Week 3 : Noise setting : $Y = X\theta + \underset{\substack{\uparrow \\ \text{Noise}}}{w}$

Last week: $\min \|\theta\|_1$ such that $\underset{\substack{\uparrow \\ \mathbb{R}^n}}{Y} = \underset{\substack{\uparrow \\ \mathbb{R}^{n\times d}}}{X}\overset{\substack{\mathbb{R}^d \\ \uparrow}}{\theta}$

Basis pursuit program

$$\Downarrow$$

This week: $\min \|\theta\|_1$ s.t. $\frac{1}{2n}\|Y - X\theta\|_2^2 \leq b^2$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T\theta)^2 \leq b^2$$

① $\min \|\theta\|_1$ s.t. $\frac{1}{2n}\|Y - X\theta\|_2^2 \leq b^2$

② $\min \frac{1}{2n}\|Y - X\theta\|_2^2$ s.t. $\|\theta\|_1 \leq R$

③ $\min\limits_{\theta\in\mathbb{R}^d} \left\{ \frac{1}{2n}\|Y - X\theta\|_2^2 + \lambda_n\|\theta\|_1 \right\}$ Ex. $\lambda_n = \frac{1}{n}$

Core question: ① $\overset{?}{\Longleftrightarrow}$ ② $\overset{?}{\Longleftrightarrow}$ ③

We show: ① $\Longleftrightarrow$ ③ and ② $\Longleftrightarrow$ ③

Lagrange Method — Convex program.

↓

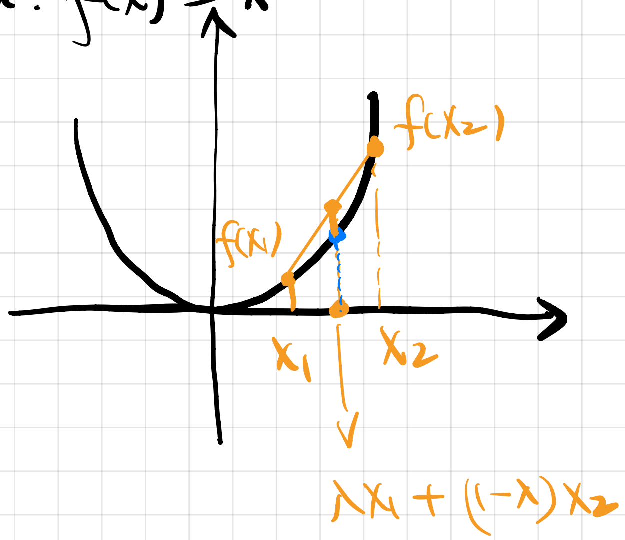
functions are convex

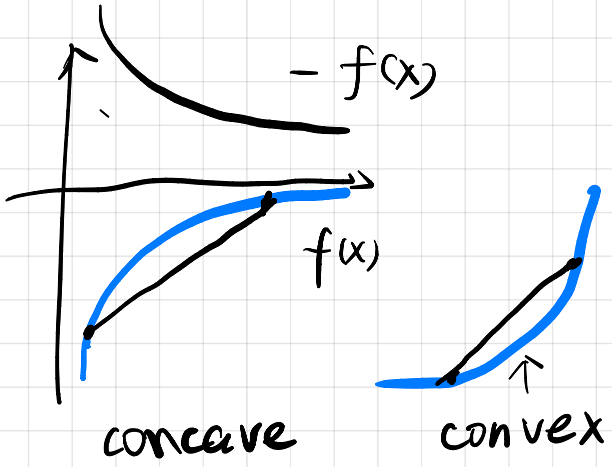
domain : convex set.

$f : \mathbb{R} \to \mathbb{R}$: convex is defined as, for $\lambda \in (0,1)$

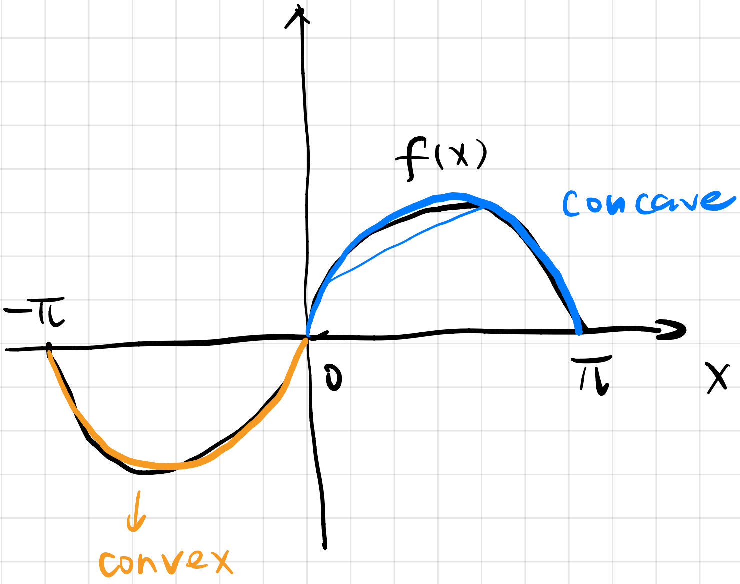
$$f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$

Ex: $f(x) = x^2$





Ex: $f(x) = \sin x$



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

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

$$\nabla \underline{\nabla f(x)} \succ 0$$

$$\Longleftrightarrow \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \\ \\ \end{bmatrix}_{d \times d}$$ is semi-positive definite (SPD)

A matrix $\Lambda \in \mathbb{R}^{d \times d}$ is SPD if for any $\underline{x, y \in \mathbb{R}^d}$

$$\underline{x^T \Lambda y \geq 0}$$

Convex optimation: $\min_x \underline{f_0(x)}$ s.t. $\boxed{\begin{array}{l} f_i(x) \leq 0, \ i \in [m] \\ h_i(x) = 0, \ i \in [n] \end{array}}$

$f_i \sim$ convex. $h_i(x) \sim$ affine: $\underline{h_i(x) = \sum_{j=1}^{d} a_{ij} x_j}$

Lagrangian function:

$$\mathcal{L}(x, \lambda, v) := \underline{f_0(x)} + \lambda^T f(x) + v^T h(x)$$

$$\mathcal{L}(x,\lambda,v) := f_0(x) + \lambda^T \underbrace{f(x)}_{\leq 0} + v^T \underbrace{h(x)}_{=0}$$

$$g(\lambda,v) = \min_{x} \mathcal{L}(x,\lambda,v)$$

$$d^* = \max_{\lambda>0,v} g(\lambda,v)$$

Question :

$$\boxed{\max_{\lambda>0,V} \min_{x} \mathcal{L}(x,\lambda,v) \quad \text{(dual)}} \overset{?}{=} \boxed{\min f_0(x) \quad \text{s.t.} \quad \begin{array}{l} h(x)=0 \\ f(x)\leq 0 \end{array} \quad \text{(primal)}}$$

$$\begin{array}{ccc} \parallel & & \parallel \\ d^* & \overset{?}{=} & p^* \end{array}$$

Theorem 3.1: If a feasible solution for (primal) exists,

then $d^* = p^*$.

(dual) $\max_{\lambda>0,V} \min_{x} \mathcal{L}(x,\lambda,v)$

$= f_0(x) + \lambda^T f(x) + v^T h(x)$

(dual) $$\max_{\underbrace{\lambda>0,\,v}_{\downarrow \atop y}} \ \min_{\underbrace{x}_{\downarrow \atop x}} \ L(x,\lambda,v) = f_0(x)+\lambda^T f(x) + v^T h(x)$$

(Theorem 3.2) If L is convex w.r.t. x

and L is concave w.r.t. y

and feasible set is not zero. then (dual) is equivalent to

$$\min_{x} \ \max_{\lambda>0,\,v} \ L(x,\lambda,v)$$

---

① $\min \|\theta\|_1$ s.t. $\frac{1}{2n}\|Y-X\theta\|_2^2 \leq b^2$

② $\min \frac{1}{2n}\|Y-X\theta\|_2^2$ s.t. $\|\theta\|_1 \leq R$

③ $\min_{\theta\in\mathbb{R}^d} \left\{ \frac{1}{2n}\|Y-X\theta\|_2^2 + \lambda_n \|\theta\|_1 \right\}$ Ex. $\lambda_n = \frac{1}{n}$

① $\min \|\theta\|_1$ s.t. $\frac{1}{2n}\|Y - X\theta\|_2^2 \le b^2$ (primal)

$$\mathcal{L}(\theta,\lambda) = \|\theta\|_1 + \lambda\left(\frac{1}{2n}\|Y-X\theta\|_2^2 - b^2\right)$$

(dual): $\max_{\lambda>0}\min_{\theta\in\mathbb{R}^d} \mathcal{L}(\theta,\lambda)$

Theorem 3.1 : (primal) ⇔ (dual)

Theorem 3.2 : (dual) ⇔ $\min_{\underline{\theta\in\mathbb{R}^d}}\max_{\lambda>0} \mathcal{L}(\theta,\lambda)$

$$\Leftrightarrow \min_{\theta\in\mathbb{R}^d} \boxed{\max_{\lambda>0} \|\theta\|_1 + \lambda\cdot\underbrace{\frac{1}{2n}\|Y-X\theta\|_2^2}_{MSE} - \lambda b^2}$$

$$\Leftrightarrow \min_{\theta\in\mathbb{R}^d} \|\theta\|_1 + \underbrace{\lambda^*}_{\text{constant}}\cdot\underbrace{\frac{1}{2n}\|Y-X\theta\|_2^2}_{MSE} \underbrace{- \lambda^* b^2}_{\text{constant}}$$

$$\Leftrightarrow \min_{\theta\in\mathbb{R}^d} \underbrace{\overset{\lambda_n}{\overset{\|}{\frac{1}{\lambda^*}}}\|\theta\|_1}_{\text{punishment}} + \underbrace{\frac{1}{2n}\|Y-X\theta\|_2^2}_{MSE}$$

$$y^* = \arg\max_{y\in Y} \phi(x^*, y)$$

$$\tilde{\phi}(y) = \min_{x\in X}\phi(x,y) = \phi(x^*, y)$$