Lecture Chapter 3

Freitag, 27. August 2021 13:55

$f^{-1}: Y \mapsto X, \; y \mapsto x = f^{-1}(y)$; function

$f^{-1}(y)$: f^{-1} evaluated at a fixed y ; $f^{-1}(y) \in X$

both these quantities may or may not exist, depends on whether f is bijective

$f^{-1}[\{y\}]$: pre-image of $\{y\}$ under f

$= \{x \in X : f(x) = y\}$; a set

the only similarity / relationship is :

$f^{-1}[\{y\}] = \{f^{-1}(y)\}$ if f is bijective / invertible

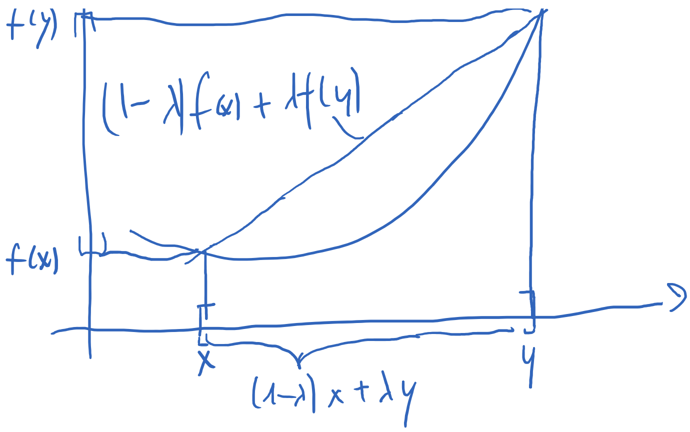
6 $\quad y \neq x$: if $y = x$

$$f(\lambda x + (1-\lambda) y) = f(x) = \lambda f(x) + (1-\lambda) f(x)$$
$$= \lambda f(x) + (1-\lambda) f(y)$$

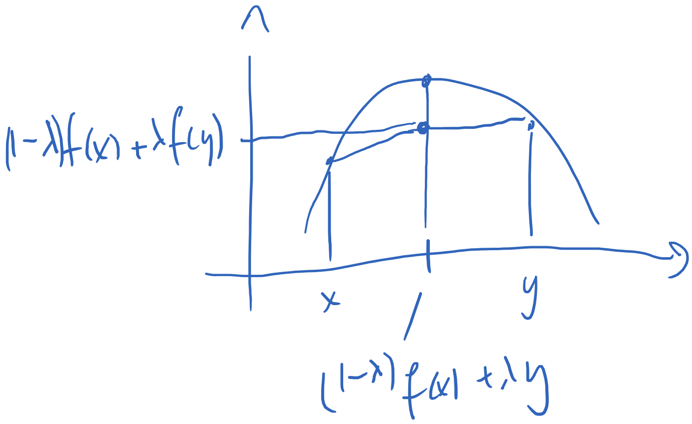
$\to$ strict ineq. always ruled out !

7 $\quad (1-\lambda) x + \lambda y = x + \lambda (y - x)$

convex



concavity



Univariate convexity

1. pick two arbitrary points x, y
2. check convexity "in between" these points: starting from x, and considering

$$\tilde{x} = (1-\lambda)x + \lambda y = x + \lambda(y - x)$$

=> move from x in direction $z = y - x$ and ask: does convexity hold in this direction?

$\underset{\sim}{g}$

Equivalence proof: we show it in two steps

Claim: $\underbrace{(f \text{ is strictly convex})}_{=: A} \iff \underbrace{\left(\forall x \in X\ \forall z \in \mathbb{R}^n \setminus \{0\}:\ g: \mathbb{R} \mapsto \mathbb{R},\ t \mapsto f(x+tz) \text{ is strictly convex.}\right)}_{=: B}$

"$\Rightarrow$" $\quad (A \Rightarrow B)$

Start by assuming the premise: suppose that A is true, i.e. that f is strictly convex.

Start by assuming the premise. Suppose that A is true, i.e. that f is strictly convex.

Claim: $\forall x \in X \ \forall z \in \mathbb{R}^n \setminus \{0\}$: g is str. convex.

Let $x \in X$, $z \in \mathbb{R}^n \setminus \{0\}$. Then,

let $t, s \in \mathbb{R}$, $t \neq s$ and $\lambda \in (0,1)$, for which

$$g(\lambda t + (1-\lambda)s) = f(x + (\lambda t + (1-\lambda)s)\, z)$$

$$= f((\lambda + 1 - \lambda)x + (\lambda t + (1-\lambda)s)\, z)$$

$$= f(\lambda(x + tz) + (1-\lambda)(x + sz))$$

$$\overset{f \text{ str. convex}}{<} \lambda f(x+tz) + (1-\lambda) f(x+sz)$$

$$= \lambda g(t) + (1-\lambda) g(s)$$

✓

$\Rightarrow g$ is str. convex.

"$\Leftarrow$" $(B \Rightarrow A)$ Suppose that B is true

Claim: f is str. convex, i.e. $\forall x, y \in X$, $x \neq y$, $\lambda \in (0,1)$:

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$$

Let $x, y \in X$, $x \neq y$ and $\lambda \in (0,1)$. Then,

$$f((1-\lambda)x + \lambda y) = f(x + \lambda \underbrace{(y-x)}_{= z \neq 0})$$

$$= g(\lambda) = g(\lambda \cdot 1 + (1-\lambda) \cdot 0)$$

$$< \lambda g(1) + (1-\lambda) g(0) \quad \text{by str. convexity of g}$$

$$= \lambda \cdot f(x + 1 \cdot (y-x)) + (1-\lambda) f(x + 0 \cdot (y-x))$$

$$= \lambda f(y) + (1-\lambda) f(x).$$

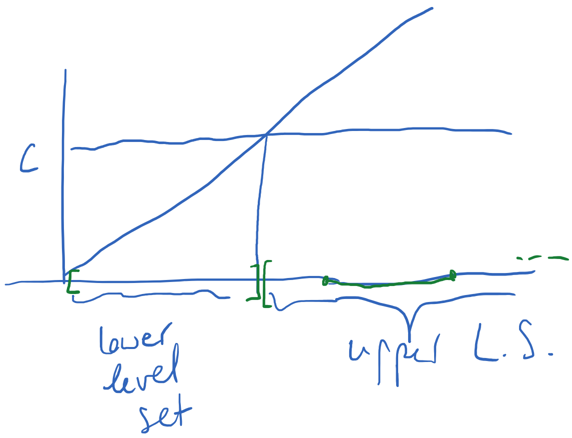
$\Rightarrow f$ is str. convex. ✓

10

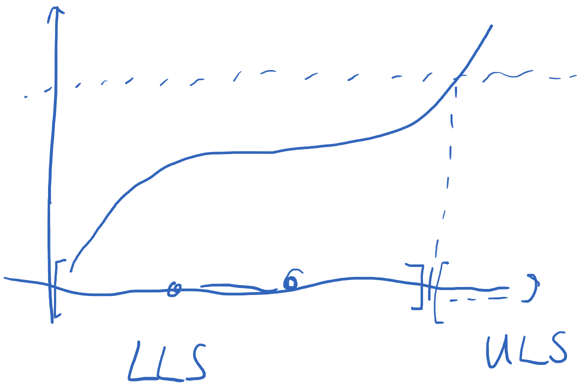
not convex: multiplicative component $\sqrt{x_n}$ is strictly concave => not convex in direction $e_n = (0, \dots, 0, 1)'$

13

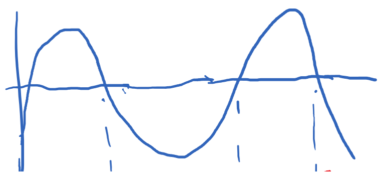
A	only quasi-concave
B	only quasi-convex
C	quasi-linear
D	none of the above



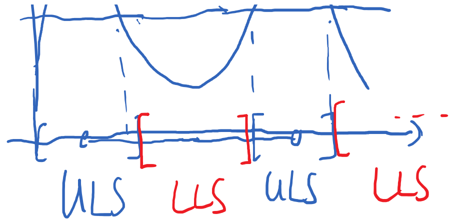
C



C



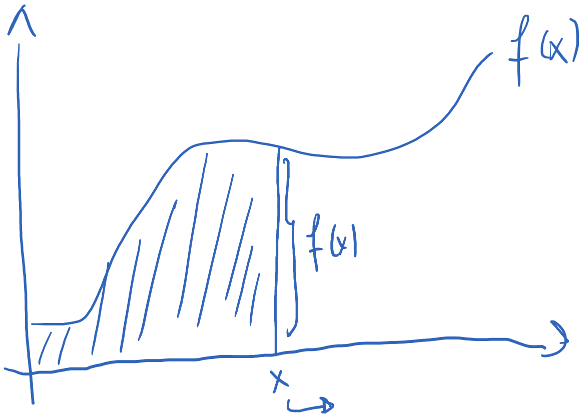
D



D

15

f = " instantaneous rate of its accumulation"



if I marginally change x, what change will I introduce to the area of accumulation?

$\to f(x)$!

marginal change of accumulation = function itself

18

Highest level: Operator level - mapping between function spaces

intermediate level: function level - mapping between sets of (real) vectors

lowest level: value level - function evaluated at specific points

26

Fact: for any norm $\|\cdot\|$ on $\mathbb{R}$, $\exists c > 0$ s.th.

$\forall x \in \mathbb{R}: \quad \|x\| = c \cdot |x|$

why: $\|x\| = \|\underbrace{x}_{\in \mathbb{R}} \cdot 1\| = |x| \cdot \underbrace{\underbrace{\|1\|}_{\neq 0}}_{=: c > 0}$

Let $\|\cdot\|$ be a norm on $\mathbb{R}$ with $\|\cdot\| = c \cdot |\cdot|$.

Then,

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

initial definition of derivative at x_0 for univariate case

$$\Leftrightarrow |0| = \left| \lim_{h \to 0} \left| \frac{f(x_0+h) - f(x_0)}{h} - d^0 \cdot \frac{h}{h} \right| \right|$$

$$= \lim_{h \to 0} \left| \frac{f(x_0+h) - f(x_0) - d^0 h}{h} \right|$$

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

$$\Leftrightarrow \underbrace{\frac{1}{c} \cdot 0}_{=0} = \lim_{h \to 0} \frac{|f(x_0+h) - f(x_0) + d^0 h|}{\|h\|} \quad \text{as } \|h\| = c \cdot |h|$$

found an **equivalent** condition that we can generalize to the multivariate case!

30 $$f(x_0 + t e_j) = f\left(\begin{pmatrix} x_{01} \\ \vdots \\ x_{0n} \end{pmatrix} + t \cdot \begin{array}{cl} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \leftarrow \text{pos. } j \end{array} \right)$$

Recap - where we stopped

1. we wanted to generalize the limit statement for the derivative to higher-dimensional spaces
2. finding an equivalent statement using norms allowed us to do so
3. concrete method: start from intuition: summarize expansion along axes of the space
4. turns out that it works

32 $$\nabla f(x_1, x_2) = \left(\frac{\partial f}{\partial x_1}(x_1, x_2), \frac{\partial f}{\partial x_2}(x_1, x_2) \right)$$

$$\nabla f^1(x_1, x_2) = (1 \; , \; 1)$$

$$\nabla f^2(x_1, x_2) = (x_2 \; , \; x_1)$$

$$\nabla f^3(x_1, x_2) = (x_2^2 - \sin(x_1) \; , \; 2x_1 x_2)$$

35: ex: if $A = \begin{pmatrix} 2 & 4 \\ 1 & -1 \end{pmatrix}$ and $f: \mathbb{R}^2 \mapsto \mathbb{R}^2, x \mapsto Ax$,

then $\forall x \in \mathbb{R}^2$:

$$f(x) = Ax = \begin{pmatrix} 2 & 4 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 + 4x_2 \\ x_1 - x_2 \end{pmatrix} = \begin{pmatrix} f^{(1)}(x_1, x_2) \\ f^{(2)}(x_1, x_2) \end{pmatrix}$$

39
$$\frac{d f \circ g}{dx} = \left(\frac{\partial f}{\partial y} \circ g\right) \frac{dy}{dx} + \frac{\partial f}{\partial x} \circ g$$ is more correct

Addition?
$$\frac{dg}{dx}(x) = \begin{pmatrix} \nabla y(x) \\ I_n \end{pmatrix} \qquad (x \in I_n)$$

$$(A_1 \;\; A_2) \frac{dg}{dx}(x) = (A_1 \;\; A_2) \begin{pmatrix} \nabla y(x) \\ I_n \end{pmatrix}$$
$$= A_1 \nabla y(x) + A_2 I_n$$
$$= A_1 \nabla y(x) + A_2$$

41

Hessian is actually nothing but a specific example of a Jacobian, namely the one you get when choosing $\nabla f'$ as the function to be differentiated

43

$$f(x_1, x_2, x_3) = x_1 x_2^2 + \cos(x_1)\, e^{x_3}$$

f is a composition of infinitely many times differentiable terms, and thus infinitely many times continuously partially differentiable.

Thus, the second derivative exists, and it is equal to a symmetric Hessian.

$$\nabla f(x_1, x_2, x_3) = \left(x_2^2 - \sin(x_1) e^{x_3},\ 2x_1x_2,\ \cos(x_1) e^{x_3}\right)$$

$$f_{11}(x_1, x_2, x_3) = -\cos(x_1) e^{x_3}$$

$$f_{12}(x_1, x_2, x_3) = 2x_2$$

$$f_{13}(x_1, x_2, x_3) = -\sin(x_1) e^{x_3}$$

$$f_{22}(x_1, x_2, x_3) = 2x_1$$

$$f_{23}(x_1, x_2, x_3) = 0$$

$$f_{33}(x_1, x_2, x_3) = \cos(x_1) e^{x_3}$$

$$H_f(x_1, x_2, x_3) = \begin{pmatrix} -\cos(x_1) e^{x_3} & 2x_2 & -\sin(x_1) e^{x_3} \\ 2x_2 & 2x_1 & 0 \\ -\sin(x_1) e^{x_3} & 0 & \cos(x_1) e^{x_3} \end{pmatrix}$$

45

$$f(x_0 + tz) = f(g(t)), \quad g(t) = x_0 + tz = \begin{pmatrix} x_{0,1} + tz_1 \\ \vdots \\ x_{0,n} + tz_n \end{pmatrix}$$

$$\frac{d}{dt} g(t) = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = z$$

Chain rule:

$$\frac{d\, f \circ g}{dt}(t) = \left(\frac{df}{dx} \circ g\right)(t) \cdot \frac{d}{dt} g(t)$$

$$= \nabla f(g(t)) \cdot z$$

$$= \nabla f(x_0 + tz) \cdot z$$

Dir. derivative:

$$\boxed{\frac{d\, f \circ g}{dt}(0) = \nabla f(x_0) \cdot z}$$

This is the general representation/formula for the directional derivative!

45 Interpretation:

fine: labor productivity increases by twice as much as capital input, then output increases by X times the amount as the capital input
=> fix capital as the "baseline change"; consider variation relative to this baseline change

not fine: labor productivity increases by two as capital input increases by 1, then output increases by X units
=> loses "marginal" nature of derivative

=> never use total derivative for fixed, non-zero variations (but only relative local variation)

49 e.g. $X = (0,1)$

$$\min\{b \in \mathbb{R} : (b \geq x \ \forall x \in X)\} = 1 = \sup(X)$$

<u>5b</u>

$$\int_{A\times B} f(x)g(y)\,d(x,y) \overset{\text{Fubini}}{=} \int_A \left(\int_B f(x) g(y)\,dy \right) dx$$

$$= \int_A f(x) \left(\int_B g(y)\,dy \right) dx$$

as $f(x)$ is const. w.r.t. y

$$= \left(\int_B g(y)\,dy \right) \cdot \int_A f(x)\,dx$$

as $\int_B g(y)dy$ is const. (w.r.t. x)