Topics in Applied Optimization

Optimization Algorithms for ML and Data Sciences

Pawan Kumar

CSTAR, IIIT-H

Remarks on First Order Condition

- The Taylor expansion of infinitely differentiable function $f(x)$ at a point $x = a$

Remarks on First Order Condition

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

- The Taylor expansion of infinitely differentiable function $f(x)$ at a point $x = a$

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

$$= f(a) + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \dots$$

Remarks on First Order Condition

- The Taylor expansion of infinitely differentiable function $f(x)$ at a point $x = a$

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$
$$= f(a) + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \dots$$

- Exercise: Expand the following functions using Taylor series
    - $f : \mathbb{R} \to \mathbb{R}$, Expand $f(x) = e^x$ at $x = 0$

Remarks on First Order Condition

- The Taylor expansion of infinitely differentiable function $f(x)$ at a point $x = a$

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

$$= f(a) + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \dots$$

- Exercise: Expand the following functions using Taylor series
  - $f : \mathbb{R} \to \mathbb{R}$, Expand $f(x) = e^x$ at $x = 0$
  - $f : \mathbb{R} \to \mathbb{R}$, Expand $f(x) = \sin x$ at $x = \pi$

Remarks on First Order Condition

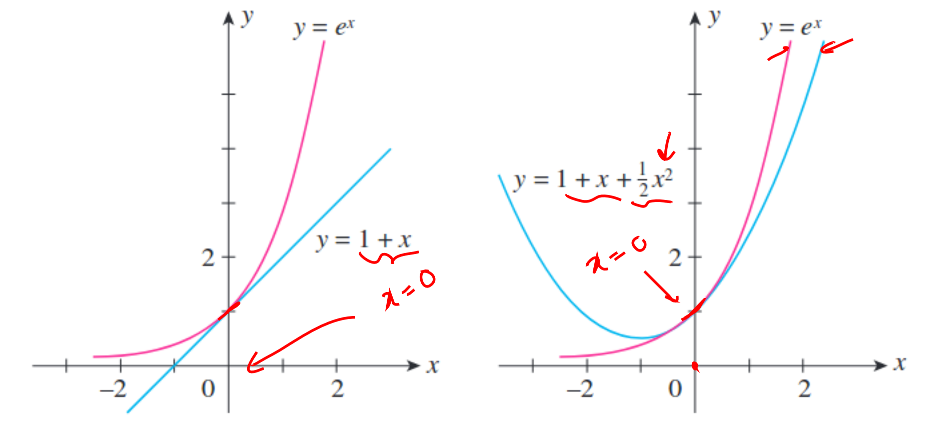
- The Taylor expansion of infinitely differentiable function $f(x)$ at a point $x = a$

$$
\begin{aligned}
f(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \\
&= f(a) + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \dots
\end{aligned}
$$

- Exercise: Expand the following functions using Taylor series
  - $f : \mathbb{R} \to \mathbb{R}$, Expand $f(x) = e^x$ at $x = 0$
  - $f : \mathbb{R} \to \mathbb{R}$, Expand $f(x) = \sin x$ at $x = \pi$
  - $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$, Expand $f(x) = \dfrac{1}{1-x}$ at $x = -1$

$\boxed{f : \mathbb{R} \to \mathbb{R}, \quad \text{Expand } f(x) = e^x \text{ at } x = 0}$

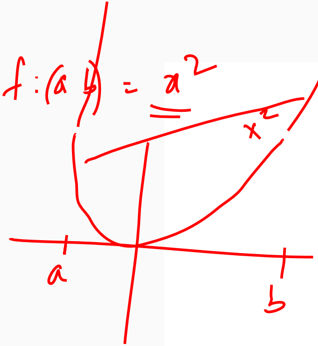
$f : \mathbb{R} \to \mathbb{R}$,   Expand $f(x) = e^x$ at $x = 0$

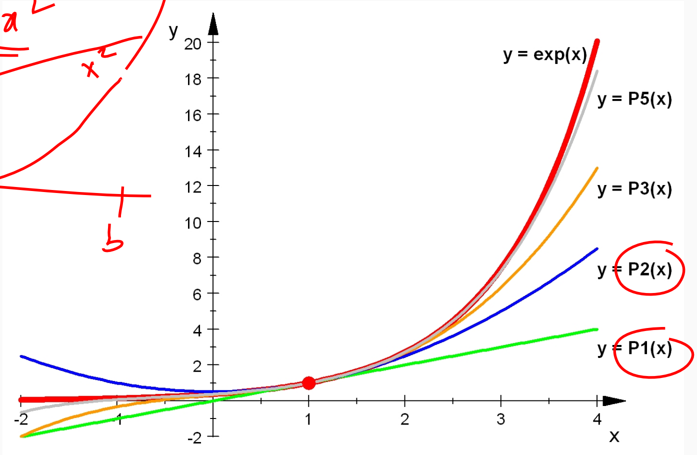


$$\boxed{f : \mathbb{R} \to \mathbb{R}, \quad \text{Expand } f(x) = e^x \text{ at } x = 1}$$

$f : \mathbb{R} \to \mathbb{R}, \quad$ Expand $f(x) = e^x$ at $x = 1$

$f : (a, b) = a^2$





$\boxed{f : \mathbb{R} \to \mathbb{R}, \quad \text{Expand } f(x) = \sin x \text{ at } x = 0}$

$f : \mathbb{R} \to \mathbb{R}$, Expand $f(x) = \sin x$ at $x = 0$

ep f

a

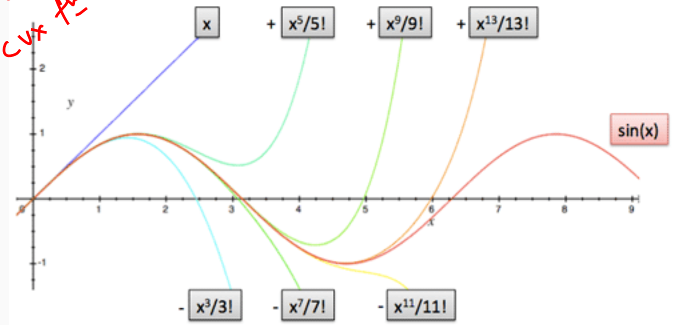
is cvx set

ep f is cvx set

$\iff$ f is cvx fn

c

Better Models of Sine



Multivariate Taylor Series

2nd order

The Taylor expansion of function $f : \mathbb{R}^2 \to \mathbb{R}$ near point (a, b) is given as follows:

$$f(x, y) \approx f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(x - b) + \frac{1}{2}(A(x - a)^2 + 2B\underbrace{(x - a)(y - b)}_{2} + C\underbrace{(y - b)^2}_{2}),$$

where

$$A = \frac{\partial^2 f}{\partial x^2}(a, b), \quad C = \frac{\partial^2 f}{\partial y^2}(a, b)$$

$$B = \frac{\partial^2 f}{\partial x \partial y}(a, b) = \frac{\partial^2 f}{\partial y \partial x}(a, b)$$

$$f(a,b) + \frac{1}{1!}\begin{bmatrix} \frac{\partial f}{\partial x} , & \frac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} x-a \\ y-b \end{bmatrix} + \frac{1}{2!}\begin{bmatrix} x-a , & y-b \end{bmatrix} \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \begin{bmatrix} x-a \\ y-b \end{bmatrix} + \cdots$$

Derivation of Multivariate Taylor Series...

From implicit fn th, $g(t) = \underbrace{f(x(t), y(t))}$ $\rightarrow$ ① (a, b) $\quad \frac{df}{dg} \cdot \frac{dg}{dt}$ chain

Need $(x, y) \sim (a, b)$

$\rightarrow x(t) = a + t(x - a)$

$\rightarrow y(t) = b + t(y - b)$

Then $x'(t) = x - a$, $\quad y'(t) = y - b$

$x''(t) = 0$, $\quad y''(t) = 0$

Using Chain rule for ①,

$$g'(t) = \frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt}$$

$x(1) = x \qquad y(1) = y$

$$f(u(v(w(t))))$$

$$\frac{df}{du} \cdot \frac{du}{dv} \cdot \frac{dv}{dw} \cdot \frac{dw}{dt}$$

$y = f(g(t))$

$$\frac{dy}{dt} = \frac{df}{dg}(g(t)) \cdot g'(t)$$

$$v(t) = f(x_1(t), x_2(t), \ldots x_n(t))$$

$$v'(t) = \frac{\partial f}{\partial x_1} \cdot \frac{dx_1}{dt} + \frac{\partial f}{\partial x_2} \cdot \frac{dx_2}{dt} + \cdots \frac{\partial f}{\partial x_n} \cdot \frac{dx_n}{dt}.$$

Derivation of Multivariate Taylor Series...

$f(x) = f(a) + f'(a)(x-a) + f''(a)(x-a)^2$

Product rule

$g'(t) = \frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt}$

$f(t) = u(t)\, v(t)$

$\frac{df}{dt} = u(t)\, v'(t) + v(t) \cdot u'(t)$

$g''(t) = \frac{\partial^2 f}{\partial x^2} (x'(t))^2 + 2 \frac{\partial^2 f}{\partial x \partial y} x'(t) y'(t) + \frac{\partial^2 f}{\partial y^2} (y'(t))^2$

A, B, C

Put $t = 0$, $x(0) = a$, $y(0) = b$

$g'(0) = \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(y-b)$

$g''(0) = A(x-a)^2 + 2B(x-a)(y-b) + C(y-b)^2$

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

$g(1) = g(0) + g'(0) + \frac{1}{2} g''(0)$ ← assum $\mathbb{R}$ result

$f(x) = f(a) + f'(a)(x-a) + f''(a)$

$x = 1$, $a = 0$

$f(x,y) = f(a,b) + \nabla f(a,b)\binom{a}{b} + \frac{1}{2}\nabla^2 f$

Derivation of Multivariate Taylor Series...

Multivariate Taylor Series Using Gradient and Hessian...

The Taylor expansion of function $f : \mathbb{R} \to \mathbb{R}$ near point (a, b) is given as follows:

$$f(x, y) = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial x}(a, b)(x - b) + \frac{1}{2}(A(x - a)^2 + 2B(x - a)(y - b) + C(y - b)^2),$$

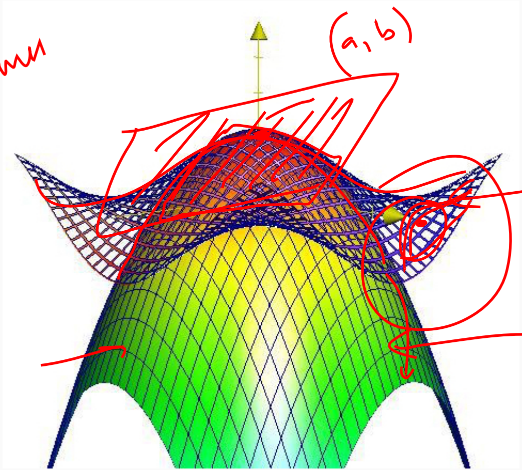
where

$$A = \frac{\partial^2 f}{\partial x^2}(a, b), \quad C = \frac{\partial^2 f}{\partial y^2}(a, b)$$

$$B = \frac{\partial^2 f}{\partial x \partial y}(a, b) = \frac{\partial^2 f}{\partial y \partial x}(a, b)$$

Taylor approximation

f(x) linear



(a, b)

original fn

Figure 1: A second order Taylor approximation of the function

Remarks on First Order Condition

- The affine function of y given by $f(x) + \nabla f(x)^T(y - x)$ is a first order Taylor approximation of f near x.

Remarks on First Order Condition

- The affine function of y given by $f(x) + \nabla f(x)^T(y - x)$ is a first order Taylor approximation of f near x.
- The inequality

$$f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

  states that for a convex function the first order Taylor approximation is in fact a global underestimator of the function

Remarks on First Order Condition

$D = \nabla^2$

1st convexity $+$ $\nabla f(x^*) = 0 \Rightarrow x^*$ is global minima.

- The affine function of y given by $f(x) + \nabla f(x)^T (y-x)$ is a first order Taylor approximation of f near x.

$f(y) \geq f(x^*)$

- The inequality

$\boxed{\Delta u = f}$ Laplace eqn

$$(1) \quad f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

0 when $\nabla f(x) = 0$

$D = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$

states that for a convex function the first order Taylor approximation is in fact a global underestimator of the function

- Conversly, if the first order Taylor approximation of a function is always a global underestimator of the function, then the function is convex
- Quiz: What does the inequality above say when $\nabla f(x) = 0$?

$f(y) \geq f(x)$

At x, when $\nabla f(x) = 0$

$(1) \Rightarrow f(y) \geq f(x) \quad \forall y \in \operatorname{dom} f$

$\forall y \Rightarrow x$ is the min val.

Remarks on First Order Condition

- The affine function of y given by $f(x) + \nabla f(x)^T(y - x)$ is a first order Taylor approximation of f near x.
- The inequality

$$f(y) \underset{>}{\geq} f(x) + \nabla f(x)^T(y - x)$$

  states that for a convex function the first order Taylor approximation is in fact a global underestimator of the function
- Conversly, if the first order Taylor approximation of a function is always a global underestimator of the function, then the function is convex
- Quiz: What does the inequality above say when $\nabla f(x) = 0$?
- A function is strictly convex if the inequality is strict

Remarks on First Order Condition

- The affine function of y given by $f(x) + \nabla f(x)^T(y - x)$ is a first order Taylor approximation of f near x.
- The inequality

$$f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

  states that for a convex function the first order Taylor approximation is in fact a global underestimator of the function
- Conversly, if the first order Taylor approximation of a function is always a global underestimator of the function, then the function is convex
- Quiz: What does the inequality above say when $\nabla f(x) = 0$?
- A function is strictly convex if the inequality is strict
- To define concave functions, the inequality is reversed

Proof of first order convexity

real valued

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

First Order Condition: Suppose f is differentiable. Then f is convex **if and only if dom** f is convex and

multivariate

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

holds for all $x, y \in$ **dom** f

Proof on Chalkboard!

Consider $n = 1$ $\qquad$ $f : \mathbb{R} \to \mathbb{R}$

1st order cond : f is cvx if & only if

$$f(y) \geq f(x) + f'(x)(y - x) \quad (1)$$

$\forall x, y \in \mathrm{dom}\, f$

Proof of first order convexity...

Necessary: Assume f is cvx, $x, y \in \operatorname{dom} f$. Since $\operatorname{dom} f$ is cvx we conclude that $(0 < t \leq 1)$, $x + t(y-x) \in \operatorname{dom} f$

$= ty + (1-t)x$ cvx

By cvxty of f

$$f(x + t(y-x)) \leq (1-t) f(x) + t f(y)$$

Dividing both sides by t:

$$f(y) \geq f(x) + \frac{f(x + t(y-x)) - f(x)}{t}$$

0/0 form

Let $t \to 0$

$$f(y) \geq f(x) + \lim_{t \to 0} \frac{f'(x + t(y-x))(y-x)}{1}$$

L'Hopital rule

$$\lim_{x \to 0} \frac{f(x)}{g(x)}$$

If $\frac{0}{0}$ form, then

$$\lim_{x \to 0} \frac{f'(x)}{g'(x)}$$

$= f'(x)(y-x)$

Proof of first order convexity...

Sufficiency: Assume that f satisfies (1) $\forall x, y \in \text{dom } f$.

Choose $x \neq y \in \text{dom } f$ & $0 \leq \theta \leq 1$, $\boxed{z = \theta x + (1-\theta)y} \in \text{dom } f$.

Claim: $\theta f(x) + (1-\theta) f(y) \geq f(\underbrace{\theta x + (1-\theta)y}_{z})$

(1) $\Rightarrow$

$\theta \times \quad f(x) \geq f(z) + f'(z)(x-z) \quad \forall x, z \in \text{dom } f$

$(1-\theta) \times f(y) \geq f(z) + f'(z)(y-z) \quad \forall y, z \in \text{dom } f$

$+$

$\theta f(x) + (1-\theta) f(y) \geq$ want $f(z)$

$\cancel{\theta f(z)} + \theta f'(z)(x-z)$

$+ \cancel{(1-\theta) f(z)} + (1-\theta) f'(z)(y-z)$

$f(z) + f'(z)\left[\cancel{\theta(x-z) + (1-\theta)(y-z)}\right]$

0

Second Order Conditions

Second Order Condition: Assume that f is twice differentiable, that is the Hessian or second derivative $\nabla^2 f$ exists at each point in **dom** f,which is open. Then f is convex if and only if **dom** f is convex and its Hessian is positive semidefinite, i.e., for all $x \in$ **dom** f,

$$\nabla^2 f(x) \succeq 0$$

+ve definite

$P = [x_i x_j]$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{n1} & \cdots & & a_{nn} \end{bmatrix}$$

For a function on $\mathbb{R}$, this reduces to

$$f''(x) \geq 0$$

The condition $\nabla^2 f(x) \succeq 0$ can be interpreted geometrically as the requirement that the graph of the function have positive curvature at x.

min $f(x)$

Example: Consider the quadratic function $f : \mathbb{R}^n \to \mathbb{R}$, with **dom** $f = \mathbb{R}^n$,

$$f(x) = (1/2)x^T Px + q^T x + r,$$

$\nabla f(x) = \cancel{2} \cdot \frac{1}{\cancel{2}} Px + q$

$Px + q$

$\boxed{\nabla^2 f(x) = P}$

with $P \in S^n$, $q \in \mathbb{R}^n$, $r \in \mathbb{R}$. **Quiz:** What is $\nabla^2 f(x)$?

symmetric

Hessian

Why need the domain of f to be convex?

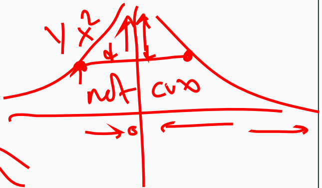
$$\boxed{\frac{d}{dx}(x^n) = n x^{n-1}}$$

Fact: The separate requirement that **dom** f be convex cannot be dropped from the first or second order charecterizations of convexity and concavity.

For example, the function

x^{-2} not def at $x = 0$

$$f(x) = 1/x^2, \quad \textbf{dom}\ f = \{x \in \mathbb{R} \mid x \neq 0\}$$



satisfies $f''(x) > 0$ for all $x \in \textbf{dom}\ f$, but it is not a convex function

$$f'(x) = (-2)\,x^{-3}$$

$$f''(x) = \underbrace{(-2)(-3)}_{>0}\,x^{-4} = \frac{6}{x^4} > 0 \quad \forall x \in \text{dom } f$$

Examples of Convex Functions

- Exponential. e^{ax} is convex on $\mathbb{R}$, for any $a \in \mathbb{R}$
  - Second derivative test

$$f(x) = e^{ax} \rightarrow 1 \rightarrow$$

$$f'(x) = a e^{ax} \rightarrow$$

$$f''(x) = \underbrace{a^2 e^{ax}}_{>0}$$

$$\geq 0 \quad > 0$$

$a > 0 \Rightarrow$ strictly cvx

Examples of Convex Functions

- Exponential. e^{ax} is convex on $\mathbb{R}$, for any $a \in \mathbb{R}$
  - Second derivative test
- Powers. x^a is convex on $\mathbb{R}_{++}$ when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$
  - Second derivative test

$$f(x) = x^a$$

$$y = x^a \qquad \log y = a \log x$$

$$\frac{1}{y}\frac{dy}{dx} = \frac{a}{x}$$

$$\frac{dy}{dx} = y\,\frac{a}{x} = \frac{x^a \cdot a}{x}$$

$$\frac{d^2y}{dx^2} = x^a\left(-\frac{a}{x^2}\right) + \left(\frac{a}{x}\right)\frac{d}{dx}\left(x^a\right)$$

$$= -a\,\frac{x^a}{x^2} + \frac{a}{x}\, y \cdot \frac{a}{x}$$

$$= x^a\left[\left(\frac{a}{x}\right)^2 - \frac{a}{x^2}\right] = \frac{x^a}{x^2}\, a(a-1)$$

$x \in R_{++}$, $x^a > 0$, $x^2 > 0$, ≤ 0

$a(a-1) \geq 0$ if $\boxed{\begin{array}{l} a \geq 1 \\ a \leq 0 \end{array}}$

Examples of Convex Functions

- Exponential. e^{ax} is convex on $\mathbb{R}$, for any $a \in \mathbb{R}$
  - Second derivative test
- Powers. x^a is convex on $\mathbb{R}_{++}$ when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$
  - Second derivative test
- Powers of absolute values. $|x|^p$, for $p \geq 1$, is convex on $\mathbb{R}$
  - Same as above with **dom** $|x|$ being $\mathbb{R}_{++}$

Examples of Convex Functions

- Exponential. e^{ax} is convex on $\mathbb{R}$, for any $a \in \mathbb{R}$
  - Second derivative test
- Powers. x^a is convex on $\mathbb{R}_{++}$ when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$
  - Second derivative test
- Powers of absolute values. $|x|^p$, for $p \geq 1$, is convex on $\mathbb{R}$
  - Same as above with **dom** $|x|$ being $\mathbb{R}_{++}$
- Logarithm. $\log x$ is concave on $\mathbb{R}_{++}$
  - Second derivative test

Examples of Convex Functions

- Exponential. e^{ax} is convex on $\mathbb{R}$, for any $a \in \mathbb{R}$
  - Second derivative test
- Powers. x^a is convex on $\mathbb{R}_{++}$ when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$
  - Second derivative test
- Powers of absolute values. $|x|^p$, for $p \geq 1$, is convex on $\mathbb{R}$
  - Same as above with **dom** $|x|$ being $\mathbb{R}_{++}$
- Logarithm. $\log x$ is concave on $\mathbb{R}_{++}$
  - Second derivative test
- Negative Entropy. $x \log x$ on $\mathbb{R}_{++}$ is convex
  - Second derivative test

Norm is a convex function. Hint: Triangle inequality.

Recall: $u, v \in \mathbb{R}^n$

Triangle ineq: $\|u+v\|_p \le \|u\|_p + \|v\|_p$

$f(x) = \|x\|_p$

$\theta f(x) + (1-\theta) f(y) = \theta \|x\|_p + (1-\theta)\|y\|_p$

$f(\theta x + (1-\theta)y) = \|\theta x + (1-\theta) y\|_p$

$\le \|\theta x\|_p + \|(1-\theta)y\|_p$

$= \theta \|x\|_p + (1-\theta)\|y\|_p$

Max Functions. $f(x) = \max\{x_1, x_2, \ldots, x_n\}$ is convex on $\mathbb{R}^n$

$$\|x\|_\infty = \max\{|x_1|, |x_2|, \ldots |x_n|\}$$

become

$\| \| \geq 0$ need $| \;\; |$