30/11/2021

Lecture - 18

$f_0 \to$ convex function.

$\mathcal{D} \to$ convex set

$f : \Omega \to \mathbb{R}$

$\Omega \subseteq \mathbb{R}^n$

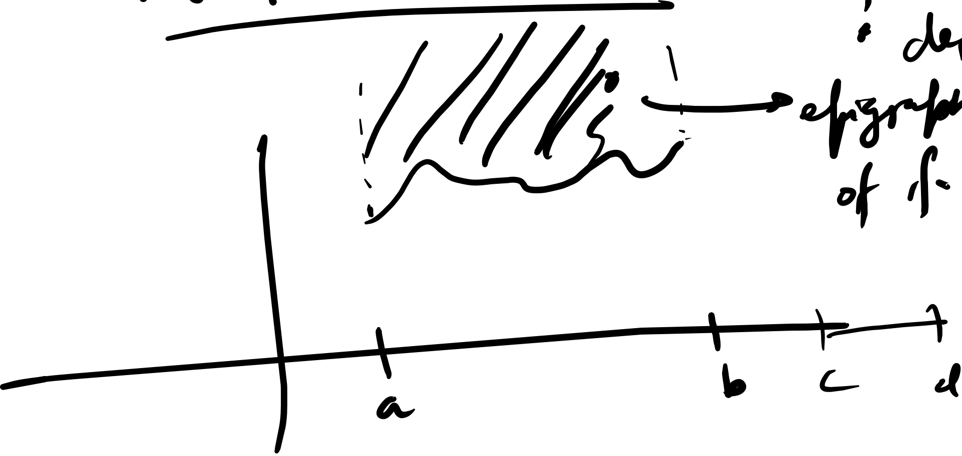
convex domain $(\mathrm{int}(\Omega) \neq \Phi)$

$$f(\lambda \vec{x} + (1-\lambda)\vec{y}) \leq \lambda f(\vec{x}) + (1-\lambda) f(\vec{y})$$

$\forall \vec{x}, \vec{y} \in \Omega$

Then f is said to be $0 \leq \lambda \leq 1$. convex function..

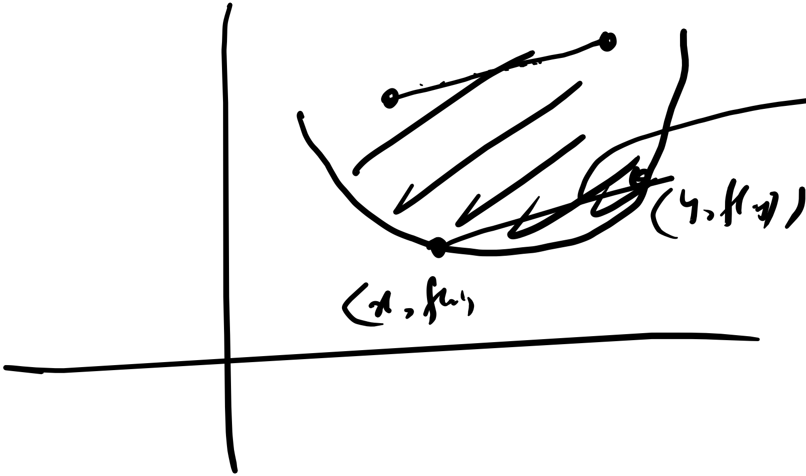
epigraph of f ; (f is a function defined on some subset of $\mathbb{R}^n$)



$f : [a,b] \to \mathbb{R}$

graph of f := $\{ (x, f(x)) : x \in [a,b] \}$

epigraph of f := $\{ (x, y) : x \in \underbrace{[a,b]}_{\mathrm{dom}(f)} , \ y \geq f(x) \}$



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

f is convex if and only if its epigraph is a convex set.

$f : \Omega \to \mathbb{R}$ (any function)

$\subseteq \mathbb{R}^n$

epigraph of f is $:= \{ (\vec{x}, y) \in \mathbb{R}^n \times \mathbb{R} ; y \geq f(\vec{x}) \}$

$\vec{x} \in \mathbb{R}^n$, $y \in \mathbb{R}$

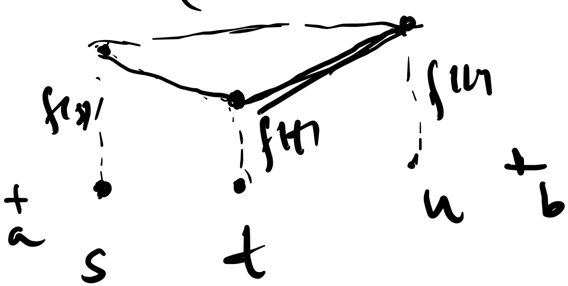
f is said to be convex iff its epigraph is convex.

Let us go back to the $\boxed{n=1}$ case

(dimension 1)

Theorem: Let $f: [a,b] \to \mathbb{R}$ be a convex function. Then for $a < s < t < u < b$, we have

$$\left(\frac{f(t)-f(s)}{t-s} \le \boxed{\frac{f(u)-f(s)}{u-s}} \le \frac{f(u)-f(t)}{u-t}. \right).$$



Proof. Let $t = \lambda u + (1-\lambda) s, \quad 0 \le \lambda \le 1.$

$$= s + \lambda (u - s)$$

$$f(\underbrace{\lambda u + (1-\lambda) s}_{t}) \le \lambda f(u) + (1-\lambda) f(s)$$

$$= f(s) + \lambda (f(u) - f(s))$$

$$\Rightarrow \quad f(t) - f(s) \le \lambda (f(u) - f(s))$$

$$= \frac{t-s}{u-s} (f(u) - f(s))$$

$$\Rightarrow \quad \frac{f(t) - f(s)}{t-s} \le \frac{f(u) - f(s)}{u-s}$$

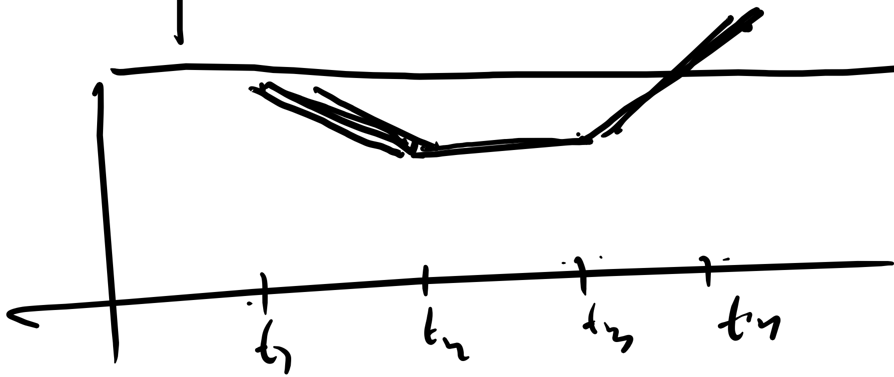
∎

<u>Corollary</u>: Let $f : [a, b] \to \mathbb{R}$ be convex,

and $a \leq t_1 < t_2 < t_3 < t_4 < b$.

Then,

$$\boxed{\frac{f(t_2) - f(t_1)}{t_2 - t_1} \leq \frac{f(t_4) - f(t_3)}{t_4 - t_3}}$$



Corollary: Let $f : [a, b] \to \mathbb{R}$ be convex and differentiable on (a, b). Then for $a < s < t < b$, we have $f'(s) \le f'(t)$.

In other words, f' is an increasing function.

Corollary: Let $f : [a, b] \to \mathbb{R}$ be convex and twice differentiable on (a, b). Then $f''(x) \ge 0$ $\forall x \in (a, b)$.

Lower bounding convex functions with affine functions

Theorem. Let $f: \Omega \to \mathbb{R}$ be convex and differentiable on $\operatorname{int}(\Omega)$. Then f is convex if and only if $\forall \vec{x}, \vec{y} \in \operatorname{int}(\Omega)$

$$f(\vec{y}) \geq f(\vec{x}) + \nabla f(\vec{x})^T (\vec{y} - \vec{x}).$$

(first-order condition for convexity)

(The hyperplane defined by $(\nabla f(\vec{x}), -1)$ supports epi(f) at the boundary point $(\vec{x}, f(\vec{x}))$, which belongs to the graph of f

Reminder:

$$\lim_{\lambda \to 0} \frac{\| f(\vec{x} + \lambda \vec{u}) - f(\vec{x}) - \lambda (\nabla f(\vec{x})^T \vec{u}) \|}{\lambda} = 0$$

if such a vector exists then it is called the total derivative of f at $\vec{x}$.

$$f(\vec{x} + \lambda \vec{u}) = f(\vec{x}) + \lambda \langle \vec{v}, \vec{u} \rangle + o(\lambda)$$

$\lambda \nabla f(\vec{x})$

$$\partial_i f(\vec{x}) = \lim_{\lambda \to 0} \frac{f(\vec{x} + \lambda e_i) - f(\vec{x})}{\lambda} =$$

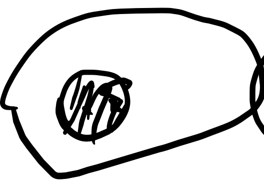
$\vec{v} = \nabla f(\vec{x}) = [\partial_1 f(\vec{x}), \ldots, \partial_n f(\vec{x})]^T$.

Theorem: Let $f : \Omega \to \mathbb{R}$ be convex and $\vec{x} \in \operatorname{int}(\Omega)$.

If f has partial derivatives (of first order) at $\vec{x}$, then f is differentiable at $\vec{x}$.

Proof:

$$\nabla f(\vec{x}) = [\partial_1 f(\vec{x}) \cdots \partial_n f(\vec{x})]$$



$$g(\vec{h}) = f(\vec{x} + \vec{h}) - f(\vec{x}) - \boxed{\langle \nabla f(\vec{x}), \vec{h} \rangle}$$

$\vec{h} \in B(0, r)$

$$\vec{h} = \sum_{i=1}^{n} \eta_i e_i.$$

$$g(\vec{h}) = g\left(\frac{1}{n}\sum_{i=1}^{n} n\eta_i e_i\right)$$

(convexity of g)

$$\leq \frac{1}{n}\left(\sum_{i=1}^{n} g(n\eta_i e_i)\right)$$

$$\langle \vec{h}, \vec{v} \rangle \leq |\vec{h}|\,|\vec{v}| \leq |\vec{h}|\left(\sum_{i=1}^{n} |v_i|\right).$$

$$\vec{v} = \sum_{i=1}^{n} v_i e_i$$

$$|\vec{v}| = (v_1^2 + \dots + v_n^2)^{1/2}$$

$$\leq |v_1| + \dots + |v_n|$$

$$g(\vec{h}) \leq \sum_{i=1}^{n} \eta_i \frac{g(n\eta_i e_i)}{n\eta_i} \to \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

$$\leq \|\vec{h}\| \sum_{i=1}^{n} \left| \frac{g(n\eta_i e_i)}{n\eta_i} \right|.$$

$$g(\vec{h}) + g(-\vec{h}) \geq 0 \quad \text{(by convexity of } g\text{).}$$

$$\underbrace{-\sum_{i=1}^{n} \left| \frac{g(n\eta_i e_i)}{n\eta_i} \right|}_{\text{LHS}} \leq -\frac{g(-\vec{h})}{\|\vec{h}\|} \leq \frac{g(\vec{h})}{\|\vec{h}\|} \leq \underbrace{\sum_{i=1}^{n} \frac{g(n\eta_i e_i)}{n\eta_i}}_{\text{RHS}}.$$

LHS $\to 0$

RHS $\to 0$ as $\|\vec{h}\| \to 0$

$g(\vec{h}) = o(\|\vec{h}\|)$.

$\blacksquare$

Theorem. Let $\Omega \subseteq \mathbb{R}^n$ be convex and open.

Let $f : \Omega \to \mathbb{R}$ be a differentiable function.

TFAE: (a) f is convex

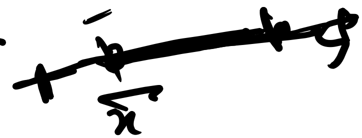
(b) $f(\vec{y}) - f(\vec{x}) \geq \langle \nabla f(\vec{x}), \vec{y} - \vec{x} \rangle$ for $\vec{x}, \vec{y} \in \Omega$

(c) $\langle \nabla f(\vec{y}) - \nabla f(\vec{x}), \vec{y} - \vec{x} \rangle \geq 0$ for $\vec{x}, \vec{y} \in \Omega$

Proof

$$g(\lambda) = f(\bar{x} + \lambda(y - \bar{x}))$$

$$= \left(f(\lambda \bar{x} + (1-\lambda) y) \right)$$



(a) $\Rightarrow$ (b)

g is convex (since f is convex).

$$\frac{g(\lambda) - g(0)}{\lambda - 0} \geq g'(0), \quad \forall \lambda \in [0,1]$$

$$\Rightarrow \underbrace{g(1)}_{f(y)} - \underbrace{g(0)}_{f(\bar{x})} \geq g'(0) = \langle \nabla f(\bar{x}), y - \bar{x} \rangle.$$

(b) $\Rightarrow$ (c),

$$g'(\lambda) - g'(0))\lambda \geq 0$$

$$\Leftrightarrow \quad g'(1) - g'(0) \geq 0$$

$$\Rightarrow \langle \nabla f(\vec{y}), \vec{y} - \vec{x} \rangle \geq \langle \nabla f(\vec{x}), \vec{y} - \vec{x} \rangle$$

$$\Rightarrow \langle \nabla f(\vec{y}) - \nabla f(\vec{x}), \vec{y} - \vec{x} \rangle \geq 0.$$

(c) $\Rightarrow$ (a), $g'_{(\vec{x},\vec{y})}$ is an increasing function, hence $g_{(\vec{x},\vec{y})}(\lambda)$ convex. $\Rightarrow$ f is convex. $\blacksquare$

Theorem : (Second-order condition for convexity)

Let $\Omega \subseteq \mathbb{R}^n$ be convex and open, and let $f : \Omega \to \mathbb{R}$ be twice differentiable.

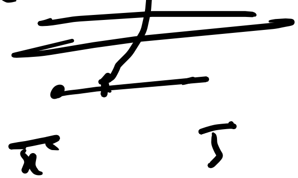
Then f is convex if and only if the Hessian matrix $\mathrm{Hess}\, f(\vec{x}) = \left(\partial_i \partial_j f(\vec{x})\right)_{i,j=1}^{n}$

is positive-semidefinite for all $\vec{x} \in \Omega$.

Proof (A is positive-semidefinite matrix, if $\langle \vec{x}, A\vec{x} \rangle \geq 0 \quad \forall \vec{x} \in \mathbb{R}^n$,

$\boxed{\langle A\vec{x}, \vec{x} \rangle \geq 0}$

$$\lim_{\lambda\to 0^+} \frac{1}{\lambda}\left\langle \nabla f(\vec{x}+\lambda(\vec{y}-\vec{x})) - \nabla f(\vec{x}),\ \vec{y}-\vec{x}\right\rangle \geq 0 \quad \forall \lambda > 0$$



$$= \lim_{\lambda\to 0} \sum_{j=1}^{n} \frac{1}{\lambda}\left[\partial_j f(\vec{x}+\lambda(\vec{y}-\vec{x})) - \partial_j f(\vec{x})\right](\eta_j - \xi_j)$$

$$= \sum_{j=1}^{n} \left(\lim_{\lambda\to 0}\frac{1}{\lambda}\left(\partial_j f(\vec{x}+\lambda(\vec{y}-\vec{x})) - \partial_j f(\vec{x})\right)\right)(\eta_j-\xi_j)$$

$$= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \partial_i\partial_j f(\vec{y})(\eta_i - \xi_i)\right)(\eta_j-\xi_j)$$

$$= \left\langle \vec{y}-\vec{x},\ \mathrm{Hess} f(\vec{x})(\vec{y}-\vec{x})\right\rangle \geq 0$$

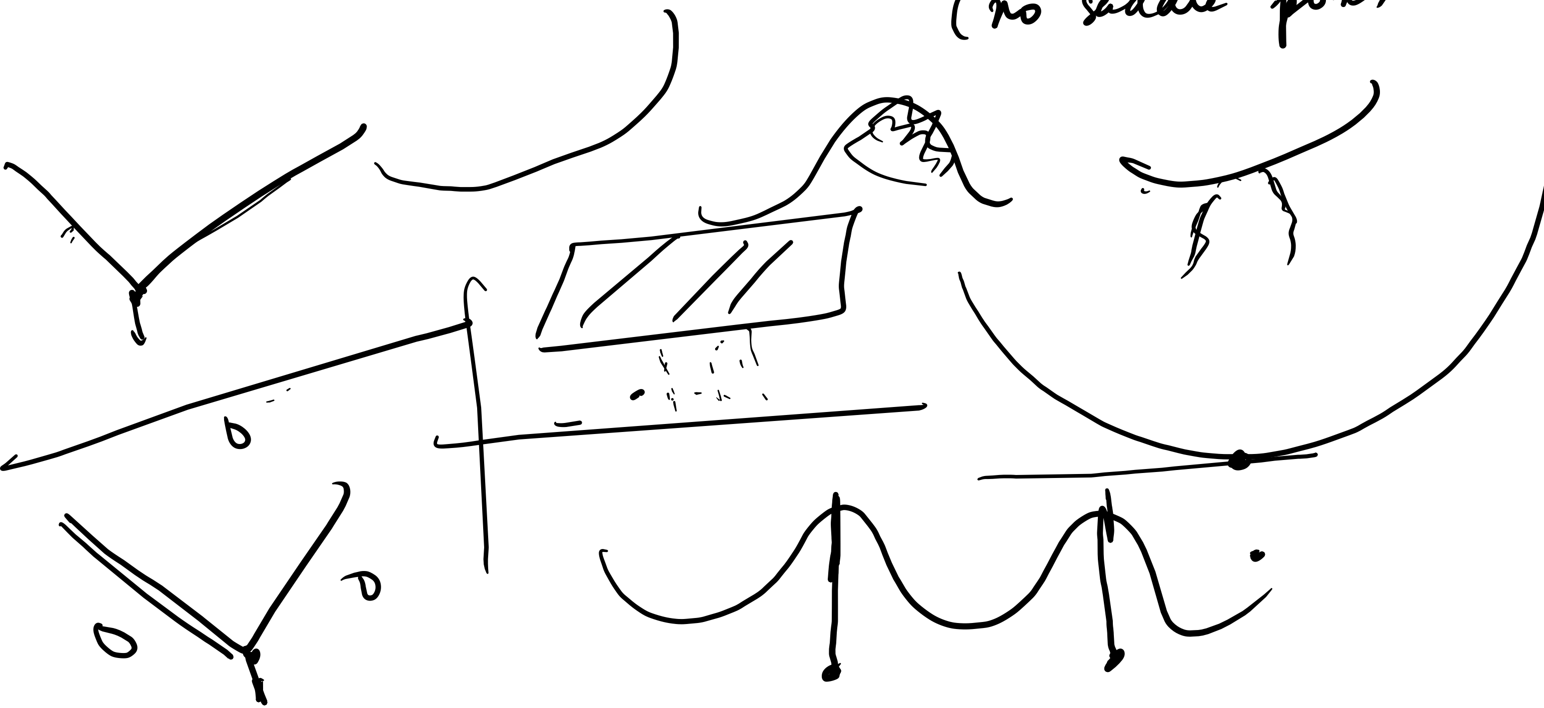
$g_i = \partial_i f$

$\nabla g_i = [\partial_i \partial_j f]_{ij=1}^{n} = \text{Hess } f(\vec{x})$ $[e_i$

$f(\vec{y}) \geq f(\vec{x}) + \nabla f(\vec{x})(\vec{y}-\vec{x}) + \langle \vec{y}-\vec{x}, \text{Hess } (\vec{y}-\vec{x}) \rangle$

$g(\lambda) = f(\vec{x} + \lambda(\vec{y}-\vec{x}))$

Boundary of
* Convex sets have "positive" curvature
(no saddle points



Examples of convex functions on $\mathbb{R}$

(1) Exponential:

$\boxed{e^{ax}}$, is convex on $\mathbb{R}$, for any $a \in \mathbb{R}$

(2) Powers:

$\boxed{x^a}$ is convex on $\mathbb{R}_+$ (set of nonnegative real numbers)

when $a \geq 1$, or $a \leq 0$

$\frac{d^2}{dx^2}(x^a) = a(a-1)\boxed{x^{a-2}}$

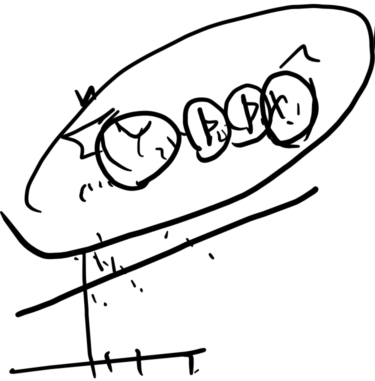
$\geq \quad \geq$

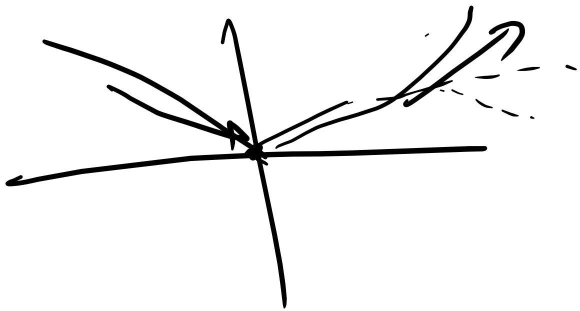
is concave on $\mathbb{R}_+$ for $0 \leq a \leq 1$.

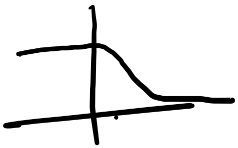
$\boxed{x^n \text{ is convex,}}$

(3) Powers of absolute values.

$|x|^p$, for $p \geq 1$, is convex on $\mathbb{R}$.







(4) Logarithm:

$-\log x$ is convex on $\mathbb{R}_+ \setminus \{0\}$.

$$\frac{d^2}{dx^2}(-\log x) = \frac{1}{x^2} \geq 0.$$

(5) Negative entropy (Shannon entropy)

→ information-theoretic entropy

$\boxed{x \log x}$ (defined as 0 for $x = 0$)

is convex on $\mathbb{R}_+$.

0, .. 1s

01

0000

1, ...

0 1

p_1 p_2

$p_1 = 0.001$

$p_2 = 0.999$

$$\boxed{-p_1 \log p_1 - p_2 \log_2 p_2}$$

$\boxed{-p \log p}$ being big ⇒ you gain a lot of information...

Examples of convex functions on $\mathbb{R}^n$

(1) Norms. $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$

$$\left(\boxed{\|\vec{x}+\vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|} \right.$$
$$\boxed{\|\lambda \vec{x}\| = |\lambda| \, \|\vec{x}\|}$$
$$\left. d(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|. \right)$$

(2) Max. function.

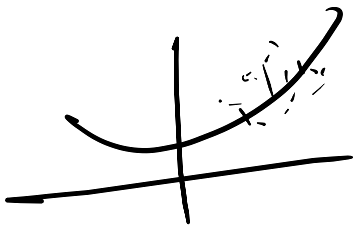
$f(\vec{x}) = \max\{x_1, \ldots, x_n\}$ is convex on $\mathbb{R}^n$.

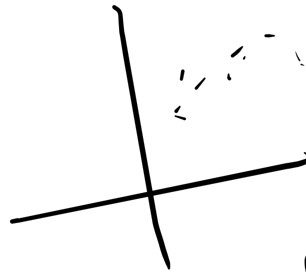
$= \vec{x}_{(1)}$

(3) log-sum-exp

$$f(\vec{x}) = \log(e^{x_1} + \dots + e^{x_n})$$

$$f(\vec{x}) - \log n = \log\left(\frac{e^{x_1} + \dots + e^{x_n}}{n}\right)$$





$$e^{n \max\{x_1, \dots, x_n\}} \le e^{x_1} + \dots + e^{x_n} \le n \max\{e^{x_1}, \dots, e^{x_n}\}$$

f is convex on $\mathbb{R}^n$,

$$\boxed{\max\{x_1, \dots, x_n\}} \le f(\vec{x}) \le \max\{x_1, \dots, x_n\} + \log n.$$

As $n \to \infty$, $\frac{\log n}{n} \to 0$

For $\log n$, $\frac{\max\{x_1, \dots, x_n\}}{n}$ is approximated by $\frac{f(x_1, \dots, x_n)}{n}$.

(4) $f(\vec{x}) = -\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$ is convex on $\mathbb{R}^n_+$.

The geometric mean is concave.

Jensen's inequality :

$f \to$ convex function on $\Omega \subseteq \mathbb{R}^n$,

$$\boxed{f(\theta_1 \vec{x}_1 + \cdots + \theta_k \vec{x}_k) \leq \theta_1 f(\vec{x}_1) + \cdots + \theta_k f(\vec{x}_k)}$$

$$\forall \; 0 \leq \theta_i \leq 1, \quad \sum_{i=1}^{k} \theta_i = 1$$

$$\vec{x}_i \in \Omega.$$

(Feels like a tautology) But the power comes from the analytic characterization of convexity.

Examples -

(1) $\boxed{-\log x}$ is convex on $\mathbb{R}_+ \setminus \{0\}$.

For $a, b > 0$

$$-\log\left(\frac{a+b}{2}\right) \overset{\text{(Jensen's)}}{\le} \frac{-\log a + \log b}{2}$$

$$\Rightarrow \quad \frac{a+b}{2} \ge \exp\left(\log\sqrt{ab}\right) = \sqrt{ab}$$

$$\boxed{AM \ge GM}$$

(2) More generally,

$$-\log(\theta a + (1-\theta) b) \overset{\text{(Jensen's)}}{\leq} -\theta \log a - (1-\theta)\log b$$

$$\underbrace{\theta}_{1/p} a + \underbrace{(1-\theta)}_{1/q} b \geq a^{\theta} b^{1-\theta}$$

$\Rightarrow$

$(x_1, \ldots, x_n)^T$, $(y_1, \ldots, y_n)^T \in \mathbb{R}^n$, $\underbrace{\frac{1}{p}}_{\theta} + \underbrace{\frac{1}{q}}_{1-\theta} = 1$

$$a_i = \frac{|x_i|^{p}}{\sum_{j=1}^{n} |x_j|^p}, \quad b_i = \frac{|y_i|^{q}}{\sum_{j=1}^{n} |y_j|^q}, \quad q = \frac{p}{p-1}, \quad p > 1.$$

With $\theta = \frac{1}{p}$,

$$a_i^{1/p} b_i^{1/q} \le \frac{a_i}{p} + \frac{b_i}{q}$$

$$\frac{|x_i|}{\left(\sum_{j=1}^n |x_j|^p\right)^{1/p}} \cdot \frac{|y_i|}{\left(\sum_{j=1}^n |y_j|^q\right)^{1/q}} \le \frac{a_i}{p} + \frac{b_i}{q}$$

$$\frac{1}{\left(\sum_{j=1}^n |x_j|^p\right)^{1/p}\left(\sum_{j=1}^n |y_j|^q\right)^{1/q}} \sum_{i=1}^n |x_i||y_i| \le \frac{1}{p}\left(\frac{\sum_{i=1}^n |x_i|^p}{\sum_{i=1}^n |x_i|^p}\right) + \frac{1}{q}(1)$$

$$= 1$$

$$\sum_{i=1}^{n} |x_i||y_i| \le \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q}$$

(Hölder's inequality)

$p = q = 2$

(Cauchy Schwarz inequality)

$\vec{x}, \vec{y} \longrightarrow \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

$$\|\vec{x}\|_p = \left(|x_1|^p + \dots + |x_n|^p\right)^{1/p}$$

$\hookrightarrow L^p$-norm

$$\|\vec{x}+\vec{y}\|_p^p = \left(\sum_{i=1}^n |x_i+y_i|^p\right)$$

$$= \sum_{i=1}^n |x_i+y_i|\,|x_i+y_i|^{p-1}$$

$$\le \left(\sum_{i=1}^n |x_i|\,|x_i+y_i|^{p-1}\right) + \left(\sum_{i=1}^n |y_i|\,|x_i+y_i|^{p-1}\right)$$

$$\le \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |x_i+y_i|^p\right)^{p-1/p} + \left(\sum_{i=1}^n |y_i|^p\right)^{1/p} \left(\sum_{i=1}^n |x_i+y_i|^p\right)^{p-1/p}$$

$$\left(\sum_{i=1}^{n} |x_i+y_i|^p\right)^{1}$$

$$\leq \left(\left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{1/p}\right)$$

$$\cdot \left(\sum_{i=1}^{n} |x_i+y_i|^p\right)^{p-1/p} \quad \to 1-\frac{1}{p}$$

$$\Longrightarrow \left(\sum_{i=1}^{n} |x_i+y_i|^p\right)^{1/p} \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{1/p}$$

$\|\cdot\|_p : \mathbb{R}^n \to \mathbb{R}$, given by $\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$ is convex function.

$$\|\vec{x}+\vec{y}\|_p \le \|\vec{x}\|_p + \|\vec{y}\|_p$$

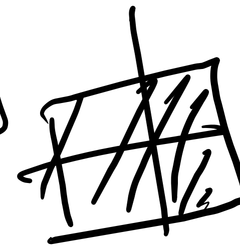
(Minkowski's inequality)

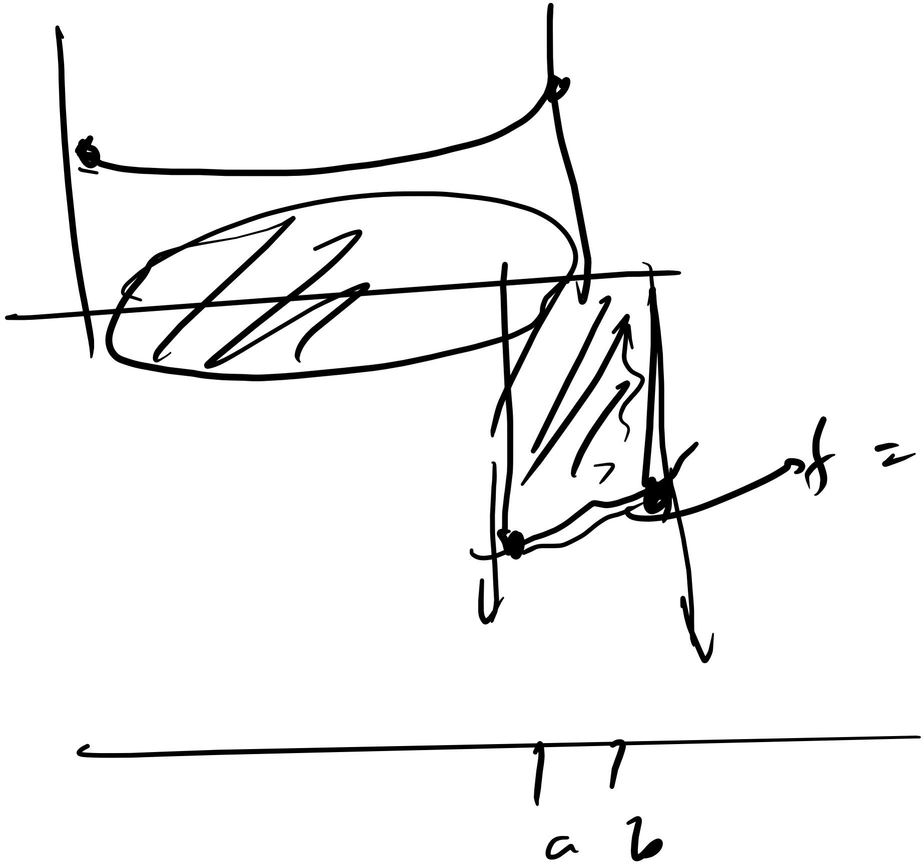
$$\|\lambda\vec{x}\|_p = |\lambda|\,\|\vec{x}\|_p$$

$$\|\vec{x}\|_1 = |x_1| + \dots + |x_n|$$

$$\|\vec{x}\|_2 = \left(|x_1|^2 + \dots + |x_n|^2\right)^{1/2}.$$

$$\|\vec{x}\|_\infty = \boxed{\max\{|x_1|, \dots, |x_n|\}}.$$





$f = \sup\{ax+b\}$

s.t. $ax+b \le f$