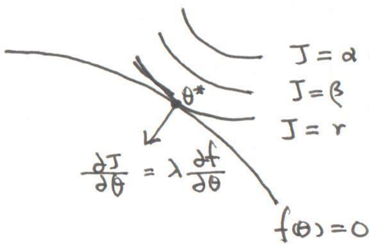
Constrained Optimization

I. Equality Constraints

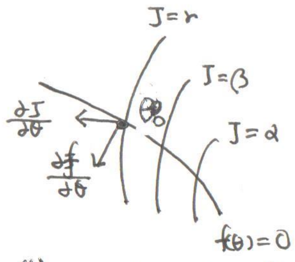
(A) min $J(\theta)$ subject to $f(\theta) = 0$

$$\Rightarrow \frac{\partial J}{\partial \theta} = \lambda \frac{\partial f}{\partial \theta}$$

① Geometric understanding



$\boxed{\alpha < \beta < \gamma}$



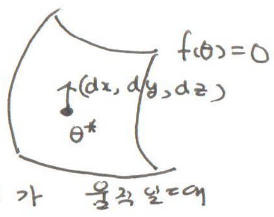
θ^* is optimal

θ' is not optimal.

② Algebraic understanding

$f(\theta) = 0$

$$\frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz = 0$$



을 만족하면서 $[dx, dy, dz]^T = d\theta$ 가 움직일때

$$\frac{\partial J}{\partial x} dx + \frac{\partial J}{\partial y} dy + \frac{\partial J}{\partial z} dz = 0$$ 이어야 함

$$\dim\left(N \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \\ \frac{\partial J}{\partial x} & \frac{\partial J}{\partial y} & \frac{\partial J}{\partial z} \end{bmatrix} \right) = 2$$

$$\therefore \frac{\partial J}{\partial \theta} = \lambda \frac{\partial f}{\partial \theta}$$

or $\mathcal{L}(\theta, \lambda) = J(\theta) - \lambda f(\theta)$ Lagrangian function

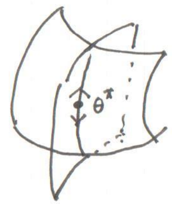
Lagrangian multiplier.

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0$$

(B) min $J(\theta)$ subject to $f_1(\theta) = 0$ and $f_2(\theta) = 0$

curve eq.

$$\begin{cases} \frac{\partial f_1}{\partial x} dx + \frac{\partial f_1}{\partial y} dy + \frac{\partial f_1}{\partial z} dz = 0 \\ \frac{\partial f_2}{\partial x} dx + \frac{\partial f_2}{\partial y} dy + \frac{\partial f_2}{\partial z} dz = 0 \end{cases}$$



일때

$$\frac{\partial J}{\partial x} dx + \frac{\partial J}{\partial y} dz + \frac{\partial J}{\partial z} dz = 0$$ 이어야 함.

$$\therefore \dim N \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ \frac{\partial J}{\partial x} & \frac{\partial J}{\partial y} & \frac{\partial J}{\partial z} \end{bmatrix} = 1$$

$$\therefore \frac{\partial J}{\partial \theta} = \lambda_1 \frac{\partial f_1}{\partial \theta} + \lambda_2 \frac{\partial f_2}{\partial \theta}$$

(C) In general

min $J(\theta)$ subject to $f_i(\theta) = 0$ $i = 1, \cdots, m$

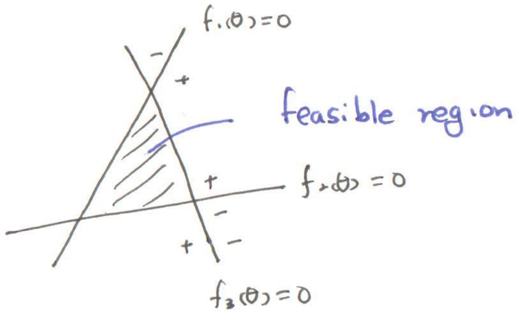
$$\mathcal{L}(\theta, \lambda) = J(\theta) - \sum_i \lambda_i f_i(\theta)$$

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0 \quad \left(\text{or } \frac{\partial J}{\partial \theta} = \sum \lambda_i \frac{\partial f_i}{\partial \theta} \right)$$

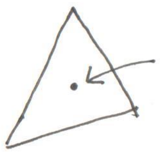
II. Inequality Constraints

Example

min $J(\theta)$ subject to $f_1(\theta) \geq 0$, $f_2(\theta) \geq 0$, $f_3(\theta) \geq 0$

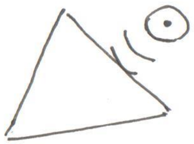


(i)



$\frac{\partial J}{\partial \theta} = 0$

(ii)

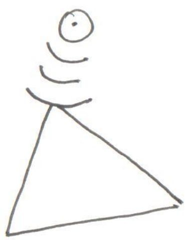


$\frac{\partial J}{\partial \theta} = \lambda_3 \frac{\partial f_3}{\partial \theta}$ f_1, f_2 inactive constraints

active constraints $\lambda_1 = \lambda_2 = 0$

$\lambda_3 \neq 0$

(iii)



$\frac{\partial J}{\partial \theta} = \lambda_1 \frac{\partial f_1}{\partial \theta} + \lambda_3 \frac{\partial f_3}{\partial \theta}$

KKT Conditions (necessary conditions)

$$\mathcal{L}(\theta, \lambda) = J(\theta) - \sum_{i=1}^{3} \lambda_i f_i(\theta)$$

① $\frac{\partial \mathcal{L}}{\partial \theta} = 0$

② $\lambda \geq 0$ (i.e. $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, $\lambda_3 \geq 0$)

③ $\lambda_i f_i(\theta) = 0$ for each i (complementary slackness conditions)

Meaning of ① & ③

(i) $\lambda_1 = \lambda_2 = \lambda_3 = 0$ all inactive

(ii) $\lambda_1 = \lambda_2 = 0$, $\lambda_3 \neq 0$

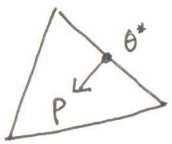
f_1, f_2 inactive f_3 active

(iii) $\lambda_2 = 0$ f_2 inactive

$f_1(\theta) = 0$ $f_3(\theta) = 0$

(가끔 λ_1 or $\lambda_3 = 0$ degenerate case)

Meaning of ② Suppose $\lambda_3 \neq 0$



$\left(\frac{\partial J}{\partial \theta}\right)^T p \geq 0$ $\left(\frac{\partial f_3}{\partial \theta}\right)^T p \geq 0$

$\because$ θ^* minimum feasible region

$$\frac{\partial J}{\partial \theta} = \lambda_3 \frac{\partial f_3}{\partial \theta}$$

$\therefore$ $\lambda_3 > 0$

In general

$\min J(\theta)$ subject to $f_i(\theta) \geq 0$ for $i = 1, \cdots, m$

KKT Conditions

$$\mathcal{L}(\theta, \lambda) = J(\theta) - \sum_i \lambda_i f_i(\theta)$$

① $\dfrac{\partial \mathcal{L}}{\partial \theta} = 0$

② $\lambda \geq 0$

③ $\lambda_i f_i(\theta) = 0$ for $i = 1, \cdots, m$

$\lambda \geq 0$ 을 가정 (KKT conditions 에 속하므로 가정할 수 있음)

Feasible region 안에서

$$\mathcal{L}(\theta, \lambda) = J(\theta) - \sum_i \lambda_i f_i(\theta) \leq J(\theta)$$

$\therefore \quad \max\limits_{\lambda \geq 0} \mathcal{L}(\theta, \lambda) = J(\theta)$ ($\because$ KKT ③을 만족할 때)

Therefore, our optimization problem =

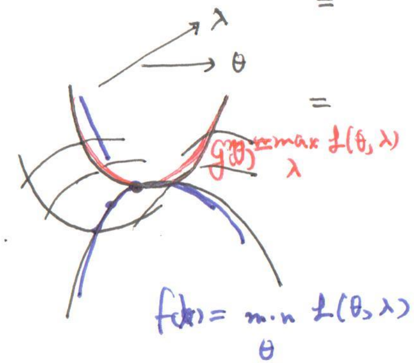
$$\min_\theta J(\theta) = \min_\theta \max_{\lambda \geq 0} \mathcal{L}(\theta, \lambda)$$

Convex Programming

(1) $J(\theta)$ is convex

(2) $f_i(\theta)$ is concave (It implies feasible region is convex)

$$\mathcal{L}(\theta^*, \lambda^*) = \text{saddle point}$$
$$= \min_\theta \max_{\lambda \geq 0} \mathcal{L}(\theta, \lambda)$$
$$= \max_{\lambda \geq 0} \min_\theta \mathcal{L}(\theta, \lambda)$$



Wolfe Dual Representation

$$\text{Our problem} = \max_{\lambda \geq 0} \min_\theta \mathcal{L}(\theta, \lambda)$$

$$= \boxed{\max_{\lambda \geq 0} \mathcal{L}(\theta, \lambda) \quad \text{subject to } \frac{\partial \mathcal{L}}{\partial \theta} = 0}$$