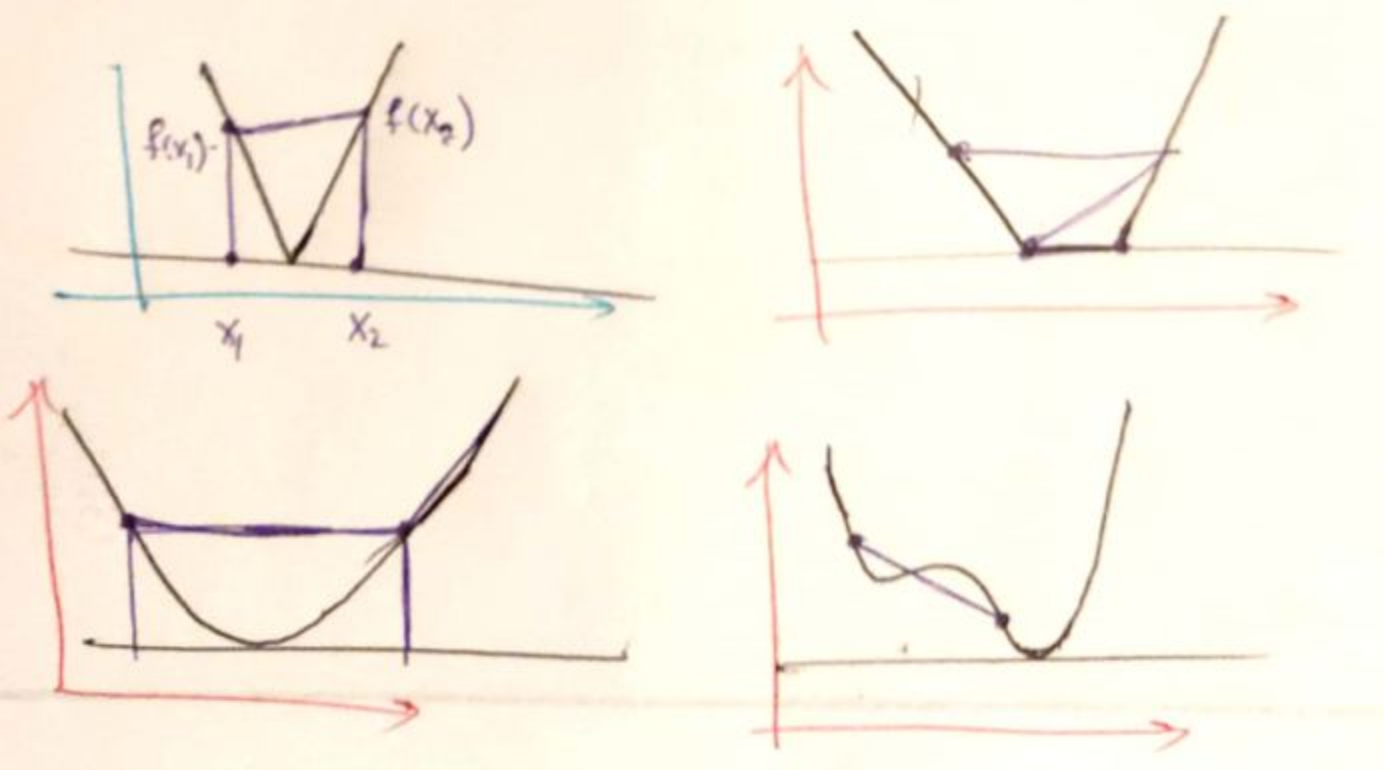
Mathematics for AI

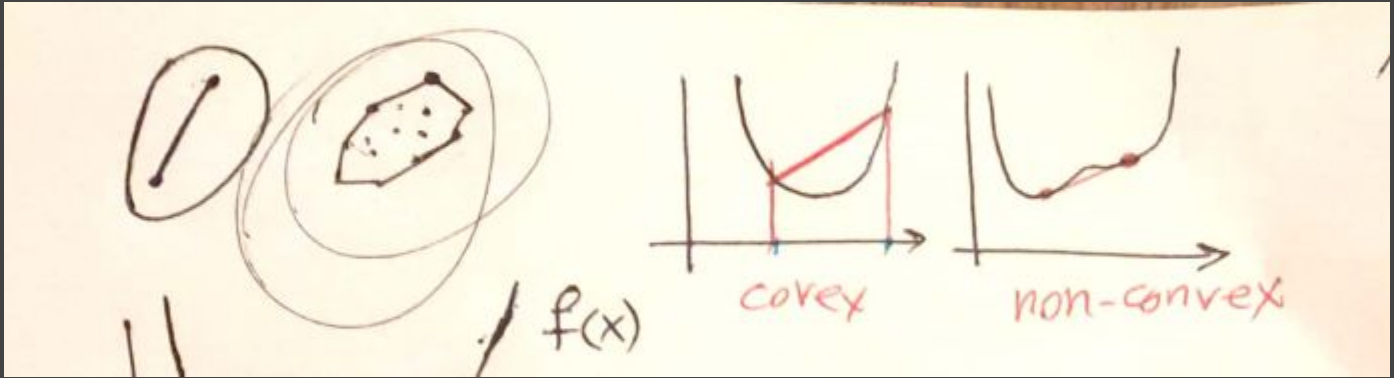
Lecture 31

Convex Functions

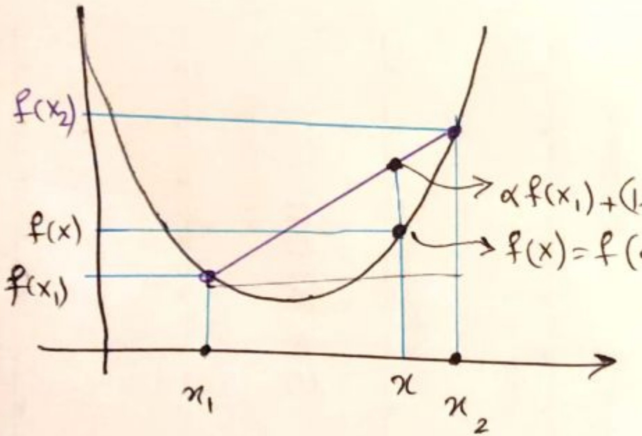
Convex Functions



Convex Functions



Convex Functions



$\rightarrow \alpha f(x_1) + (1-\alpha) f(x_2)$

$\rightarrow f(x) = f(\alpha x_1 + (1-\alpha) x_2)$

$x = \alpha x_1 + (1-\alpha) x_2$

for some $\alpha \in [0,1]$

$x = \alpha x_1 + (1-\alpha) x_2$ for some $\alpha \in [0,1]$

If for any $x_1, x_2 \in \mathbb{R}$

$$f(\alpha x_1 + (1-\alpha) x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$$

then f is for all $\alpha \in [0,1]$

or for all $\alpha \in (0,1)$

Convex Functions

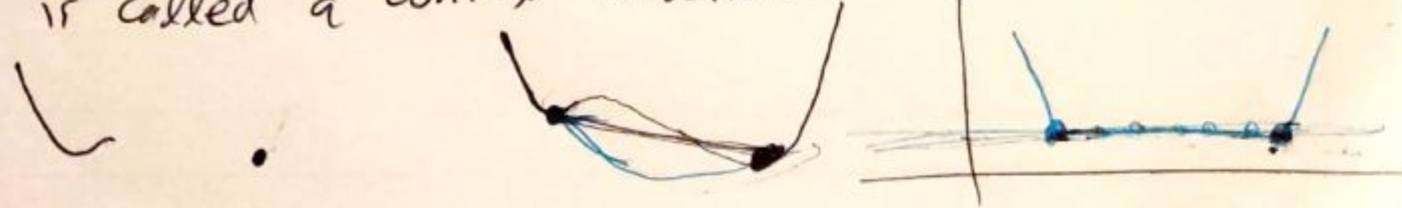
If for any $x_1, x_2 \in \mathbb{R}$

$$f(\alpha x_1 + (1-\alpha) x_2) \leqslant \alpha f(x_1) + (1-\alpha) f(x_2)$$

~~then f~~ is for all $\alpha \in [0,1]$

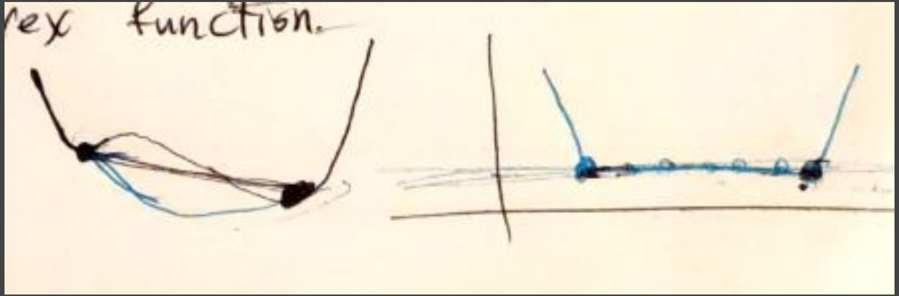
or for all $\alpha \in (0,1)$

$\Longrightarrow$ f is called a convex function.



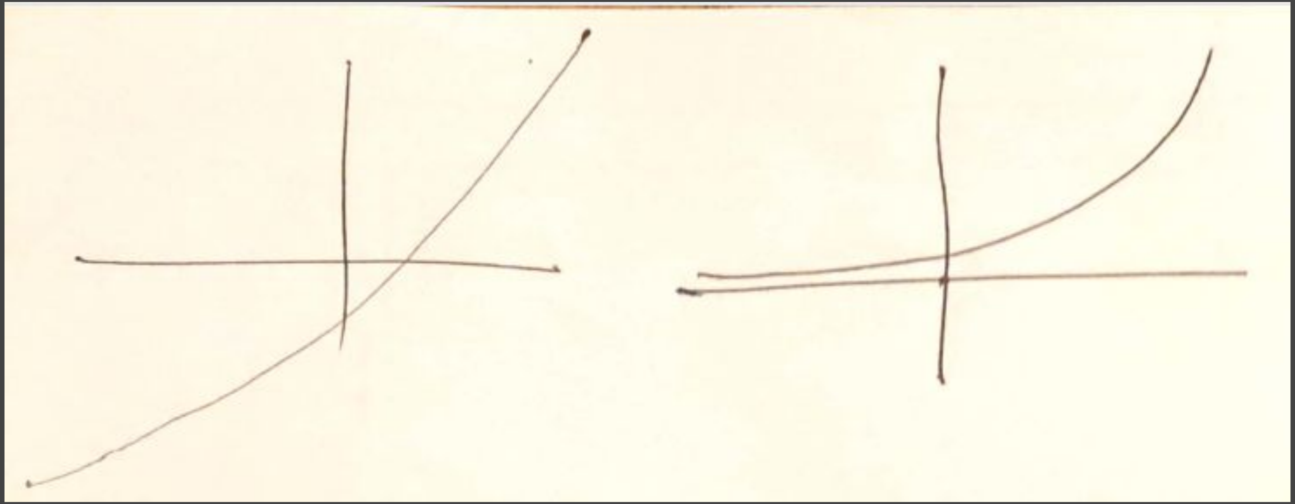
Convex Functions

For convex functions every local minimum is a global minimum.



Convex Functions

Convex functions might not have a minimum, or the minimum (infimum) might not be achieved (there is no argmin).



Convex Functions must have a convex domain

K. N. Toosi
University of Technology

$f: D \longrightarrow \mathbb{R}$ D is a convex set.

Strictly Convex Functions

→ A function $f: \mathbb{R} \to \mathbb{R}$ is called strictly convex if D convex

$$f(\alpha x_1 + (1-\alpha) x_2) < \alpha f(x_1) + (1-\alpha) f(x_2)$$

for all $x_1, x_2 \in \mathbb{R}$ and all $\alpha \in (0,1)$.
D

For strictly convex functions the minimum point is unique (if exists).

Strictly Convex Functions

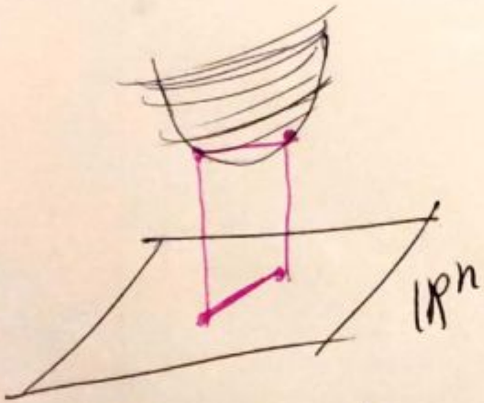
K. N. Toosi
University of Technology

$f: D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ D: convex $\boxed{\beta = 1-\alpha}$

$\mathbb{R}^n \longrightarrow \mathbb{R}^n$

convex $f(\alpha \vec{x}_1 + \beta \vec{x}_2) \leq \alpha f(\vec{x}_1) + \beta f(\vec{x}_2)$

strictly convex $f(\alpha \vec{x}_1 + \beta \vec{x}_2) < \alpha f(x_1) + \beta f(x_2)$



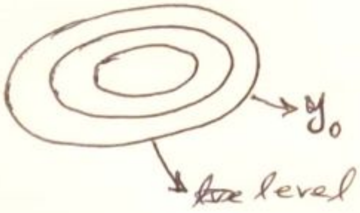
$\alpha \in (0,1)$

$\beta = 1 - \alpha$

f is strictly convex $\Longrightarrow$ $\underset{x \in D}{\operatorname{argmin}} f(x)$ is unique or non-existent

Sublevel sets of a convex function

level set, level curve



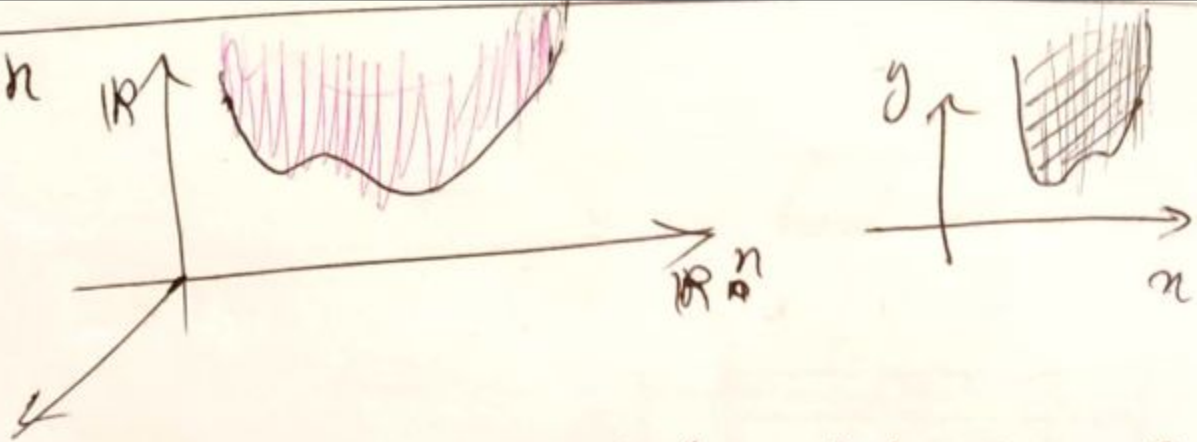
level set $\{x \mid f(x) = y_0\}$

sub-level set $\{x \mid f(x) \leqslant y_0\}$

For convex functions sub-level sets are convex.

Epigraph sets of a convex function

epigraph



$f: D \longrightarrow \mathbb{R}$ $\qquad$ $\text{epigraph}(f) = \{(x, y) \mid x \in D, y \in R, y \geq f(x)\}$

f: convex function $\Longrightarrow$ epigraph(f) is a convex set

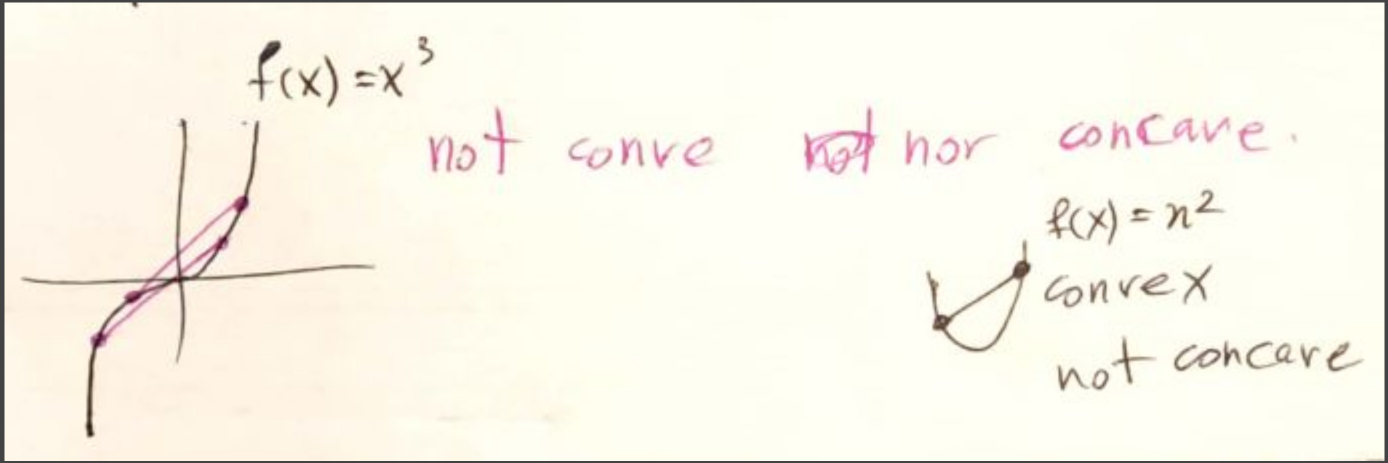
Concave functions

concave function  $f(\alpha x_1+(1-\alpha)x_2) \geq \alpha f(x_1)+(1-\alpha)f(x_2)$

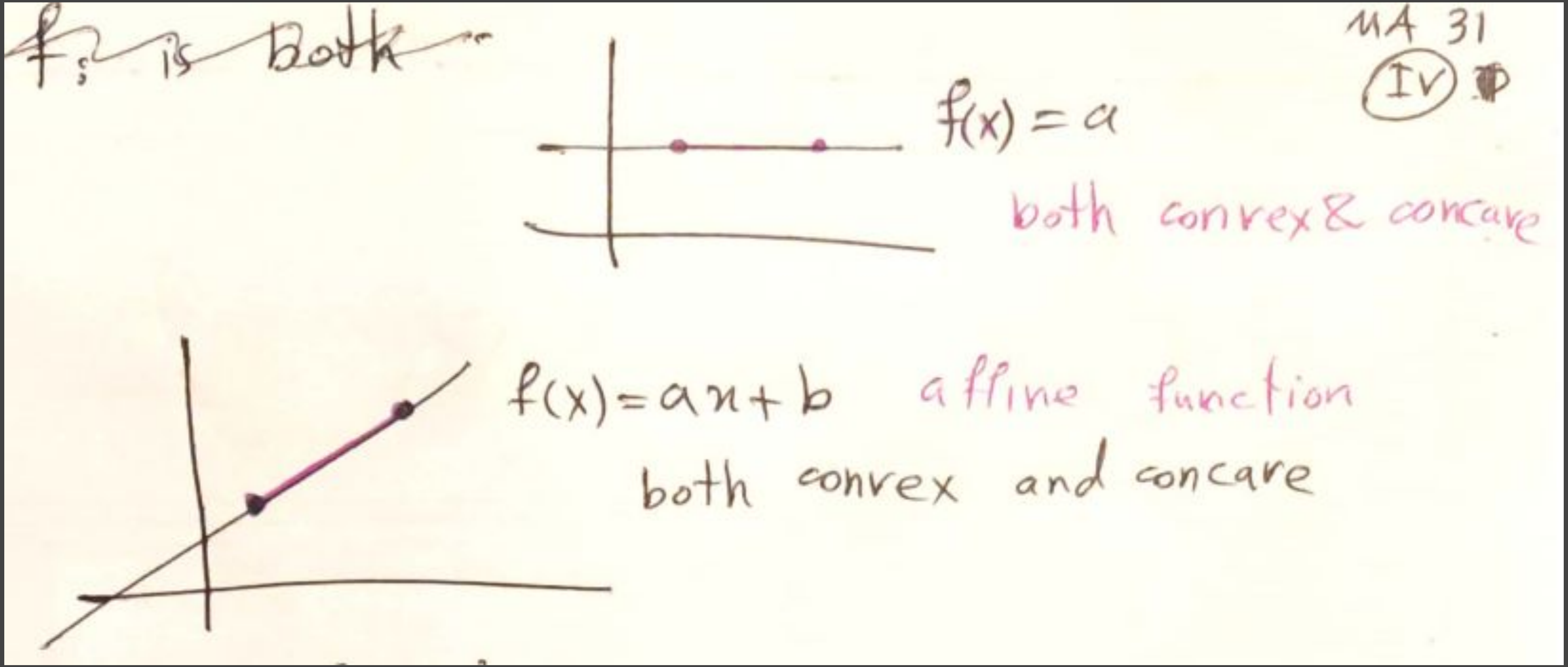
strictly concave.

f: convex $\Longleftrightarrow$ $-f$ concave

Example:



Affine functions are both convex and concave

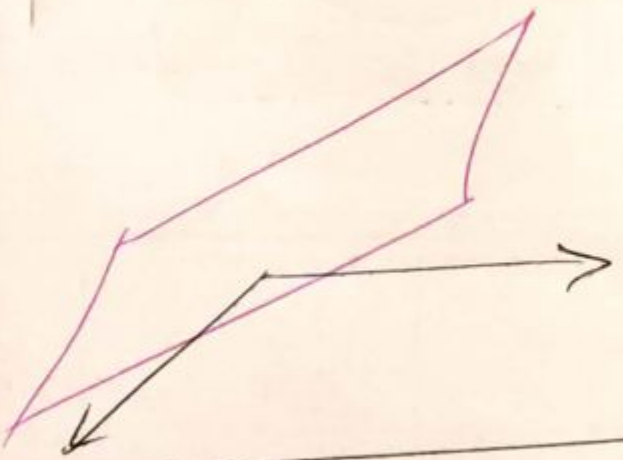


Affine functions are both convex and concave

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

$f(x) = a^T x + b$ affine

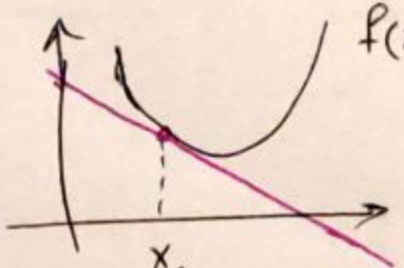
both convex & concave.

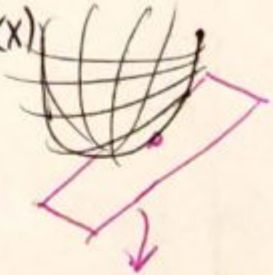


Gradient and Convexity

$f: \mathbb{R}^n \longrightarrow \mathbb{R}$ convex & differentiable at x_0.

$$f(x_0) + \nabla_f(x_0)^T(x - x_0) \leq f(x)$$



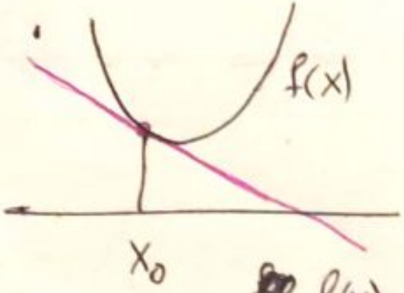


$$l(x) = f(x_0) + (x - x_0) f'(x_0) \leq f(x)$$

$$f(x_0) + \nabla_f(x_0)^T(x - x_0)$$

Gradient and Convexity (first order convexity condition)

K. N. Toosi
University of Technology



f is convex & differentiable

MA31

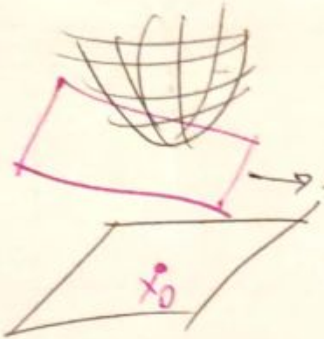
∇

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

$\exists\, m \in \mathbb{R}^n$ s.t. $f(x) \geq f(x_0) + m^T(x - x_0)$ for all $x \in \mathbb{R}^n$

$\ell(x) = f(x_0) + m(x - x_0)$

$\Longrightarrow m = \nabla_f (x_0)$



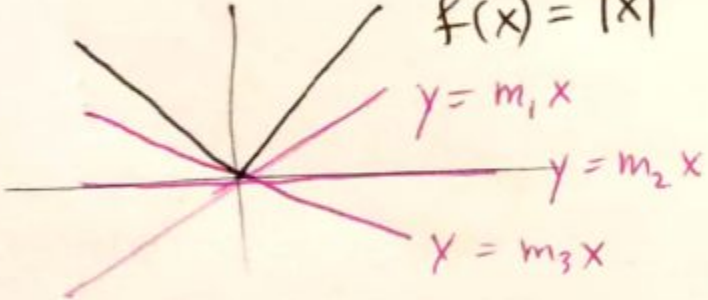
$\ell(x) = f(x_0) + m^T(x - x_0)$

$$\boxed{f(x) - f(x_0) \geq m^T(x - x_0)}$$

What about non-differentiable convex functions?

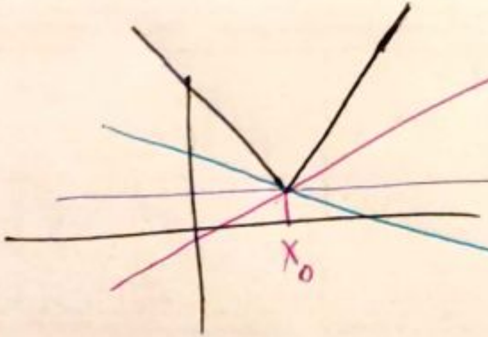
What if $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ is convex, but not differentiable at x_0

$f(x) = |x|$ not differentiable at 0.



Subgradient

K. N. Toosi
University of Technology



$\ell_1(x) = f(x_0) + m_1(x - x_0)$

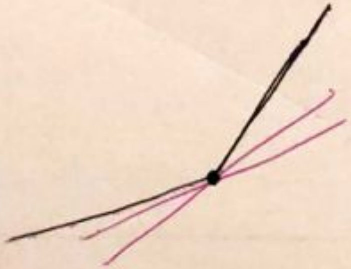
$\ell_3(x) = f(x_0) + 0\,(x - x_0)$

$\ell_2(x) = f(x_0) + m_2(x - x_0)$

$m_1, m_2, 0$ are subgradients of f.

---

f convex, $f : \mathbb{R}^n \longrightarrow \mathbb{R}$



subgrad f at x_0

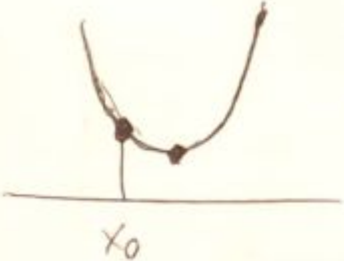
$$= \{\, m \in \mathbb{R}^n \mid f(x_0) + m^T(x - x_0) \leqslant f(x) \quad \forall x \,\}$$

1D convex functions and second derivative

$f:\mathbb{R} \to \mathbb{R}$ convex & twice differentiable at x_0.

$\Longrightarrow$ $f''(x_0)$ exists

$f''(x_0) \geq 0$



Hessian and convexity (second order convexity condition)

$f: \mathbb{R}^n \longrightarrow \mathbb{R}$ convex & twice differentiable

at $x_0 \Longrightarrow$ Hessian $H_f(x_0)$ exists.

$\Longrightarrow$ curvature is positive in all directions

$$\forall u \in \mathbb{R}^n \quad u^T H_f(x_0)\, u \geqslant 0$$

$\forall u \quad u^T H u \geqslant 0 \Longrightarrow H$ is positive semi-definite

Hessian and convexity (second order convexity condition)

K. N. Toosi
University of Technology

$f: D \to \mathbb{R}$ Hessian exists for all $x \in D$

and is PSD $\Rightarrow$ f convex

is PD $\Rightarrow$ f strictly convex

Hessian and convexity (second order convexity condition)

How to check if a function is convex. MA 371 VII

$\Longrightarrow$ compute Hessian

$f(x) = \frac{1}{2}x^T A x + b^T x + c$ quadratic

$H_f(x) = A$

A positive definite $\Longrightarrow f$ convex

A negative " $\Longrightarrow f$ concave

Convex functions properties

f convex $\Rightarrow -f$ concave

Convex functions properties

$f_1, f_2, \dots, f_n$ convex

$\sum_i \alpha_i f_i(x)$ convex $\quad \alpha_i \in \mathbb{R}$

provided that $\alpha_i \geqslant 0$

Convex functions properties

K. N. Toosi
University of Technology

when is $f \circ g(x)$ convex? $g: \mathbb{R}^n \to \mathbb{R}$

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

$\beta = 1-\alpha$

$\cancel{f(g(x))}$ $f(g(\alpha x_1 + \beta x_2)) \leq \alpha f(g(x_1)) + \beta f(g(x_2))$

g convex $\quad g(\alpha x_1 + \beta x_2) \leq \alpha g(x_1) + \beta g(x_2)$

f: ~~decreasing~~ non-decreasing

$$f(g(\alpha x_1 + \beta x_2)) \leq f(\alpha g(x_1) + \beta g(x_2)).$$

f: convex

$$f(\alpha g(x_1) + \beta (g(x_2)) \leq \alpha f(g(x_1)) + \beta f(g(x_2))$$