LEC 1 Optimization for Data Modelling

$\begin{cases} \text{Optimization Model} \longrightarrow \text{ML model} \begin{cases} \text{Logistic Reg} \\ \text{PCA} \end{cases} \\ \text{Optimization Algorithm} \longrightarrow \text{more related to "optimization"} \begin{cases} \text{SGD} \\ \text{ADMM} \end{cases} \end{cases}$

1. Motivating examples → formulate real-life problems into mathematics forms

①
$$\begin{cases} \max_{K,L} & KL \\ \text{s.t} & 4K + L \leq 8000 \end{cases}$$

②
$$\begin{cases} \max_{P} & D(P)\cdot(P-c) \\ \text{s.t} & P \geq 0 \end{cases}$$

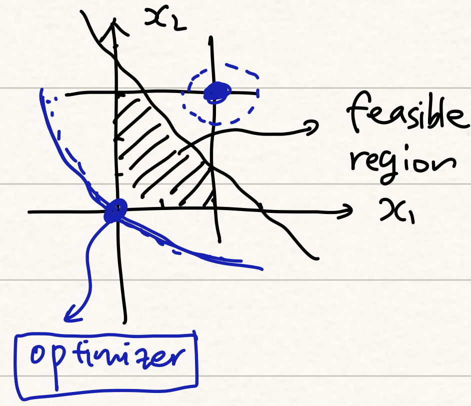
③
$$\begin{cases} \min_{x_{ij}} & \sum_i \sum_j c_{ij} x_{ij} & \\ \text{s.t} & \sum_j x_{ij} = a_i & \forall i \\ & \sum_i x_{ij} = b_j & \forall j \\ & x_{ij} \geq 0 & \forall i,j \end{cases}$$

2. Graphical Solution

① $f(x) = x\sqrt{24-x}$ $\begin{cases} \text{Necessary Cond.} \\ \text{Sufficient Cond.} \\ (\text{Convex \& Differentiable function}) \end{cases}$

$\tilde{x}$ is maximizer $\Leftrightarrow$ $f'(\tilde{x}) = 0$

② $\begin{cases} \min & (x_1-4)^2+(x_2-6)^2 \\ \text{s.t} & 0 \le x_1 \le 4 \\ & 0 \le x_2 \le 6 \\ & 3x_1 + 2x_2 \le 18 \end{cases}$



① & ② are low-dimension example → $\boxed{\text{good for intuition}}$

→ Q: What about General Non-Linear Programming ? → $\boxed{\text{high-dimension}}$

→ A: We want some $\boxed{\text{Optimality Condition}}$ to characterize the optimal $\text{sol}^{\underline{n}}$.

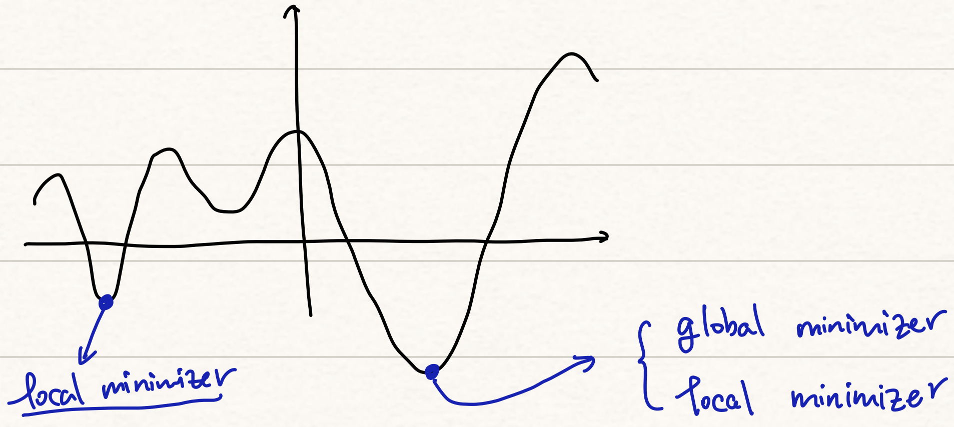
$\boxed{\text{General Setting}}$ for Non-linear Programming

① $\begin{cases} \min\limits_{x} & f(x) \\ \text{s.t} & x \in S \end{cases}$

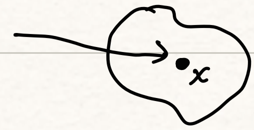
$S = \{ x : g_i(x) = 0 \,,\, h_j(x) \le 0 \}$ → feasible region

② $\begin{cases} \text{global optimizer} & \forall x \in S \,, \quad f(x') \le f(x) \\ \text{local optimizer} & \exists B(x', \varepsilon) \cap S \,, \quad f(x') \le f(x) \end{cases}$

↳ neighbourhood (small) around x'



3. Basic Calculus & Linear Algebra

① interior point x 

(Boundary point)

② Gradient $\nabla f(x) := [\partial_{x_1} f, \ldots\ldots, \partial_{x_n} f]^T$

$\boxed{-\nabla f(x)} \longrightarrow$ the direction decreases most rapidly

③ Hessian Matrix $H_f(x) := \nabla^2 f(x) := \nabla[\nabla f(x)]$

$$= \left[\frac{\partial^2}{\partial x_i \partial x_j} f\right]_{i,j}$$

$\boxed{\text{eigenvalue}(A) > 0}$

④ Positive Definite (PD) $\rightarrow$ A is PD $\Leftrightarrow$ $\forall x \neq 0,\ x^T A x > 0$

Positive Semi-Definite (PSD) $\rightarrow$ A is PSD $\Leftrightarrow$ $\forall x,\ x^T A x \geq 0$

eigenvalue(A) ≥ 0

⑤ eigenvalue calculation (Power Method)

4. Optimality Condition $\begin{cases} \text{Necessary Condition} \\ \text{Sufficient Condition} \end{cases}$

① Unconstrained Programming

a) Necessary → shrink the candidates for searching

x' is minimizer $\Rightarrow$ $\nabla f(x') = 0$

$\Rightarrow$ $H_f(x') \succeq 0$

b) Sufficient → Verification

$$\left.\begin{array}{l} \nabla f(x') = 0 \\ H_f(x') \succ 0 \end{array}\right\} \Rightarrow x' \text{ is minimizer (locally)}$$

② Constrained Programming $\Longleftrightarrow$ $\boxed{\text{KKT}}$ cond.

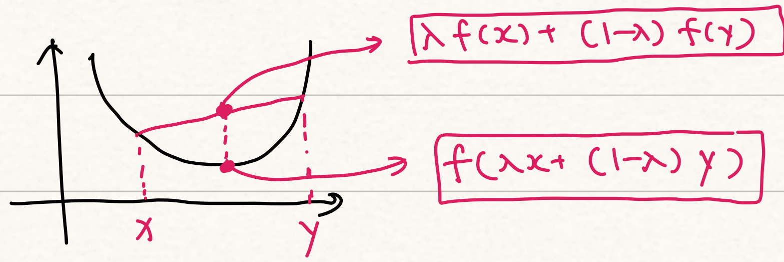
5. Convex Set → $\boxed{\text{Defn}}$ $\forall x, y \in S$, $\lambda x + (1-\lambda) y \in S$

$\forall \lambda \in [0, 1]$

Convex function → Consider function f defined on one convex set S

$\boxed{\text{Defn}}$ → $\forall x, y \in S$, $f(\lambda x + (1-\lambda) y)$

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



$\boxed{\text{strictly convex}} \longrightarrow \forall x, y \in S, \quad f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$

Q: What is the Benefits for Convexity ?

A: Local minimizer = Global minimizer

(if f is strictly convex, then have unique minimizer)

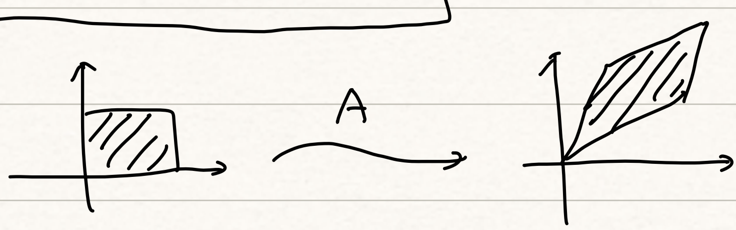
PCA Recap

1. $\boxed{\text{Eigenvalue Decomposition}}$ for symmetry matrix

$$A = Q \Lambda Q^T$$

$$\begin{cases} Q = [Q_{\cdot 1}, \dots, Q_{\cdot n}] \longrightarrow \text{eigenvector (orthogonal)} \\ \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n) \longrightarrow \text{eigenvalue} \end{cases}$$

2. $\boxed{\text{Linear Transformation}}$ A



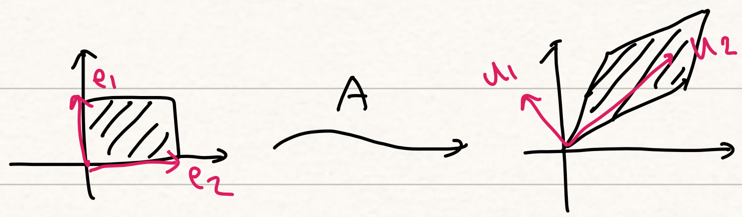
Q: How to interpret this? (Linear Transformation)

A: Eigenvalue Decomposition of A

$$\boxed{A = Q \Lambda Q^T}$$

$\rightarrow$ $Q = [\text{eigenvector } 1, \text{ eigenvector } 2]$

$\rightarrow$ $\Lambda = \text{diag}(\text{eigenvalue } 1, \text{ eigenvalue } 2)$



$\Leftarrow$ $\boxed{u_1, u_2 \text{ is the eigenvector}}$