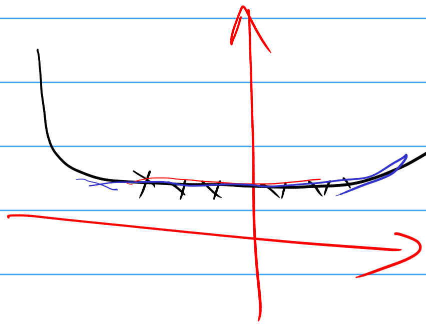
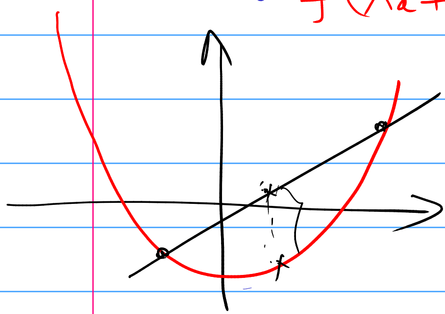
* Convex functions → Every local minima is also a global minima.



* "Strictly" Convex Functions:

✓ $f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$ [Convex functions]

✓ $f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$, $\forall \lambda \in (0,1)$, $x \neq y$ [strictly convex]



$\Downarrow$

if the minimizer exists, it is unique

* "Sufficient" cond^n for strict convexity.

Assume f is twice differentiable and $\nabla^2 f > 0$, then f is strictly convex

eg: $f(x) = x^4 \rightarrow$ strictly convex

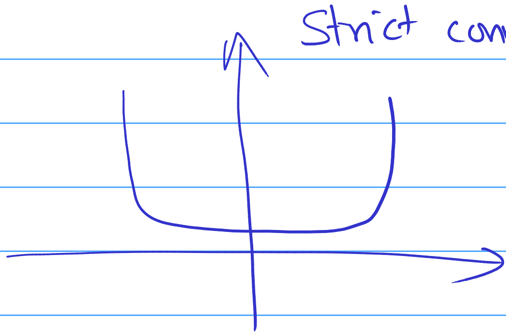
$\nabla^2 f = 12x^2 \geq 0$

$f(x) = x^2$

$\nabla^2 f = 2 > 0$

Strict convexity $\Rightarrow$ Convexity

$\not\Leftarrow$



$\nabla^2 f(x) > 0 \; \forall x \Rightarrow$ Strict convexity

$\not\Leftarrow$

* Strongly Convex functions:

$$f(\lambda x+(1-\lambda)y) \le \lambda f(x)+(1-\lambda) f(y) \quad \text{[Convex]}$$

$$f(\lambda x+(1-\lambda)y) < \lambda f(x)+(1-\lambda) f(y) \quad \text{[Strictly Convex]}$$

$$f(\lambda x+(1-\lambda)y) \le \lambda f(x)+(1-\lambda)f(y) - \frac{\lambda(1-\lambda)\mu}{2}\|x-y\|^2$$

eg: x^2

$\mu>0$

↳ Modulus of Strong Convexity

Strong convexity ⇒ Strict convexity ⇒ Convexity

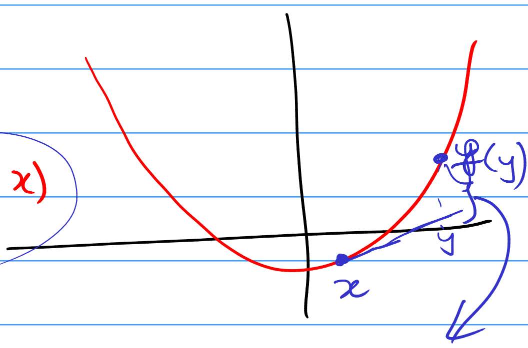
* Assume f is twice continuously differentiable, then

f is μ-Strongly convex (SC) $\iff$ $\nabla^2 f \succeq \mu I$ $\succ 0$

$f(x)=x^2$, ⇒ $f(x)$ is SC with $\mu=2$

1st order cond^n for convexity:

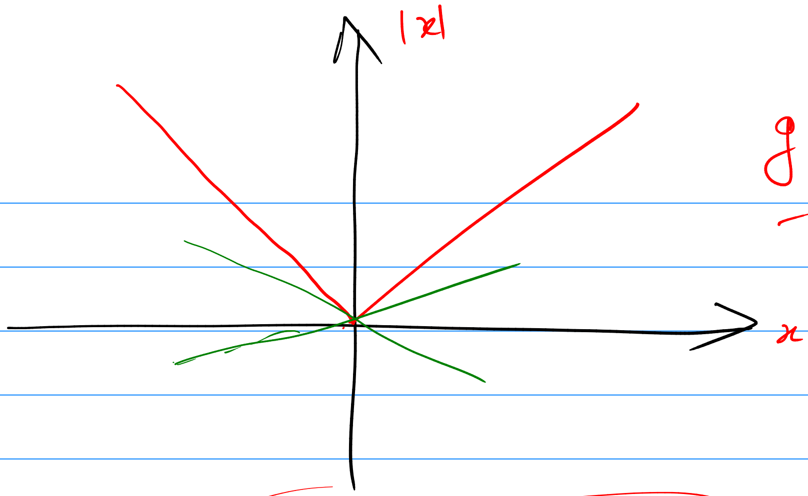
$$f(y) \ge f(x) + \nabla f(x)^T (y-x)$$



1st order condition for SC:

$$f(y) \ge f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|x-y\|^2$$

Remark: When f is not differentiable, we use "subgradients."



$$g = \begin{cases} +1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ [-1, 1] & \text{if } x = 0 \end{cases}$$

$x = 0$

$$f(y) \geq f(x) + \underline{g}^T (y - x)$$

$$f(y) \geq g^T y$$

Strongly convex functions satisfy PL-inequality

Polyak - Lojasiewicz inequality
(Gradient dominance condition)

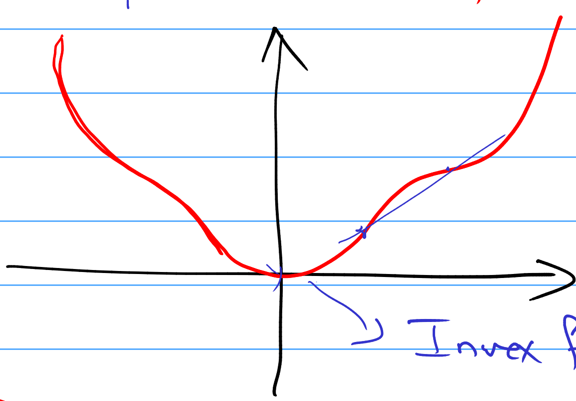
<u>PL-inequality</u>

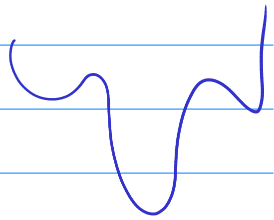
$$\frac{1}{2} \|\nabla f(x)\|^2 \geq \mu (f(x) - f_*)$$

f_* → optimal value of the function

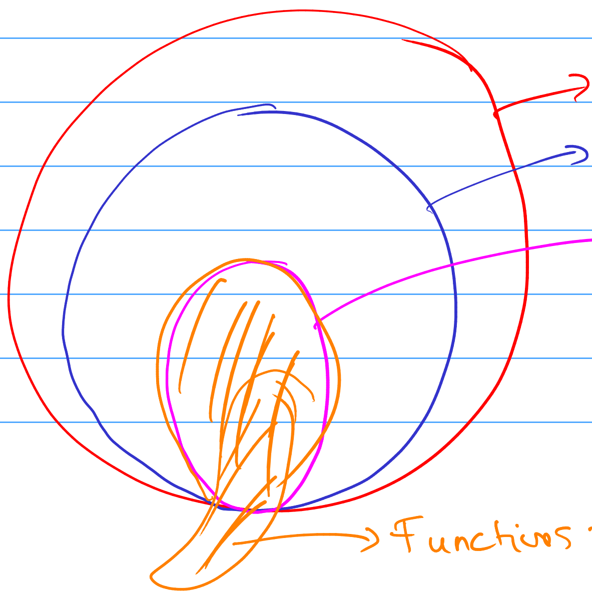
Ex: <u>$f(x) = x^2$ is SC</u>

$g(x) = x^2 + 3\sin^2 x$





Invex functions



Eg: $\frac{1}{2}\|x\|^2 \to$ is SC.

$\frac{1}{2}\|Ax-b\|^2$ is not SC. always.

$\hookrightarrow$ when A is full-row-rank, then this function is SC.

* Implications of SC:

Proposition: Let f be strongly convex with $\mu > 0$, then the following are equivalent:

(i) $f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|x-y\|^2$

(ii) $g(x) := f(x) - \frac{\mu}{2}\|x\|^2$ is convex

(iii) $(\nabla f(x) - \nabla f(y))^T (x-y) \geq \mu \|x-y\|^2 \to \|\nabla f(x) - \nabla f(y)\| \geq \mu \|x-y\|$

(iv) $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{\lambda(1-\lambda)}{2}\|x-y\|^2$

Proof: (i) $\Leftrightarrow$ (ii) $\Rightarrow$ (iii)

$\Downarrow$

(iv)

(i) $\Leftrightarrow$ (ii) $\qquad \nabla g(x) = \nabla f(x) - \mu x$

g is convex $\Leftrightarrow$

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

$$f(y) - \frac{\mu}{2}\|y\|^2 \geq f(x) - \frac{\mu}{2}\|x\|^2 + (\nabla f(x) - \mu x)^T (y-x)$$

$$f(y) \geq f(x) - \frac{\mu}{2}\|x\|^2 + \frac{\mu}{2}\|y\|^2 + \nabla f(x)^T (y-x) + \mu\|x\|^2 - \mu x^T y$$

$$= f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|x-y\|^2$$

(ii) $\Rightarrow$ (iii)

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

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

$$0 \geq (\nabla g(x) - \nabla g(y))^T (y-x)$$

$$(\nabla g(x) - \nabla g(y))^T (x-y) \geq 0$$

(ii) $\Rightarrow$ (iv): - g is convex

$$g(\lambda x+(1-\lambda)y) \le \lambda g(x) + (1-\lambda) g(y)$$

$$f(\lambda x+(1-\lambda)y) - \frac{\mu}{2}\|\lambda x+(1-\lambda)y\|^2 \le \lambda f(x) - \lambda\frac{\mu}{2}\|x\|^2 + (1-\lambda) f(y) - (1-\lambda)\frac{\mu}{2}\|y\|^2$$

$$f(\lambda x+(1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y) - \frac{\lambda(1-\lambda)\mu}{2}\|x-y\|^2$$

$\blacksquare$

* Lipschitz continuous functions:

f is L-Lipschitz

$$|f(y)-f(x)| \le L\|x-y\|$$

$$\|\nabla f(x)\| \le L$$

$\rightarrow$ L-smooth

Lipschitz gradient

$$\mu\|x-y\| \le \|\nabla f(x) - \nabla f(y)\| \le L\|x-y\|$$

$L \ge \mu$

* Let f be μ-SC, g be convex $\}$ $h := f+g$ $\hookrightarrow$ h is μ-SC

Proof:

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

$$\le \lambda f(x) + (1-\lambda) f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2 + \lambda g(x) + (1-\lambda) g(y)$$

$$= \lambda h(x) + (1-\lambda) h(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2$$

$\blacksquare$

* Strong Convexity $\Rightarrow$ PL-inequality.

Propn: f is μ-SC $\Rightarrow$ f satisfies.

$$\frac{1}{2}\|\nabla f(x)\|^2 \ge \mu(f(x) - f_*), \text{ where } f_* := f(x_*)$$

Proof: Let us fix y.

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|x-y\|^2 \quad \forall x.$$

$$f(y) \geq \min_y \left\{ f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|x-y\|^2 \right\} \quad \forall x$$

$$\nabla f(x) + \mu(y^* - x) = 0$$

$$y - x = -\frac{1}{\mu}\nabla f(x)$$

$$f(y) \geq f(x) + \nabla f(x)^T \left(-\frac{1}{\mu}\nabla f(x)\right) + \frac{\mu}{2}\frac{1}{\mu^2}\|\nabla f(x)\|^2$$

$$f(y) \geq f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2$$

$$\min_y f(y) \geq f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2$$

PL inequality

$$f_* \geq f(x) - \frac{1}{2\mu}\|\nabla f(x)\|^2 \Rightarrow \boxed{\frac{1}{2}\|\nabla f(x)\|^2 \geq \mu(f(x) - f_*)}$$

$x^2 + 3\sin^2 x \longrightarrow$ PL-inequality

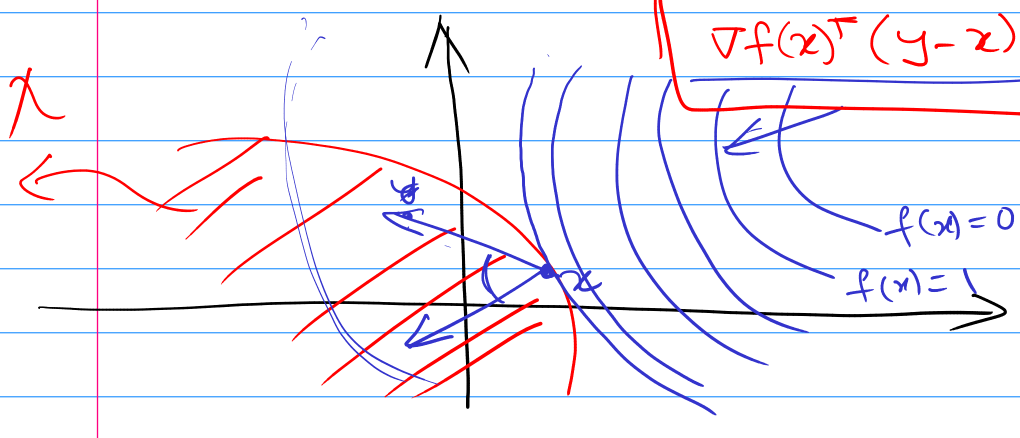
Optimization Problem:

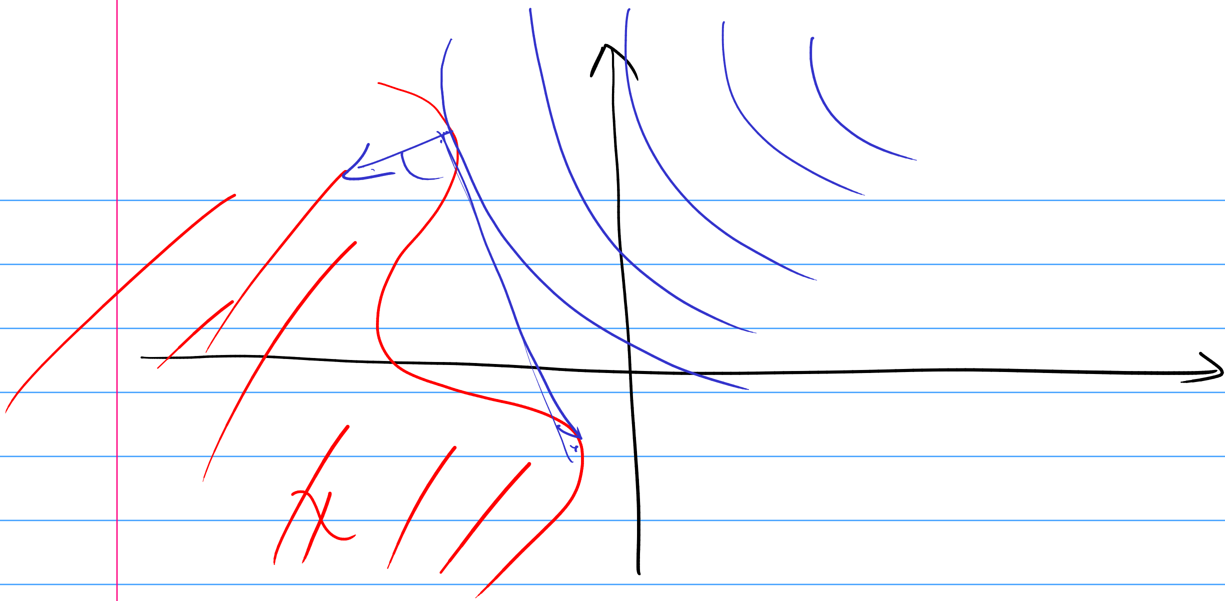
$$\min_{x \in \mathcal{X}} f(x) \quad , f \text{ is convex}$$

$\mathcal{X}$ is a convex set

1st order condition for optimality: x minimizes f if

$$\boxed{\nabla f(x)^T (y-x) \geq 0 \quad \forall y \in \mathcal{X}}$$





non-convex

For unconstrained optimization:

$$\nabla f(x^*) = 0 \quad \text{if } x^* \text{ is an optimizer}$$

$$\boxed{x \text{ is optimal} \iff \nabla f(x)^T (y - x) \geq 0 \quad \forall y \in \mathcal{X}}$$

Ex: Quadratic optimization:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T Q x + c^T x \quad \} \; Q \succeq 0 \quad Q \succ 0$$

$$\text{s.t. } Ax = b$$

$$f(x) := \frac{1}{2} x^T Q x + c^T x$$

$$\nabla f(x) = Qx + c$$

x_* is optimal

1st order condition for optimality

$$\langle Q x_* + c, y - x_* \rangle \geq 0 \quad \forall y \in \mathcal{X}$$

$$Ay = b$$

$$Ax_* = b$$

$$A\underbrace{(y - x_*)}_{z} = 0$$

$$\langle Q x_* + c, z \rangle \geq 0$$

$$z \in \mathrm{Null}(A)$$

$$Q x_* + c = 0$$

Unconstrained optimization

$$\boxed{x_* = -Q^{-1} c}$$

Unconstrained optimiza