Lagrange Multiplier

Hanwool Jeong

hwjeong@kw.ac.kr

Implicit Function Expression & Concept of Gradient

- Plane or line
- Circle

Minimizing or Maximizing Objective w/ Constraint

- Suppose that we have the following problem

$$\text{Minimize } f(x,y) = ax+by$$

$$\text{subject to } g(x,y) = x^2 + y^2 = r$$

- Can you see the relationship between f(x,y) and g(x,y) = 0

Lagrange Multiplier

- Let's introduce an auxiliary function

$$L = f(x,y) - \lambda g(x,y)$$

- Then gradient of this function should be **0** for solving the optimization problem.

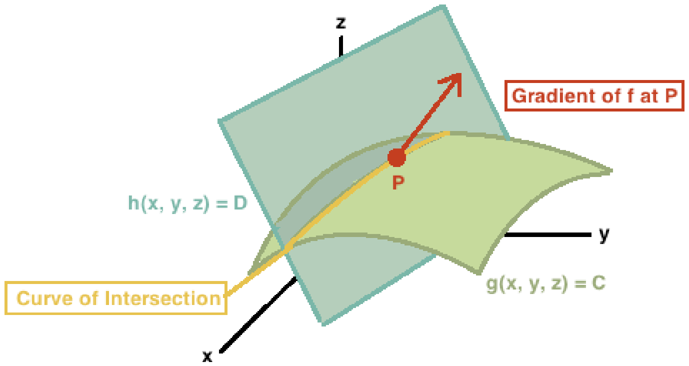
$$\nabla_{x,y,\lambda} L = 0$$

- Generally, for N dimensional vector case,

$$\begin{aligned} g\left(x_1, x_2, \ldots, x_N\right) &= 0 \\ \frac{\partial f}{\partial x_1}\left(x_1, x_2, \ldots, x_N\right) - \lambda \frac{\partial g}{\partial x_1}\left(x_1, x_2, \ldots, x_N\right) &= 0 \\ \frac{\partial f}{\partial x_2}\left(x_1, x_2, \ldots, x_N\right) - \lambda \frac{\partial g}{\partial x_2}\left(x_1, x_2, \ldots, x_N\right) &= 0 \\ &\vdots \\ \frac{\partial f}{\partial x_N}\left(x_1, x_2, \ldots x_N\right) - \lambda \frac{\partial g}{\partial x_N}\left(x_1, x_2, \ldots, x_N\right) &= 0. \end{aligned}$$

Multiple Constraints?

- Suppose that we should minimize or maximize f(x,y,z) subject to h(x,y,z) = D and g(x,y,z) = C



- Can you revise Lagrange multiplier method?

$$\nabla f(x_0, y_0, z_0) = \lambda \nabla g(x_0, y_0, z_0) + \mu \nabla h(x_0, y_0, z_0)$$

$$\frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x} + \mu \frac{\partial h}{\partial x}$$

$$\frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y} + \mu \frac{\partial h}{\partial y}$$

$$\frac{\partial f}{\partial z} = \lambda \frac{\partial g}{\partial z} + \mu \frac{\partial h}{\partial z}$$

For Multiple Constraints

- If there are M constraints, there are scalars λ_1, λ_2, …. ,λ_M such that

$$\nabla f(\mathbf{x}) = \sum_{k=1}^{M} \lambda_k \nabla g_k(\mathbf{x}) \quad \Longleftrightarrow \quad \nabla f(\mathbf{x}) - \sum_{k=1}^{M} \lambda_k \nabla g_k(\mathbf{x}) = 0.$$

- Then auxiliary function can be generalized to

$$\mathcal{L}\left(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_M\right) = f\left(x_1, \ldots, x_n\right) - \sum_{k=1}^{M} \lambda_k g_k\left(x_1, \ldots, x_n\right)$$

- Then, solve

$$\nabla_{x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_M} \mathcal{L}(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_M) = 0 \iff \begin{cases} \nabla f(\mathbf{x}) - \sum_{k=1}^{M} \lambda_k \nabla g_k(\mathbf{x}) = 0 \\ g_1(\mathbf{x}) = \cdots = g_M(\mathbf{x}) = 0 \end{cases}$$

Handling Inequality Constraints

- Suppose that

$$\text{Minimize } f(\mathbf{x})$$

$$\text{subject to } g(\mathbf{x}) \le 0$$

- Auxiliary function

$$L = f(\mathbf{x}) + \mu g(\mathbf{x})$$

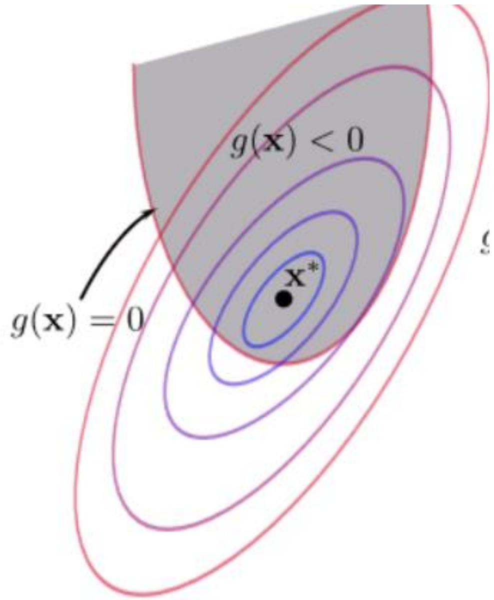
- Then solve not only $\nabla_{x,y,\lambda} L = 0$ but also

$$\mu g(\mathbf{x^*}) = 0$$

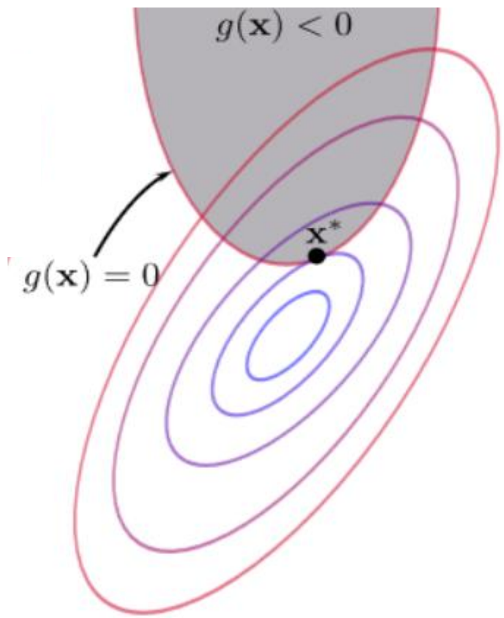
$$\mu \ge 0$$

Meaning of $\mu \cdot g(x) = 0$

1) g(x) constraint has no meaning ➔ $\mu = 0$



2) Optimal point is $g(\mathbf{x}) = 0$

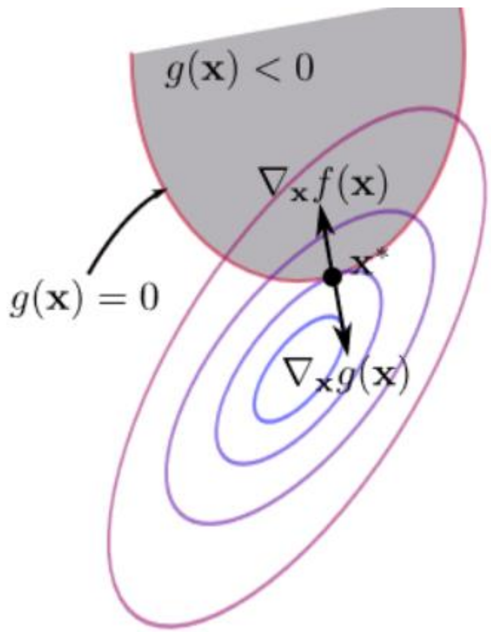


Meaning of μ ≥ 0

- When g(x*) = 0, that is 2) case in the previous slide,
- From ∇L = 0,

$$\nabla f(\mathbf{x})+\mu\nabla g(\mathbf{x}) = 0$$

$$\nabla f(\mathbf{x}) = -\mu\nabla g(\mathbf{x})$$



Handling "Multiple" Inequality Constraints

- Suppose that

$$\text{Minimize } f(\mathbf{x})$$

$$\text{subject to } \mathbf{g_i(x) \le 0} \quad \text{for } i = 1, \ldots, M$$

- Auxiliary function

$$L = f(\mathbf{x}) + \sum_i^M \mu_i g(\mathbf{x})$$

- Then solve not only **$\nabla_{x,y,\lambda} L = 0$** but also

$$\boldsymbol{\mu_i g(x^*) = 0}$$

$$\boldsymbol{\mu_i \ge 0}$$

Karush–Kuhn–Tucker (KKT) conditions

Checkpoints

✓Intro to kernel trick

✓Lagrange multiplier & KKT condition

✓SVM classification model & kernel trick application