1. Intro to Optimization

1.1 Basic formulations

- $\min_{x \in R^n} f(x)$ st $c_i(x) = 0 \quad \forall i \in \mathcal{E}$
  $c_i(x) \geq 0 \quad \forall i \in \mathcal{I}$

  $\mathcal{E}$: sets of equality constraints.

  $\mathcal{I}$: sets of inequality constraints.

- ~~min~~ A more general formulation: $C = \{ x \mid c_i(x) = 0 \ \forall i \in \mathcal{E}, \ c_i(x) \geq 0 \ \forall i \in \mathcal{I} \}$

  $\min_x f(x)$ s.t. $x \in C$ ← constraints.

  (objective function)

  C : the set of "feasible solutions"

- Example: Linear regression

  Input: training data $(a_1, y_1), (a_2, y_2) \ldots \ldots (a_n, y_n)$

  $a_i \in R^d$: feature vector

  $y_i \in R$: output

  Goal: Find x st $x^T a_i \approx y_i$

  optimization problem:

  $$\min_x \underbrace{\left\{ \sum_i (a_i^T x - y_i)^2 \right\}}_{f(x)} \quad \text{... linear regression.}$$

  unconstrained

If $C = \{R^d\}$:

Unconstrained optimization

1.2 Different types of optimization problems.

a. Constrained optimization vs
unconstrained optimization

b. Continuous vs Discrete optimization
e.g. $x_i \in \{0,1\}$
$x_i \in \mathbb{Z}$

c. Deterministic vs Stochastic optimization

$(a,y) \sim D$ (a distribution)

regression to minimize expected error:

$$\min_x E_{(a,y)\sim D}\left[(a^T x - y)^2\right]$$ ----- stochastic optimization

regression by minimizing empirical error:

$$\min_x \sum_{i=1}^n (a_i^T x - y_i)^2$$ ---- Deterministic optimization

where $\{(a_i, y_i)\}_{i=1}^n$ are training data generated from D.

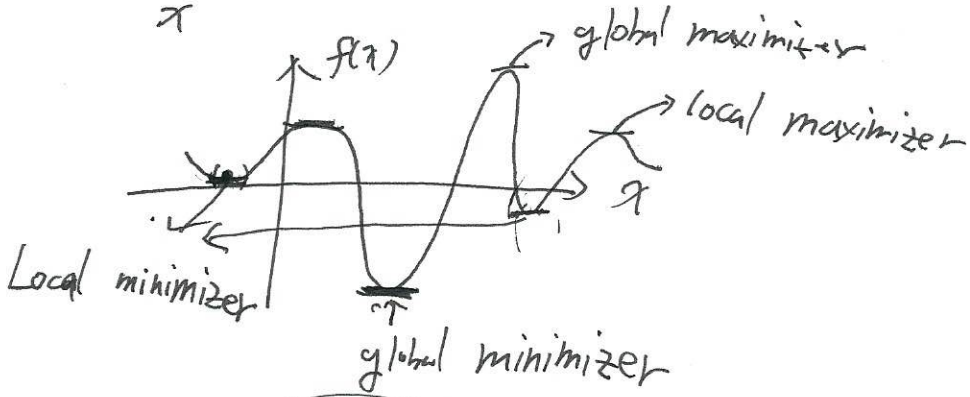
d. Smooth vs Nonsmooth objective function.

Smooth: continuous differentiable up to degree 2.

Nonsmooth: e.g. $\|x\|_1 = |x_1| + |x_2| + \cdots + |x_d|$.

1.3 Optimizer

$$\min_{x} f(x) \text{ st } x \in C$$



- $\boxed{\text{Global minimum:}}$

  x^* is a global minimizer iff

  $f(x^*) \leq f(x) \quad \forall x \in C$

- Local minimizer:

  x^* is a local minimizer iff

  $\exists\, \varepsilon > 0$ st $f(x^*) \leq f(x) \quad \forall x \in C$ and $\|x - x^*\| < \varepsilon$

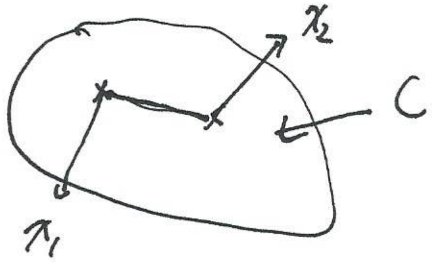
- strict global / local minimizer

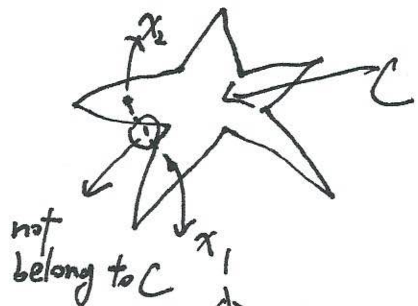
  replace "$\leq$" by "$<$" for the above definitions.

1.3 Convex Set

Def: A set C is convex iff

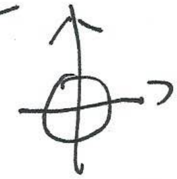
$$\forall x_1, x_2 \in C, \quad \alpha x_1 + (1-\alpha) x_2 \in C \quad \forall \alpha \in [0, 1]$$



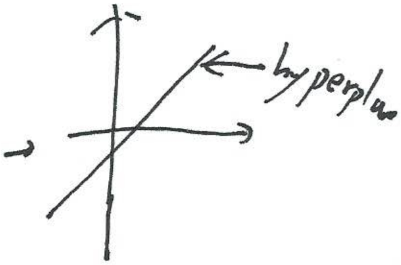


Examples:

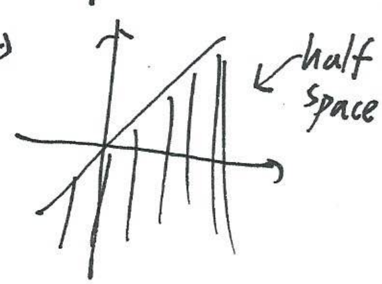
Unit ball : $\{ x \in R^n \mid \|x\| \leq 1 \}$



Hyperplane : $\{ x \in R^n \mid s^T x = b \}$ $(s \neq 0)$ →



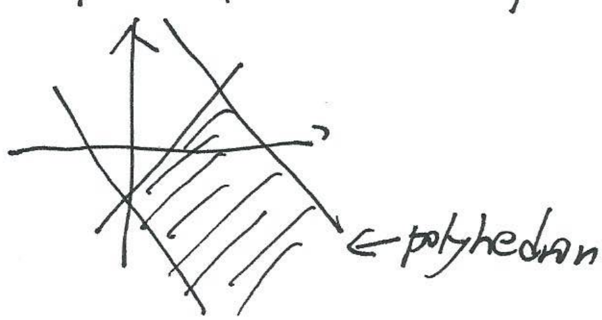
Halfspace : $\{ x \in R^n \mid s^T x \leq b \}$ $(s \neq 0)$ →



polyhedron : $\{ x \in R^n \mid Ax = b, \ Cx \leq d \}$

↓

intersection of halfspaces and hyperplanes.



Thm: If C_1, C_2 are convex sets.

$\Rightarrow C_1 \cap C_2$ is convex.

pf: $\forall x_1, x_2 \in C_1 \cap C_2$, want to show $\alpha x_1 + (1-\alpha) x_2 \in C_1 \cap C_2$

C_1 is convex $\Rightarrow \alpha x_1 + (1-\alpha) x_2 \in C_1$

C_2 is convex $\Rightarrow \alpha x_1 + (1-\alpha) x_2 \in C_2$

$\Downarrow$

$\alpha x_1 + (1-\alpha) x_2 \in C_1 \cap C_2$

#

1.4 Convex functions:

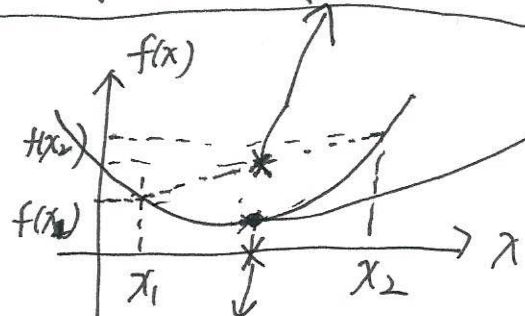
Def 1: A function f is a convex function iff

① the domain of f is a convex set

$(\mathrm{dom}(f) := \{x \mid f(x) < \infty\})$

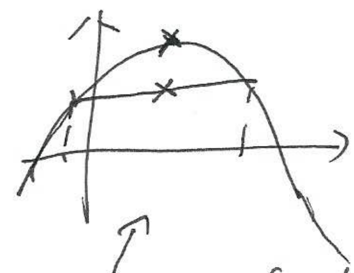
☆② $\forall x_1, x_2 \in \mathrm{dom}(f)$, $\forall \alpha \in (0, 1]$,

$$\alpha f(x_1) + (1-\alpha) f(x_2) \geq f(\alpha_1 x_1 + (1-\alpha) x_2)$$



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

$\Leftrightarrow$



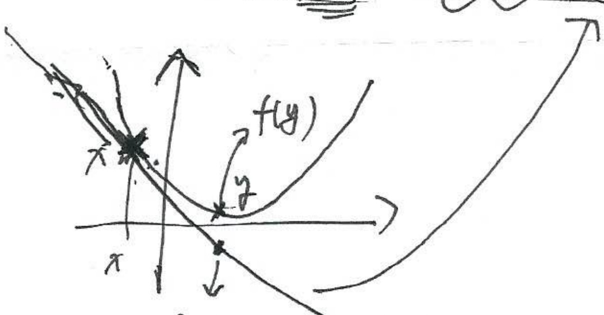
cocave function

convex

Other definitions of convex functions.

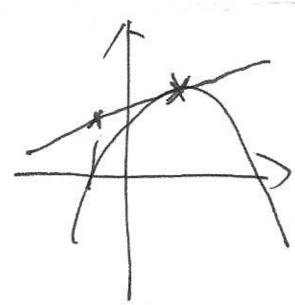
Def. 2 ① Thm: If a function f is differentiable, then f is convex iff

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



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

(first order approximation)



pf: $\forall x_1, x_2$

($\Leftarrow$) pick $\bar{x} = \alpha x_1 + (1-\alpha) x_2$

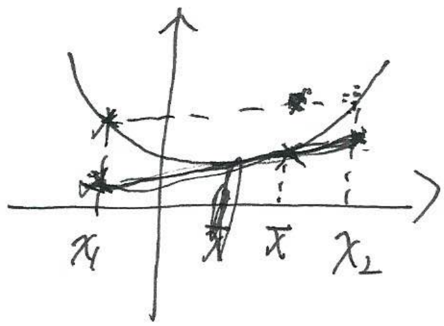
$$f(x_1) \ge f(\bar{x}) + \nabla f(\bar{x})^T (x_1 - \bar{x})$$

$$f(x_2) \ge f(\bar{x}) + \nabla f(\bar{x})^T (x_2 - \bar{x})$$

$$\Rightarrow \alpha f(x_1) + (1-\alpha) f(x_1) \ge f(\bar{x}) + \nabla f(\bar{x})^T \underbrace{(\alpha x_1 + (1-\alpha) x_2 - \bar{x})}_{=0}$$

$$= f(\bar{x})$$

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



Def 3: Thm: If f is twice differentiable, then f is convex iff $\nabla^2 f(x)$ is positive ~~defini~~ semidefinite $\forall x \in dom(f)$.

$\nabla^2 f(x)$: Hessian matrix

pf: ($\Leftarrow$) $\forall x, y$

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \underbrace{\frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x)}_{\ge 0}$$

for some z

$\nabla^2 f(z) \succeq 0 \Rightarrow$

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

($\Rightarrow$) If $\nabla^2 f(\bar{x})$ is not psd.

$\exists\, d$ st $d^T \nabla^2 f(\bar{x}) d = -c < 0$

Since $\nabla^2 f(x)$ is continuous, $\exists\, \varepsilon > 0$ st

$d^T \nabla^2 f(x) d < 0 \quad \forall x \in B_\varepsilon(\bar{x})$

1-17

$\min_{x} f(x)$

$x \in \mathrm{dom}(f)$

$$f(\bar{x} + \varepsilon \cdot \frac{d}{\|d\|}) = f(\bar{x}) + \nabla f(\bar{x})^T (\frac{\varepsilon d}{\|d\|}) + \frac{1}{2} (\frac{\varepsilon d}{\|d\|})^T \nabla^2 f(z) (\frac{\varepsilon d}{\|d\|})$$

for some $z \in B_\varepsilon(\bar{x})$ $\quad < 0$

$$\Rightarrow d^T \nabla^2 f(z) d < 0$$

$$f(\bar{x} + \varepsilon \frac{d}{\|d\|}) < f(\bar{x}) + \nabla f(\bar{x})^T (\frac{\varepsilon d}{\|d\|})$$

#

Summary:

3 ways to define convexity

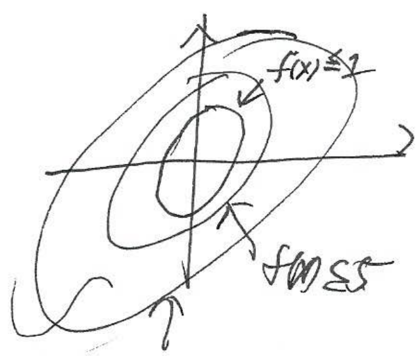
① $\forall x_1, x_2$, $f(\alpha x_1 + (1-\alpha) x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$

② $\forall x, y$, $f(y) \geq f(x) + \nabla f(x)^T (y - x)$

③ $\forall x$, $\nabla^2 f(x) \succeq 0$

Connection to Convex set

Thm: If f is convex function, then the level set $\{x \mid f(x) \leq b\}$ is a convex set $\forall b$.



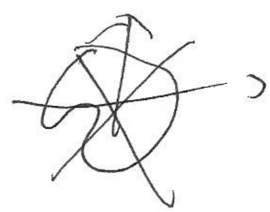
pf: If $x_1, x_2 \in C := \{x \mid f(x) \leq b\}$.

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

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

$\leq \alpha b + (1-\alpha) b = b$

#

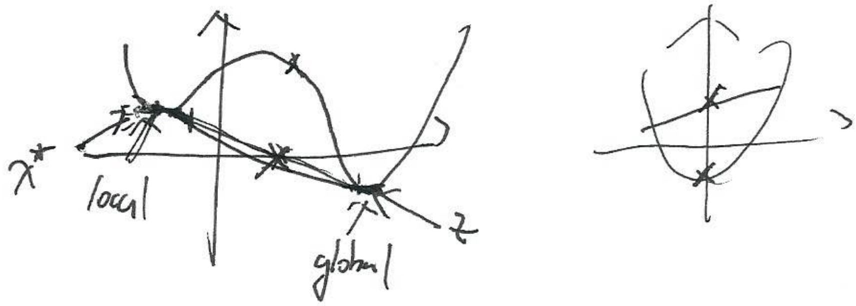


1.5 Convex Optimization

The convex optimization problem:

$$\boxed{\begin{array}{l} \min_{x} f(x) \quad \text{st } x \in C \\ - f(x) \text{ is a convex function} \\ - C \text{ is a convex set.} \end{array}}$$

Thm: For a convex optimization prob,
x^* is a local minimizer $\Leftrightarrow$ x^* is a global minimizer.



pf: If x^* is a local but not global minimizer

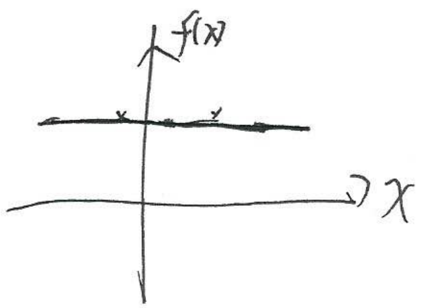
$\exists z$ st $f(z) < f(x^*)$

then $\forall \alpha \in (0,1)$

$$f(\alpha z + (1-\alpha)x^*) \leq \alpha f(z) + (1-\alpha) f(x^*)$$
$$< f(x^*)$$

Every neighborhood of x^* contains a point
$\alpha z + (1-\alpha) x^*$, so x^* is not local minimizer

#



$f(x) = a$

convex

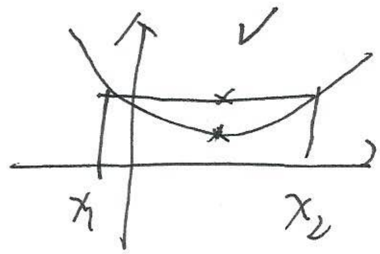
Strict convex function.

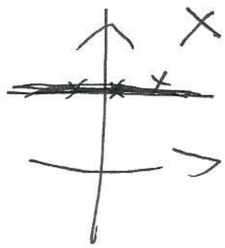
① Def: $f: R^n \to R$ is strictly convex iff

$\forall x_1 \neq x_2,\ \alpha \in (0,1)$

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

$\Updownarrow$





② Thm: f is strictly convex and differentiable, then

$f(y) > f(x) + \nabla f(x)^T (y-x) \quad \forall x, y$

$\Updownarrow$

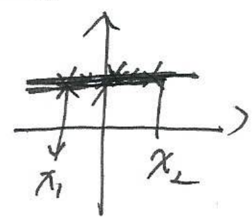
③ Thm: f is stricty convex and twice differentiable, then

$\nabla^2 f(x) > 0 \quad \forall x$

Thm: If f is strictly convex, then

$\min_x f(x)$ st $x \in C$

has a unique solution. (If it has a solution)

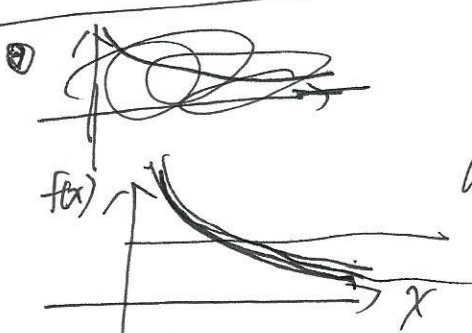


pf: If x_1, x_2 are two minimizers. Then

$f\left(\frac{x_1+x_2}{2}\right) < \frac{1}{2} f(x_1) + \frac{1}{2} f(x_2) = f(x_1) = f(x_2)$ ※

---

x^*



Thm: If $C \cap \{x \mid f(x) \leq a\}$ is closed and bounded (compact), then there exists a solution.

$\min x^2$ st $x \geq 0$

Strong Convexity

Def: A function f is strongly convex iff

$$\underbrace{f(y) \geq f(x) + \nabla f(x)^T (y-x)}_{\text{convexity}} + \underbrace{\frac{m}{2} \|y-x\|_2^2}_{\text{strong convexity}} \quad \forall x, y$$

Thm: f is twice differentiable, then

$$\underbrace{f \text{ is strongly convex}} \Leftrightarrow \nabla^2 f(x) \succeq mI \quad \forall x$$

$$m > 0$$

Thm: f is strongly convex $\Rightarrow \underbrace{(\nabla f(x) - \nabla f(y))^T (x-y) \geq m\|x-y\|^2}$

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

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

$\Rightarrow \cancel{f(x)+f(y)} \geq \cancel{f(x)+f(y)} + (\nabla f(x) - \nabla f(y))^T (y-x) + m\|x-y\|^2$

$\cancel{0 \geq} \; (\nabla f(x) - \nabla f(y))^T (x-y) \geq m\|x-y\|^2$ #

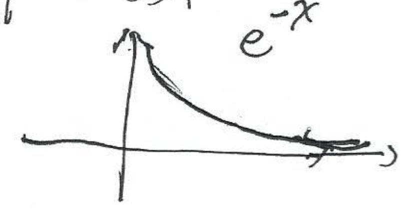
$\|\nabla f(x) - \nabla f(y)\| \|x-y\| \geq (\nabla f(x) - \nabla f(y))^T (x-y) \geq m\|x-y\|^2$

$\Rightarrow \underline{\|\nabla f(x) - \nabla f(y)\| \geq m\|x-y\|}$ .... strongly convexity

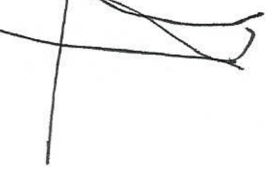
---

✓ Convex function $\Leftrightarrow \nabla^2 f(x) \succeq 0$

✓ strict convex $\Leftrightarrow \nabla^2 f(x) \succ 0$ $\longrightarrow$



✓ Strong convex $\Leftrightarrow \nabla^2 f(x) \succeq mI$, $m>0$ $\longrightarrow$



$\Downarrow$

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

$\Rightarrow \sigma_{min}(\nabla^2 f(x)) \geq m$

Why do we need strong convexity?

We will show it is ~~requi~~ usually needed when we want to show "linear" convergence rate.

Intuition (why need strong convexity)?

Q: If $\|\nabla f(x) - f(x^*)\| < \varepsilon$, can we say anything about $\|x - x^*\|$?

- $\nabla_i f(x) - \nabla_i f(x^*) = (\nabla^2 f(z)(x - x^*))_i$

$\Rightarrow \|\nabla f(x) - \nabla f(x^*)\| \geq \sigma_{min}(\nabla^2 f(z)) \|x - x^*\|$ ~~for some z~~

$\Rightarrow \|x - x^*\| \leq \|\nabla f(x) - \nabla f(x^*)\| / \sigma_{min}(\nabla^2 f(z))$

If f is convex:

σ_{min} can be 0 $\Rightarrow$ no lower bound on $\|x - x^*\|$

If f is strictly convex:

σ_{min} can be arbitrary closed to $0 \Rightarrow$ no lower bound on $\|x - x^*\|$

If f is strongly convex:

$\sigma_{min} \geq m > 0$

$$\Rightarrow \|x - x^*\| \leq \frac{\|\nabla f(x) - \nabla f(x^*)\|}{m}$$

$$\Rightarrow \text{If } \|\nabla f(x) - \nabla f(x^*)\| < \varepsilon, \ \|x - x^*\| < \frac{\varepsilon}{m}$$