_m$ □

Lemma: Let $A \in M_{m,m}(\mathbb{R})$ be a symmetric matrix then

① A positive definite $\Leftrightarrow \exists \lambda > 0, \forall h \in \mathbb{R}^m, h^t A h \geqslant \lambda \|h\|^2$

② A negative definite $\Leftrightarrow \exists \lambda < 0, \forall h \in \mathbb{R}^m, h^t A h \leqslant \lambda \|h\|^2$

Δ ① $\Leftarrow$ let $h \in \mathbb{R}^m \setminus \{\vec{0}\}$ then $h^t A h \geqslant \lambda \|h\|^2 > 0$

$\Rightarrow$ let $(u_1, \dots, u_m)$ be an orthogonal basis of eigenvectors (exists since A is symmetric)
let $h \in \mathbb{R}^m$ then $h = \sum_{i=1}^{m} \alpha_i u_i$, $\alpha_i \in \mathbb{R}$

$$h^t A h = \sum_{i,j} \alpha_i \alpha_j u_i^t A u_j$$

$$= \sum_{i,j} \alpha_i \alpha_j \lambda_j u_i^t u_j \quad \text{where } A u_j = \lambda_j u_j, \ \lambda_j > 0$$

$$= \sum_i \alpha_i^2 \lambda_i u_i^t u_i \quad \text{since } i \neq j \Rightarrow u_i^t u_j = u_i \cdot u_j = 0$$

$$\geq \min(\lambda_i) \sum_i \alpha_i^2 u_i^t u_i$$

$$= \min(\lambda_i) \sum_{i,j} \alpha_i \alpha_j u_i^t u_j \quad \text{since } u_i^t u_j = u_i \cdot u_j = 0$$

$$= \min(\lambda_i) \left(\sum \alpha_i u_i\right) \cdot \left(\sum \alpha_j u_j\right)$$

$$= \min(\lambda_i) \|h\|^2$$

② Apply ① to $-A$

□

ie $\nabla f(a) = \vec{0}$

Theorem: (Second derivative test)

$U \subset \mathbb{R}^n$ open, $f: U \to \mathbb{R}$ of class C^2, $a \in U$ critical point

↳ don't forget this assumption

① If $\mathcal{H}_f(a)$ is positive definite then a is a local min

② If $\mathcal{H}_f(a)$ is negative definite then a is a local max

③ If $\mathcal{H}_f(a)$ is indefinite then a is neither a local max nor a local min

Remark: in all the other cases we can't conclude about the nature of "a" simply from $\mathcal{H}_f(a)$

Ex: $f: \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^2$ has a local min at $(0,0)$

$g: \mathbb{R}^2 \to \mathbb{R}$, $g(x,y) = x^3$ has no local extremum at $(0,0)$

$\mathcal{H}_f(\vec{0})$ and $\mathcal{H}_g(\vec{0})$ are both non-negative definite

Proof: ① $f(a+h) - f(a) = \frac{1}{2} h^t \mathcal{H}_f(a) h + E(h)$ by ⊛

$$\geqslant \frac{1}{2}\lambda \|h\|^2 + E(h) \text{ for some } \lambda > 0 \text{ by the lemma}$$

$$= \|h\|^2 \underbrace{\left(\underset{>0}{\frac{\lambda}{2}} + \frac{E(h)}{\|h\|^2} \xrightarrow[h\to 0]{} 0 \right)}_{\text{hence } >0 \text{ for } \|h\| \text{ small enough}}$$

② apply ① to "$-f$"

③ $\exists h,k$ s.t. $h^t \mathcal{H}_f(a) h < 0 < k^t \mathcal{H}_f(a) k$

$f(a+th) - f(a) = t^2 \left(h^t \mathcal{H}_f(a) h + \|h\|^2 \frac{E(th)}{\|th\|^2} \right) < 0$ for t small enough

$f(a+tk) - f(a) > 0$ for t small enough

So f takes some values $> f(a)$ and $< f(a)$ in any ball centered at a

□

Comment: in the case ③ $g(t) = f(a+th)$ has a local max at 0

and $h(t) = f(a+tk)$ has a local min at 0

$\rightsquigarrow$ looks like a saddle.

⚠ don't forget this assumption

Def: $U \subset \mathbb{R}^n$ open, $f : U \longrightarrow \mathbb{R}$ of class C^2, $a \in U$ $\boxed{\text{critical point}}$

We say that a is a saddle point if $Hf(a)$ is indefinite

The two-variable case

don't forget this assumption

$\mathcal{U} \subset \mathbb{R}^2$ open, $f : \mathcal{U} \to \mathbb{R}$ of class C^2, $a \in \mathcal{U}$ $\underbrace{\text{critical}}$ point of f.

Set $\alpha = \frac{\partial^2 f}{\partial x^2}(a)$, $\beta = \frac{\partial^2 f}{\partial x \partial y}(a) = \frac{\partial^2 f}{\partial y \partial x}(a)$, $\gamma = \frac{\partial^2 f}{\partial y^2}(a)$

by Clairaut's theorem

then $\mathcal{H}f(a) = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$

Since $\mathcal{H}f(a)$ is a symmetric 2×2-matrix, it has two eigenvalues λ_1, λ_2 (that may be equal) and then its determinant is the product of its eigenvalues, ie: $\lambda_1 \lambda_2 = \det(\mathcal{H}f(a)) = \alpha\gamma - \beta^2$

- $\alpha\gamma - \beta^2 > 0$: then either λ_1, λ_2 are both positive and $\mathcal{H}f(a)$ is positive definite or they are both negative and $\mathcal{H}f(a)$ is negative definite.

  ie either $h \neq 0 \Rightarrow h^t \mathcal{H}f(a) h > 0$ (positive definite)
  or $h \neq 0 \Rightarrow h^t \mathcal{H}f(a) h < 0$ (negative definite)

  But $e_1^t \mathcal{H}f(a) e_1 = (1\ 0)\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \alpha$

  So if $\alpha > 0$ then $\mathcal{H}f(a)$ is positive definite and a is a local min of f
  and if $\alpha < 0$ then $\mathcal{H}f(a)$ is negative definite and a is a local max of f

  Comment: $\alpha = 0$ is impossible since $\mathcal{H}f(a)$ is positive or negative definite but you can double check: $\alpha = 0 \Rightarrow \det \mathcal{H}f(a) = -\beta^2 \leq 0$ impossible

- $\alpha\gamma - \beta^2 < 0$: one eigenvalue is > 0 and the other < 0
  then $\mathcal{H}f(a)$ is indefinite and a is a saddle point

- $\alpha\gamma - \beta^2 = 0$: one eigenvalue is 0 so $\mathcal{H}f(a)$ can't be positive/negative definite and since there are only two eigenvalues we can't have a positive and a negative eigenvalue so $\mathcal{H}f(a)$ is not indefinite.

  $\Rightarrow$ we can't conclude

We just proved:

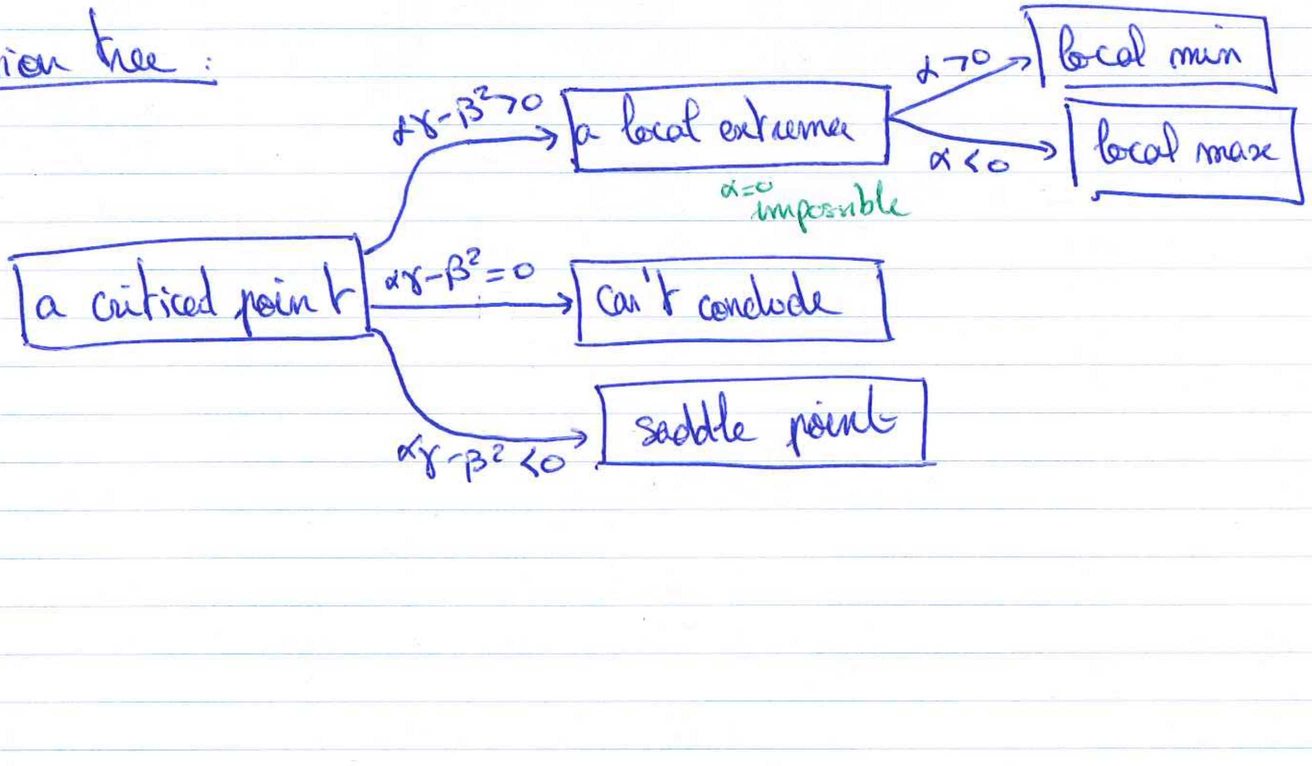
Theorem (Monge)

$U \subset \mathbb{R}^2$ open, $f : U \to \mathbb{R}$ of class C^2, $\overbrace{a \in U \text{ critical point}}^{\text{ie } \frac{\partial f}{\partial x}(a) = \frac{\partial f}{\partial y}(a) = 0}$ of f

We denote $Hf_f(a) = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$, $\alpha = \partial_x^2 f(a)$, $\beta = \partial_{xy}^2 f(a) = \partial_{yx}^2 f(a)$, $\gamma = \partial_y^2 f(a)$

- If $\begin{cases} \alpha\gamma - \beta^2 > 0 \\ \alpha > 0 \end{cases}$ then a is a local min of f
- If $\begin{cases} \alpha\gamma - \beta^2 > 0 \\ \alpha < 0 \end{cases}$ then a is a local max of f
- If $\alpha\gamma - \beta^2 < 0$ then a is a saddle point of f
- If $\alpha\gamma - \beta^2 = 0$ then we can't determine the nature of a from $Hf_f(a)$

Decision tree :



Constrained optimization: Lagrange multipliers.

A linear algebra lemma. (You can safely skip it)

Let $\varphi_1, \dots, \varphi_p, \psi : \mathbb{R}^m \longrightarrow \mathbb{R}$ be linear

Then

$$\bigcap_{i=1}^{p} \ker(\varphi_i) \subset \ker(\psi) \iff \exists a_1, \dots, a_p \in \mathbb{R},\ \psi = \sum_{i=1}^{p} a_i \varphi_i$$

Δ $\underline{\Leftarrow}$: Assume that $\psi = \sum_{i=1}^{p} a_i \varphi_i$ for some $a_i \in \mathbb{R}$

Let $x \in \bigcap_{i=1}^{p} \ker \varphi_i$ then

$$\psi(x) = \sum_{i=1}^{p} a_i \varphi_i(x) = \sum_{i=1}^{p} a_i 0 = 0$$

Hence $x \in \ker \psi$

We proved that $\bigcap_{i=1}^{p} \ker(\varphi_i) \subset \ker(\psi)$

$\underline{\Rightarrow}$: We define $\underline{\Phi} : \mathbb{R}^m \to \mathbb{R}^p$ by $\underline{\Phi}(x) = (\varphi_1(x), \dots, \varphi_p(x))$

Notice that $\underline{\Phi}$ is linear since the φ_i are

Claim 1: $\ker \underline{\Phi} \subset \ker \psi$

Indeed, let $x \in \ker \underline{\Phi}$, then $\vec{0} = \underline{\Phi}(x) = (\varphi_1(x), \dots, \varphi_p(x))$ and $x \in \bigcap_{i=1}^{p} \ker \varphi_i \subset \ker \psi$

Hence $\ker \underline{\Phi} \subset \ker \psi$ as claimed.

Claim 2: $\exists f : \mathbb{R}^p \to \mathbb{R}$ linear such that $\psi = f \circ \underline{\Phi}$

Set $r = \operatorname{rank}(\underline{\Phi})$, then by the rank-nullity theorem, $\dim \ker \underline{\Phi} = m - r$

Hence we may find a basis $(v_1, \dots, v_m)$ of $\mathbb{R}^m$ such that $(v_{r+1}, \dots, v_m)$ is a basis of $\ker \underline{\Phi}$

Then $u_1 = \Phi(v_1), \dots, u_r = \Phi(v_r)$ are linearly dependent,

indeed $\sum_{i=1}^{r} a_i \Phi(v_i) = 0 \Rightarrow \Phi\left(\sum_{i=1}^{r} a_i v_i\right) = 0$

$$\Rightarrow \sum_{i=1}^{r} a_i v_i \in \ker \Phi$$

$$\Rightarrow \forall i,\ a_i = 0 \quad \text{since } \mathbb{R}^m = \langle v_1, .., v_r \rangle \oplus \ker \Phi$$

So we can extend $(u_1, .., u_r)$ in a basis $(u_1, .., u_r, u_{r+1}, .., u_p)$ of $\mathbb{R}^p$.

Now we define $f : \mathbb{R}^p \to \mathbb{R}$ linear by:

$$f(u_1) = \psi(v_1), .., f(u_r) = \psi(v_r),\ f(u_{r+1}) = \dots = f(u_p) = 0$$

Let's check that $\psi = f \circ \Phi$

let $x \in \mathbb{R}^m$, then $x = \sum_{i=1}^{m} x_i v_i$, and

$$\begin{aligned}
f \circ \Phi(x) &= f\left(\sum_{i=1}^{m} x_i \Phi(v_i)\right) \\
&= f\left(\sum_{i=1}^{r} x_i u_i\right) \quad \text{since } \begin{cases} \Phi(v_i) = u_i \text{ for } i = 1, .., r \\ \Phi(v_i) = 0 \text{ for } i = r+1, .., m \end{cases} \\
&= \sum_{i=1}^{r} x_i f(u_i) \\
&= \sum_{i=1}^{r} x_i \psi(v_i) \\
&= \sum_{i=1}^{m} x_i \psi(v_i) \quad \text{since for } i \geq r+1,\ v_i \in \ker \Phi \subset \ker \psi \text{ by claim 1} \\
&= \psi\left(\sum x_i v_i\right) \\
&= \psi(x)
\end{aligned}$$

And the claim is proved

Now, since $f : \mathbb{R}^p \to \mathbb{R}$ is linear, $f(y_1, .., y_p) = \sum_{i=1}^{p} y_i f(e_i)$

and $\psi(x) = f(\Phi(x)) = f(\varphi_1(x), .., \varphi_p(x)) = \sum_{i=1}^{p} f(e_i)\varphi_i(x) = \sum_{i=1}^{p} a_i \varphi_i(x)$

for $a_i = f(e_i)$

□

(Extra curricular) (You can safely skip it)

Comment: If you are familiar with duality then the proof of "$\underline{\Leftarrow}$" is very natural:

Δ $\varphi_1, \dots, \varphi_p$ are vectors of the n-dim space $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$

so we may find a linearly independant subfamily $\varphi_1, \dots, \varphi_q$ in $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$

such that $\mathrm{Vect}(\varphi_1, \dots, \varphi_q) = \mathrm{Vect}(\varphi_1, \dots, \varphi_p)$

Then we extend $(\varphi_1, \dots, \varphi_q)$ in a basis $(\varphi_1, \dots, \varphi_n)$ of $\mathcal{L}(\mathbb{R}^n, \mathbb{R})$

Hence $\psi = \sum_{i=1}^{n} a_i \varphi_i$

Let $(e_1, \dots, e_n)$ the basis of $\mathbb{R}^n$ dual to $(\varphi_1, \dots, \varphi_n)$, ie $\varphi_i(e_j) = \begin{cases} 1 & \text{for } i=j \\ 0 & \text{otherwise} \end{cases}$

For $j \geq q+1$, $e_j \in \ker\left(\bigcap_{i=1}^{q} \varphi_i\right) = \ker\left(\bigcap_{i=1}^{p} \varphi_i\right) \subset \ker \psi$

Hence $\quad 0 = \psi(e_j) = \sum_{i=1}^{n} a_i \varphi_i(e_j) = a_j \Rightarrow \forall j \geq q+1, \; a_j = 0$

and $\psi = \sum_{j=1}^{q} a_j \varphi_j$

$\square$

Theorem : (Lagrange multipliers) ⚠ Main result of this chapter

$\mathcal{U} \subset \mathbb{R}^m$ open, $f, g_1, \dots, g_p : \mathcal{U} \longrightarrow \mathbb{R}$ of class C^1.

Define $X = \{x \in \mathcal{U} : g_1(x) = \dots = g_p(x) = 0\}$

If : $\begin{cases} f_{|X} : X \to \mathbb{R} \text{ has a local extremum at } a \in X \\ \text{and} \\ \nabla g_1(a), \dots, \nabla g_p(a) \text{ are linearly independant} \end{cases}$

then there exist $\lambda_1, \dots, \lambda_p \in \mathbb{R}$ s.t. $\nabla f(a) = \sum_{i=1}^{p} \lambda_i \nabla g_i(a)$

Comment :

⚠ $x \in X$ not $x \in \mathcal{U}$

- $f_{|X}$ has a local min at $a \in X$ means $\exists r > 0, \forall x \in X, \|x-a\| < r \Rightarrow f(a) \leq f(x)$
- $f_{|X}$ has a local max at $a \in X$ means $\exists r > 0, \forall x \in X, \|x-a\| < r \Rightarrow f(a) \geq f(x)$

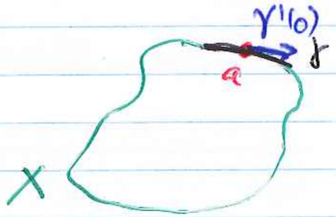
Δ **Sketch of proof : the geometric idea** (You can safely skip it)

Fact : $\bigcap_{i=1}^{p} \ker(d_a g_i) = \left\{ v \in \mathbb{R}^m : v = \gamma'(0) \text{ for a } C^1\ \gamma : (-1,1) \to \mathbb{R}^m \text{ s.t. } \forall t \in (-1,1),\ \gamma(t) \in X \text{ and } \gamma(0) = a \right\}$

We admit this fact, but you can convince yourself that these two sets describe the tangent space of X at a

- v is tangent to $g_i = 0$ at a means $0 = \nabla g_i(a) \cdot v = d_a g_i(v)$, ie $v \in \ker d_a g_i$

  so v is tangent to X at a if v is tangent to all the $g_i = 0$, ie $v \in \bigcap \ker d_a g_i$
- v is tangent to X at a if $v = \gamma'(0)$ for γ has above



Let $\gamma : (-1,1) \to \mathbb{R}^n$ s.t. $\gamma(0) = a$, $\forall t \in (-1,1), \gamma(t) \in X$ and γ C^1

if $f_{|X}$ has an extremum at a, then $f \circ \gamma$ has an extremum at 0

So

$$0 = (f\circ\gamma)'(0) = d_0(f\circ\gamma)(1) = d_{\gamma(0)} f \circ d_0\gamma(1) = d_a f(\gamma'(0))$$

hence $\gamma'(0) \in \ker d_a f$ for any γ as above

By the fact $\bigcap_{i=1}^{p} d_a g_i = \{\gamma'(0) : \gamma \text{ as above}\} \subset \ker d_a f$

Hence by the linear algebra lemma:

$$d_a f = \sum_{i=1}^{p} \lambda_i \, d_a g_i \quad \text{for some } \lambda_i \in \mathbb{R}$$

i.e $$\nabla f(a) = \sum_{i=1}^{p} \lambda_i \nabla g_i(a)$$

$\square$

Remark: the assumption $(\nabla g_1(a), .., \nabla g_p(a))$ linearly independant ensures that $X = g_1^{-1}(0) \cap \cdots \cap g_p^{-1}(0)$ is a "submanifold" at a so that the tangent space of X at a is well defined in the above proof:

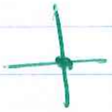
Ex: $X = \{x^2 + y^2 = 1\}$

$a = (0,1)$

Non-ex: $X = \{x^3 - y^2 = 0\}$

$a = (0,0)$

Non-ex: $X = \{xy = 0\}$ $a = (0,0)$



Special case: $p=1$

Theorem:

$\mathcal{U} \subset \mathbb{R}^n$ open, $f, g : \mathcal{U} \to \mathbb{R}$ C^1

Let $X = g^{-1}(0) := \{x \in \mathcal{U},\ g(x) = 0\}$

If $\begin{cases} f_{|X} \text{ has a local extremum at } a \in X \\ \nabla g(a) \neq \vec{0} \end{cases}$

then $\nabla f(a) = \lambda \nabla g(a)$ for some $\lambda \in \mathbb{R}$

If the constraint is given by an inequality:

$\mathcal{U} \subset \mathbb{R}^n$ open, $f, g : \mathcal{U} \to \mathbb{R}$ C^1

$X = g^{-1}((-\infty, 0)) := \{x \in \mathcal{U},\ g(x) \leq 0\}$

We look for local extrema of f on X

Notice that $X = \{x \in \mathcal{U} : g(x) = 0\} \cup \{x \in \mathcal{U} : g(x) < 0\}$

$$= \overset{\shortparallel}{X_1} \cup \overset{\shortparallel}{X_2}.$$

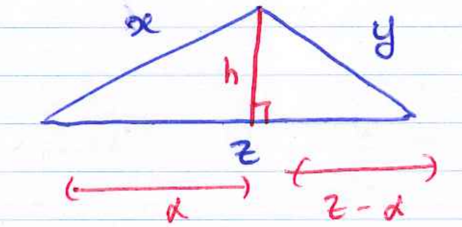
Step 1: Use Lagrange multipliers on X_1

Step 2: Notice that X_2 is open so you can use the results from the previous chapters here.

Homework: examples + questions from 2.8.

A first application

What's the largest area that we can obtain with a triangle of perimeter p?



$$\begin{cases} x^2 = h^2 + \alpha^2 \\ y^2 = h^2 + (z-\alpha)^2 \end{cases} \Rightarrow \quad x^2 - y^2 = \alpha^2 - (z-\alpha)^2 = 2\alpha z - z^2$$

$$\Rightarrow \quad \alpha = \frac{x^2 - y^2 + z^2}{2z}$$

$$h^2 = x^2 - \alpha^2 = x^2 - \frac{(x^2-y^2+z^2)^2}{(2z)^2}$$

$$= \frac{(2xz)^2 - (x^2-y^2+z^2)}{(2z)^2}$$

$$= \frac{(2xz - x^2 + y^2 - z^2)(2xz + x^2 - y^2 + z^2)}{4z^2}$$

$$= \frac{(y^2 - (x-z)^2)((x+z)^2 - y^2)}{4z^2}$$

$$= \frac{(y-x+z)(y+x-z)(x+z-y)(x+z+y)}{4z^2}$$

$$= \frac{P(P-2x)(P-2y)(P-2z)}{4z^2}$$

$$A = \frac{zh}{2} \Rightarrow A^2 = \frac{z^2h^2}{4} = \frac{p(p-2x)(p-2y)(p-2z)}{16}$$

Since $t \mapsto \sqrt{t}$ is increasing on $[0,+\infty)$, it is enough to maximize A^2 with the constraint $x+y+z=p$

(Get rid of the square root when you can...)

$$f(x,y,z) = p(p-2x)(p-2y)(p-2z)$$

$$g(x,y,z) = x+y+z-p$$

So we want to maximize f with the constraint $g=0$ for $x \geq 0, y \geq 0, z \geq 0$.

$S = \{(x,y,z) \in \mathbb{R}^3 : x \geq 0, y \geq 0, z \geq 0, x+y+z=p\}$

is compact hence f has a max on S.

If x, y or $z=0$ then $A=0$, hence we study $f: U \to \mathbb{R}$ on the open set $\{(x,y,z) : x>0, y>0, z>0\}$ with the constraint $g(x,y,z)=0$

By Lagrange multipliers theorem, at a local max $\underset{(x_0,y_0,z_0)}{a}$ on $U \cap g^{-1}(0)$ we have

$\nabla f(a) = \lambda \nabla g(a)$ for some λ, assuming $\nabla g(a) \neq \vec{0}$ which is the case

$$\Leftrightarrow \begin{pmatrix} -2(p-2y_0)(p-2z_0) \\ -2(p-2x_0)(p-2z_0) \\ -2(p-2x_0)(p-2z_0) \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Hence we have to solve

$$\begin{cases} (p-2y_0)(p-2z_0) = \mu \\ (p-2x_0)(p-2z_0) = \mu \\ (p-2x_0)(p-2y_0) = \mu \\ p = x_0+y_0+z_0 \end{cases}$$

here $\mu = -\frac{\lambda}{2}$

$\underline{\text{eg}}$: $(p-2y_0)(p-2z_0) = \mu = (p-2y_0)(p-2x_0)$

$\Rightarrow p - 2z_0 = p - 2x_0$

$\Rightarrow x_0 = z_0$

$$\Rightarrow \begin{cases} x_0 = y_0 = z_0 \\ p = x_0+y_0+z_0 \end{cases}$$

$$\Rightarrow x_0 = y_0 = z_0 = P/3$$

it is the only local extremum in $\{(x,y,z) : x>0, y>0, z>0, x+y+z=p\}$ and $f(P/3,P/3,P/3) > 0$ the min on S is 0 for $x=0$ or $y=0$ or $z=0$

So the only local max is at $(P/3, P/3, P/3)$ and it has to be a global max

$\rightsquigarrow$ We get the max area for an equilateral triangle and

$$A^2 = \frac{p\left(p-\frac{2}{3}p\right)^3}{16} = \frac{p^4}{27\times16}$$

$\underline{\text{ie}}$ $$A = \frac{p^2}{12\sqrt{3}}$$

<u>Homework</u>: questions from section 2.8.

A fancy proof of the AM-GM inequality

$\forall x_1,\dots,x_n \in \mathbb{R}_{\geqslant 0}$, $\sqrt[n]{x_1\cdots x_n} \leqslant \frac{x_1+\cdots+x_n}{n}$

Δ $\Gamma = \{(x_1,\dots,x_n)\in\mathbb{R}^n, x_i \geqslant 0, x_1+\dots+x_n=1\}$ is closed and bounded

Hence it is compact and $f:\Gamma\longrightarrow\mathbb{R}$ defined by

$f(x_1,\dots,x_n)=x_1\cdots x_n$ has a max on Γ since it is C^0 on a compact set.

If one of the $x_i=0$ then $f(x_1,\dots,x_n)=0$ so the max of f on Γ must be in $U\cap X$ where $U=\{(x_1,\dots,x_n)\in\mathbb{R}^n, x_i>0\}$ is open and $X=\{(x_1,\dots,x_n)\in\mathbb{R}^n : g(x_1,\dots,x_n)=0\}$

where $g(x_1,\dots,x_n)=x_1+\dots+x_n-1$

Notice that $\nabla g(x_1,\dots,x_n)=(1,\dots,1)\neq\vec{0}$ hence, by Lagrange's multipliers theorem, if a is a max of f on X then $\exists\lambda\in\mathbb{R}$ s.t.

$\nabla f(a)=\lambda\nabla g(a)$ ie $(f(a)/a_1,\dots,f(a)/a_n)=\lambda(1,\dots,1)$

Hence $\forall i,j$, $a_i=a_j$.

Moreover $g(a)=0$ ie $a_1+\dots+a_n=1 \Rightarrow \forall i, a_i=1/n$.

Hence $f(1/n,\dots,1/n)=\frac{1}{n^n}$ has to be the max of f on X

(it is >0 and the only local extremum here, the min on Γ is 0 when some $x_i=0$)

Now let $x_1,\dots,x_n\in\mathbb{R}_{>0}$ and set $x_i'=\frac{x_i}{\sum_{j=1}^{n} x_j}$ then $\sum_{i=1}^{n} x_i'=1$

(if a $x_i=0$ then the statement is obvious)

so that $\underset{\displaystyle x_1'\cdots x_n' = \frac{x_1\cdots x_n}{(\Sigma x_i)^n}}{\underset{\shortparallel}{f(x_1',\dots,x_n')}} \leqslant \frac{1}{n^n} \Rightarrow x_1\cdots x_n \leqslant \frac{(\Sigma_i x_i)^n}{n^n} \Rightarrow \sqrt[n]{x_1\cdots x_n} \leqslant \frac{\Sigma x_i}{n}$

$\square$

Ex: Let $L = \{(x,y,z) \in \mathbb{R}^3 : x+y+z+\frac{7}{2}=0,\ x-y+2z=0\}$

Find $p \in L$ minimizing the distance to the origin

Δ Notice that for $g_1(x,y,z) = x+y+z+\frac{7}{2}$ and $g_2(x,y,z) = x-y+2z$

$\nabla g_1(x,y,z) = (1,1,1)$ and $\nabla g_2(x,y,z) = (1,-1,2)$ are linearly independent

hence L is a line

Let $f(x,y,z) = x^2+y^2+z^2$

By Lagrange multipliers theorem, if $p = (x_0, y_0, z_0)$ is a min of $f_{|L}$

then $\exists \lambda_1, \lambda_2 \in \mathbb{R}$ s.t. $\nabla f(p) = \lambda_1 \nabla g_1(p) + \lambda_2 \nabla g_2(p)$

$$\Rightarrow \begin{pmatrix} 2x_0 \\ 2y_0 \\ 2z_0 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$$

then $x_0 = \dfrac{\lambda_1+\lambda_2}{2}$, $y_0 = \dfrac{\lambda_1-\lambda_2}{2}$, $z_0 = \dfrac{\lambda_1+2\lambda_2}{2}$

and $\begin{cases} g_1(x_0,y_0,z_0) = 0 \\ g_2(x_0,y_0,z_0) = 0 \end{cases} \Rightarrow \begin{cases} 3\lambda_1 + 2\lambda_2 = -7 \\ \lambda_1 + 3\lambda_2 = 0 \end{cases} \Rightarrow \begin{cases} \lambda_1 = -3 \\ \lambda_2 = +1 \end{cases}$

Therefore $(x_0,y_0,z_0) = (-1,-2,-\frac{1}{2})$ is the only possible $p \in L$ minimizing the distance to the origin

□

Ex : Find the min and max of $f(x,y,z) = x + 2y + z$.

on $X = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2=1, \ y-z=2\}$

Δ X is compact and f continuous hence $f_{|X}$ has a min and a max

Define $g_1(x,y,z) = x^2+y^2-1$ $\quad g_2(x,y,z) = y-z-2$

then $Dg_1(x,y,t) = (2x, 2y, 0)$ $\quad \nabla g_2(x,y,z) = (0,1,-1)$

Hence $Dg_1(x,y,z)$ and $Dg_2(x,y,z)$ are linearly independant on X

Let $p = (x_0,y_0,z_0)$ be a local extremum of f on X then

by Lagrange multipliers theorem, $\exists \lambda_1, \lambda_2 \in \mathbb{R}$ s.t.

$$\nabla f(x_0,y_0,z_0) = \lambda_1 Dg_1(x_0,y_0,z_0) + \lambda_2 Dg_2(x_0,y_0,z_0)$$

$$\Rightarrow \quad (1,2,1) = \lambda_1(2x_0,2y_0,0) + \lambda_2(0,1,-1)$$

from the last component, we get that $\lambda_2 = -1$

Hence $\lambda_1(2x_0,2y_0,0) = (1,2,1) + (0,1,-1) = (1,3,0)$

and $x_0 = \frac{1}{2\lambda_1}, \ y_0 = \frac{3}{2\lambda_1}, \ z_0 = y_0 - 2 = \frac{3-4\lambda_1}{2\lambda_1}$ or $\frac{3}{2\lambda_1} - 2$

From $x_0^2+y_0^2=1$ we get $\lambda_1 = \pm\frac{\sqrt{10}}{2}$

the local extrema has to be at $P_1 = (1/\sqrt{10}, \frac{3}{\sqrt{10}}, 3/\sqrt{10} - 2)$

and $P_2 = (-1/\sqrt{10}, -3/\sqrt{10}, -2-3/\sqrt{10})$

$f(P_1) = \sqrt{10}-2 > -2-\sqrt{10} = f(P_2)$

↳ max $\qquad$ ↳ min

$\square$

Theorem (Spectral theorem)

Let $A \in M_{n,n}(\mathbb{R})$ be a symmetric matrix (ie. $A^t = A$)

Then there is an orthogonal basis of $\mathbb{R}^n$ made of eigenvectors of A

Δ Proof by induction on n: if $n = 1$ OK.

Assume that the statement holds for $n-1$.

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be defined by $f(x) = x^t A x$

then f is differentiable and $\nabla f(x) = 2Ax$

Indeed: $f(x+h) = (x+h)^t A (x+h)$

$$= x^t A x + h^t A x + x^t A h + h^t A h$$

$$= f(x) + (Ax) \circ h + (Ax)^t h + h^t A h$$

$$= f(x) + 2(Ax) \circ h + h^t A h$$

and $|h^t A h| = |h \circ (Ah)| \leq \|h\| \cdot \|Ah\|$ by Cauchy-Schwarz

hence $\dfrac{|h^t A h|}{\|h\|} = \|Ah\| \xrightarrow[h \to 0]{} 0$ by continuity.

Define $g : \mathbb{R}^n \to \mathbb{R}$ by $g(x) = \|x\|^2 = x^t x$

Then $X = \{x \in \mathbb{R}^n : g(x) = 1\}$ is compact and $f_{|X}$ has a max v

Recall that $\nabla g(x) = 2x \neq \vec{0}$ for $x \neq \vec{0} \notin X$, hence, by Lagrange multipliers theorem $\exists \lambda \in \mathbb{R}$, $\nabla f(v) = \lambda \nabla g(v)$

$$\Rightarrow 2Av = 2\lambda v$$

$$\Rightarrow Av = \lambda v$$

hence v is an eigenvector of A

Now, if $x \in \langle v \rangle^{\perp}$ then $(Ax) \cdot v = (Ax)^{t} v = x^{t} A^{t} v = x A v = \lambda x \cdot v = 0$

so $x \in \langle v \rangle^{\perp} \Rightarrow Ax \in \langle v \rangle^{\perp}$

Hence, in a basis w.r.t. $\mathbb{R}^{n} = \langle v \rangle \oplus \langle v \rangle^{\perp}$

$$A = \begin{pmatrix} \lambda & 0 \text{ — } 0 \\ \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} & B \end{pmatrix}$$

with $\underset{\substack{\cap \\ M_{n-1,n-1}(\mathbb{R})}}{B}$ symmetric, so we may conclude by the induction hypothesis □

The Implicit Function Theorem.

Theorem: $U \subset \mathbb{R}^m$ open, $V \subset \mathbb{R}^p$ open $\left(\Rightarrow U \times V \subset \mathbb{R}^{m+p} \text{ open}\right)$

$$F : U \times V \longrightarrow \mathbb{R}^p \text{ of class } C^1$$
$$(x, y) \longmapsto F(x,y) \qquad \text{ie } x \in \mathbb{R}^m, y \in \mathbb{R}^p$$

Let $(x_0, y_0) \in U \times V$.

If $D_yF(x_0,y_0)$ is invertible (ie $\det(D_yF(x_0,y_0)) \neq 0$)

then $\exists r, s > 0$ s.t. $B(x_0,r) \subset U$, $B(y_0,s) \subset V$ and $\varphi : B(x_0,r) \to B(y_0,s)$ of class C^1

such that $\forall (x,y) \in B(x_0,r) \times B(y_0,s)$, $F(x,y) = F(x_0,y_0) \Leftrightarrow y = \varphi(x)$.

Remark 0: $F(x,y) = F(x_0,y_0)$ defines implicitely a function $y = \varphi(x)$ around (x_0,y_0)

Remark 1: $D_yF(x_0,y_0)$ is the Jacobian matrix of $\begin{matrix} \overset{\subset \mathbb{R}^p}{V} \longrightarrow \mathbb{R}^p \\ y \longmapsto F(x_0,y) \end{matrix}$

$$\text{ie } D_yF(x_0,y_0) = \begin{pmatrix} \frac{\partial F_1}{\partial y_1}(x_0,y_0) & \cdots & \frac{\partial F_1}{\partial y_p}(x_0,y_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_p}{\partial y_1}(x_0,y_0) & \cdots & \frac{\partial F_p}{\partial y_p}(x_0,y_0) \end{pmatrix} \in M_{p,p}(\mathbb{R})$$

Similarly we define

$$D_xF(x_0,y_0) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x_0,y_0) & \cdots & \frac{\partial F_1}{\partial x_m}(x_0,y_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_p}{\partial x_1}(x_0,y_0) & \cdots & \frac{\partial F_p}{\partial x_m}(x_0,y_0) \end{pmatrix} \in M_{p,m}(\mathbb{R})$$

$$\underset{\in M_{p,m+p}(\mathbb{R})}{DF(x_0,y_0)} = \left(\underbrace{\begin{matrix} \frac{\partial F_1}{\partial x_1}(x_0,y_0) & \cdots & \frac{\partial F_1}{\partial x_m}(x_0,y_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_p}{\partial x_1}(x_0,y_0) & \cdots & \frac{\partial F_p}{\partial x_m}(x_0,y_0) \end{matrix}}_{D_xF(x_0,y_0)} \; \underbrace{\begin{matrix} \frac{\partial F_1}{\partial y_1}(x_0,y_0) & \cdots & \frac{\partial F_1}{\partial y_p}(x_0,y_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_p}{\partial y_1}(x_0,y_0) & \cdots & \frac{\partial F_p}{\partial y_p}(x_0,y_0) \end{matrix}}_{D_yF(x_0,y_0)} \right)$$

Remark: $F(x_0, y_0) = F(x_0, y_0)$ so $y_0 = \varphi(x_0)$ by $(*)$

Remark: $F(x, \varphi(x)) = F(x_0, y_0) \quad \forall\, x \in B(x_0, r)$

$$\Rightarrow DF(x_0, \underset{\overset{\shortparallel}{y_0}}{\varphi(x_0)}) \begin{pmatrix} I_{m,m} \\ D\varphi(x_0) \end{pmatrix} = 0$$

↳ the RHS is constant

↳ by the chain rule applied to $F \circ G(x)$ where $G(x) = (x, \varphi(x))$

$$\Rightarrow \begin{pmatrix} D_x F(x_0,y_0) & D_y F(x_0,y_0) \end{pmatrix} \begin{pmatrix} I_{m,m} \\ D\varphi(x_0) \end{pmatrix} = 0$$

$$\Rightarrow D_x F(x_0,y_0) + D_y F(x_0,y_0)\, D\varphi(x_0) = 0$$

$$\Rightarrow D\varphi(x_0) = -\left[D_y F(x_0,y_0)\right]^{-1} D_x F(x_0,y_0)$$

recall that $D_y F(x_0,y_0)$ is invertible

Ccl: We know how to compute $D\varphi(x_0)$ in terms of F

$$\boxed{D\varphi(x_0) = -\left[D_y F(x_0,y_0)\right]^{-1} D_x F(x_0,y_0)}$$

You should know this formula (or better: be able to quickly recover it)

Special case of the IFT when $p = 1$

Theorem: $\mathcal{U} \subset \mathbb{R}^m$ open, $I = (a,b)$, $F : \begin{matrix} \mathcal{U} \times I \longrightarrow \mathbb{R} \\ (x_1, \dots, x_m, y) \mapsto F(x_1, \dots, x_m, y) \end{matrix}$ C^1

If $\frac{\partial F}{\partial y}(x_0, y_0) \neq 0$ then there exist $r, s > 0$ with $B(x_0, r) \subset \mathcal{U}$ and $(y_0 - s, y_0 + s) \subset I$ and $\varphi : B(x_0, r) \to (y_0 - s, y_0 + s)$ C^1 s.t. $\forall (x,y) \in B(x_0, r) \times (y_0 - s, y_0 + s)$, $F(x,y) = F(x_0, y_0) \Leftrightarrow y = \varphi(x)$

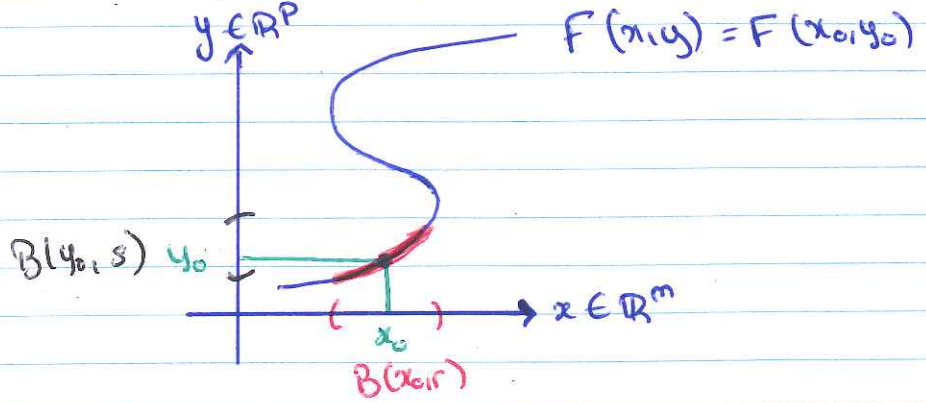
Remark: by computing the $\frac{\partial}{\partial x_i}$'s derivative at x_0 of $F(x, \varphi(x)) = F(x_0, y_0)$

we get:

$$\boxed{\frac{\partial \varphi}{\partial x_i}(x_0, y_0) = -\frac{\frac{\partial F}{\partial x_i}(x_0, y_0)}{\frac{\partial F}{\partial y}(x_0, y_0)}}$$

⚠

Geometric interpretation:

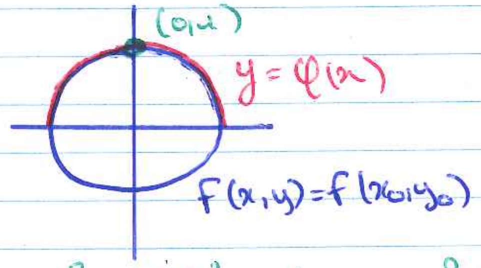


Under the assumptions of the IFT, the level set $F(x,y) = F(x_0,y_0)$ define locally around (x_0,y_0) a function $y = \varphi(x)$ of class C^1

Example:

$$F(x,y) = x^2+y^2, \quad (x_0,y_0) = (0,1), \quad \frac{\partial F}{\partial y}(0,1) = 2 \neq 0$$

$$F(x,y) = F(x_0,y_0) \Leftrightarrow x^2+y^2 = 1$$



$$\varphi : \begin{array}{ccc} (-1,1) & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \sqrt{1-x^2} \end{array}$$

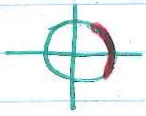
$$F(x,\varphi(x)) = 1 \Rightarrow x^2 + \varphi(x)^2 = 1 \Rightarrow 2x + 2\varphi(x)\varphi'(x) = 0$$

$$\Rightarrow 2\varphi(0)\varphi'(0) = 0$$

$$\Rightarrow \varphi'(0) = 0$$

Remark: at $(1,0)$ $\frac{\partial F}{\partial y}(1,0) = 0$ but $\frac{\partial F}{\partial x}(1,0) = 2 \neq 0$

so we can express $F(x,y) = 1$ as a function $x = \varphi(y)$



Homework: Questions from 3.1

⚠

Heuristic behind the IFT (it's not a proof!)

We want to solve $F(x,y) = F(x_0,y_0)$ around (x_0,y_0)

where the unknown is y (ie we want $F(x,\varphi(x)) = F(x_0,y_0)$)

↳ y in terms of x

By Taylor's theorem,

I am forgetting these terms: that's where I cheat...

$$F(x,y) = F(x_0,y_0) + D_xF(x_0,y_0)(x-x_0) + D_yF(x_0,y_0)(y-y_0) + \cdots$$

Hence $F(x_0,y_0) = F(x,y)$ becomes

$$\cancel{F(x_0,y_0)} = \cancel{F(x_0,y_0)} + D_xF(x_0,y_0)(x-x_0) + D_yF(x_0,y_0)(y-y_0) + \cdots$$

$$\Rightarrow 0 = D_xF(x_0,y_0)(x-x_0) + D_yF(x_0,y_0)y - D_yF(x_0,y_0)y_0 + \cdots$$

$$\Rightarrow D_yF(x_0,y_0)y = D_yF(x_0,y_0)y_0 + D_xF(x_0,y_0)x_0 - D_xF(x_0,y_0)x + \cdots$$

mult by $(D_yF)^{-1}$

$$\Rightarrow y = y_0 + [D_yF(x_0,y_0)]^{-1} D_xF(x_0,y_0)x_0 - [D_yF(x_0,y_0)]^{-1} D_xF(x_0,y_0)x + \cdots$$

↳ since $D_yF(x_0,y_0)$ is invertible

so we expressed y in terms of x (modulo some small errors in the ...)

and the linear part gives the differential of $y = \varphi(x)$:

$$D\varphi(x_0) = -[D_yF(x_0,y_0)]^{-1} D_xF(x_0,y_0)$$

Ex: Prove that $x+y+z+\sin(xyz)=0$ define z as a function of x and y in a neighborhood of $(0,0,0)$

what are $\frac{\partial z}{\partial x}(0,0)$ and $\frac{\partial z}{\partial y}(0,0)$?

Δ $f(x,y,z)=x+y+z+\sin(xyz)$ is C^1

$$\frac{\partial f}{\partial z}(0,0,0)=1\neq 0$$

Hence according to the IFT, $\exists\, r,\delta>0$ and $g:\overset{B_1}{\overbrace{B(0,0;r)}}\to\overset{B_2}{\overbrace{B(0;\delta)}}$

s.t. $\forall((x,y),(z))\in B_1\times B_2$, $g(x,y)=z \Leftrightarrow f(x,y,z)=0$

By the formula given in class:

$$\frac{\partial g}{\partial x}(0,0)=-\frac{\frac{\partial f}{\partial x}(0,0)}{\frac{\partial f}{\partial z}(0,0)}=-1$$

$$\frac{\partial g}{\partial y}(0,0)=-\frac{\frac{\partial f}{\partial y}(0,0)}{\frac{\partial f}{\partial z}(0,0)}=-1$$

$\square$

Proof of Lagrange multipliers theorem:

$f, g_1, \dots, g_p : \mathcal{U} \longrightarrow \mathbb{R}$ C^1, $\mathcal{U} \subset \mathbb{R}^m$ open

$X = \bigcap_{i=1}^{p} g_i^{-1}(0)$

If a is a local extremum of $f_{|X}$ and $\nabla g_1(a), \dots, \nabla g_p(a)$ are linearly independant

Then $\exists \lambda_1, \dots, \lambda_p \in \mathbb{R}$ s.t. $\nabla f(a) = \sum_{i=1}^{m} \lambda_i \nabla g_i(a)$

Δ We may assume that $p < m$. (if $p = m$ then $\nabla g_i(a)$ is a basis of $\mathbb{R}^m$ and $\nabla f(a) \in \mathbb{R}^m$, so there is nothing to prove, for $p > m$, $\nabla g_i(a)$ can't be lin indep)

Define $g : \mathcal{U} \longrightarrow \mathbb{R}^p$ by $g(x) = (g_1(x), \dots, g_p(x))$

$$Dg(a) = \begin{pmatrix} \frac{\partial g_1(a)}{\partial x_1} & \cdots & \frac{\partial g_1(a)}{\partial x_{m-p}} & \frac{\partial g_1(a)}{\partial x_{m-p+1}} & \cdots & \frac{\partial g_1(a)}{\partial x_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_p(a)}{\partial x_1} & \cdots & \frac{\partial g_p(a)}{\partial x_{m-p}} & \underbrace{\frac{\partial g_p(a)}{\partial x_{m-p+1}}} & \underbrace{\cdots} & \underbrace{\frac{\partial g_p(a)}{\partial x_m}} \end{pmatrix}$$

invertible (up to reording the variables) by the assumption that $\nabla g_1(a), \dots, \nabla g_p(a)$ lin indep

By the IFT, $\exists \varphi : B(\overbrace{a_1, \dots, a_{m-p}; r}^{B_1}) \longrightarrow \overbrace{B(a_{m-p+1}, \dots, a_m; \delta)}^{B_2}$

s.t. $g(u, v) = 0 \Leftrightarrow v = \varphi(u)$

hence $X \cap (B_1 \times B_2) = \{ (u, v) \in B_1 \times B_2 : v = \varphi(u) \}$ $(*)$

Define $h : B_1 \to \mathbb{R}$ by $h(x_1, \dots, x_{m-p}) = f(\underbrace{x_1, \dots, x_{m-p}, \varphi(x_1, \dots, x_{m-p})}_{\in X \text{ by } (*)})$

then $x = (a_1, \dots, a_{m-p})$ is a local extremum of h and by the first derivative test and the chain rule

$\forall i = 1, .., m-p$

$$0 = \frac{\partial h}{\partial x_i}(\alpha) = \frac{\partial f}{\partial x_i}(a) + \sum_{j=1}^{p} \frac{\partial f}{\partial x_{m-p+1}}(a) \, \frac{\partial \varphi_j}{\partial x_i}(\alpha) \qquad Ⓐ$$

From $g(x_1, .., x_{m-p}, \varphi(x_1, .., x_{m-p})) = 0$, we obtain

$\forall i = 1, .., m-p$

$\forall s = 1, .., p$

$$0 = \frac{\partial g_s}{\partial x_i}(a) + \sum_{j=1}^{p} \frac{\partial g_s}{\partial x_{m-p+1}}(a) \, \frac{\partial \varphi_j}{\partial x_i}(\alpha) \qquad Ⓑ$$

Since the relations Ⓐ and Ⓑ are similar, the following matrix is of rank $\leq p$

$$\begin{pmatrix} \frac{\partial f}{\partial x_1}(a) & \cdots & \frac{\partial f}{\partial x_{m-p}}(a) & \frac{\partial f}{\partial x_{m-p+1}}(a) & \cdots & \frac{\partial f}{\partial x_m}(a) \\ \frac{\partial g_1}{\partial x_1}(a) & \cdots & \frac{\partial g_1}{\partial x_{m-p}}(a) & \frac{\partial g_1}{\partial x_{m-p+1}}(a) & \cdots & \frac{\partial g_1}{\partial x_m}(a) \\ \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ \frac{\partial g_p}{\partial x_1}(a) & \cdots & \frac{\partial g_p}{\partial x_{m-p}}(a) & \frac{\partial g_p}{\partial x_{m-p+1}}(a) & \cdots & \frac{\partial g_p}{\partial x_m}(a) \end{pmatrix} \in M_{p+1,m}(\mathbb{R})$$

hence the rows are linearly dependant : $\exists \lambda_1, .., \lambda_p, \mu$ st.

$$\sum \lambda_i \nabla g_i(a) + \mu \nabla f(a) = 0$$

and $\mu \neq 0$ since the family $(\nabla g_i(a))$ is linearly independant $\square$