Constrained Optimization

Hoonyoung Jeong
Department of Energy Resources Engineering
Seoul National University

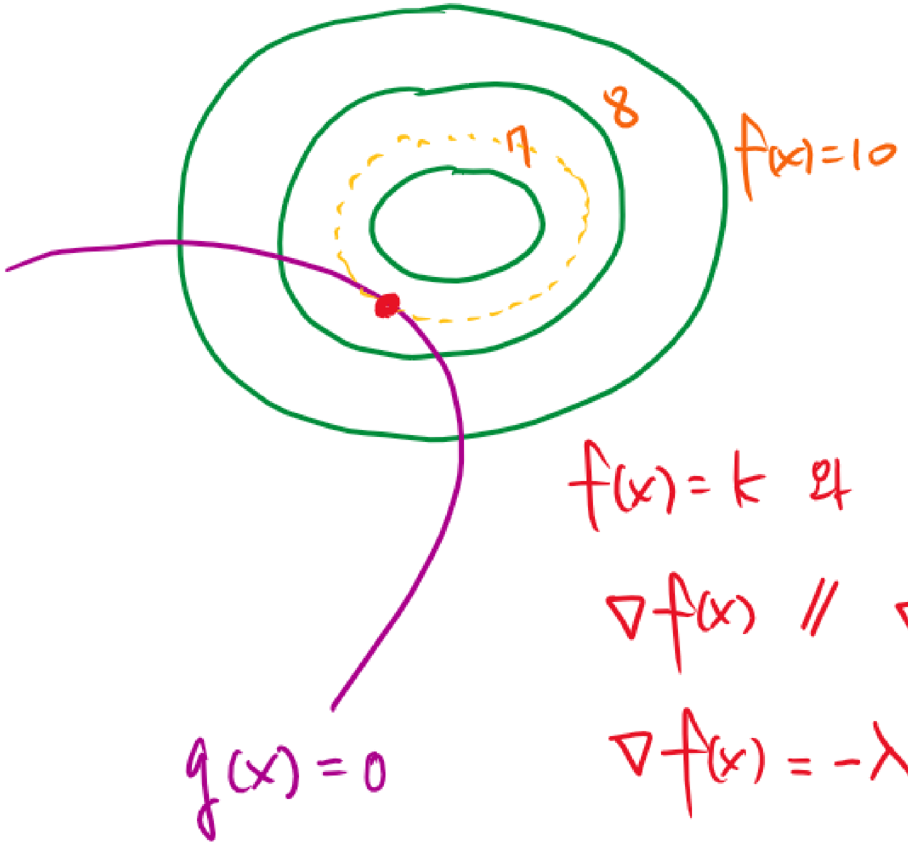
Key Questions

- Explain the Lagrange multiplier method intuitively for optimization under equality constraints
- Explain how inequality constraints are considered using the Lagrange multiplier method
- Exaplain reduced gradient
- Explain the KKT conditions
- Explain the active-set algorithm
- Explain a barrior function
- Explain a penalty function
- Explain SQP

Equality Constraints (1)

$x = \underset{x}{\operatorname{argmin}} f(x) \ , \quad x = [x_1 \cdots x_N]^T$

$g_j(x) = 0 \ , \quad j = 1, \cdots, M$



$f(x) = k$ 와 $g(x) = 0$ 이 접함.

$\nabla f(x) \ // \ \nabla g(x)$ 평행

$\nabla f(x) = -\lambda \nabla g(x)$

$$\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{bmatrix} = -\lambda \begin{bmatrix} \frac{\partial g}{\partial x_1} \\ \vdots \\ \frac{\partial g}{\partial x_N} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial f}{\partial x_1} + \lambda \frac{\partial g}{\partial x_1} = 0 \\ \vdots \\ \frac{\partial f}{\partial x_N} + \lambda \frac{\partial g}{\partial x_N} = 0 \end{bmatrix}$$

$+$

$g(x) = 0$

Equality Constraints (2)

$$\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{bmatrix} = -\lambda \begin{bmatrix} \frac{\partial g}{\partial x_1} \\ \vdots \\ \frac{\partial g}{\partial x_N} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\partial f}{\partial x_1} + \lambda \frac{\partial g}{\partial x_1} = 0 \\ \vdots \\ \frac{\partial f}{\partial x_N} + \lambda \frac{\partial g}{\partial x_N} = 0 \end{bmatrix}$$

$$+$$

$$g(x) = 0$$

This is equivalent to minimize $L(x,\lambda) = f(x) + \lambda g(x)$

$$\nabla L(x,\lambda) = 0$$

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} + \lambda \frac{\partial g}{\partial x_1} = 0$$

$$\vdots$$

$$\frac{\partial L}{\partial x_N} = \frac{\partial f}{\partial x_N} + \lambda \frac{\partial g}{\partial x_N} = 0$$

$$\frac{\partial L}{\partial \lambda} = g(x) = 0$$

$\Rightarrow$ Lagrange multiplier method

Generalization of Equality Constraints

Generalized form for $g_1, g_2, \cdots, g_M$

$$L(x, \lambda) = f(x) + \sum_{j=1}^{M} \lambda_j g_j(x)$$

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} + \sum_{j=1}^{M} \lambda_j \frac{\partial g_j}{\partial x_1}$$

$$\vdots$$

$$\frac{\partial L}{\partial x_N} = \frac{\partial f}{\partial x_N} + \sum_{j=1}^{M} \lambda_j \frac{\partial g_j}{\partial x_N}$$

$$\frac{\partial L}{\partial \lambda_1} = g_1$$

$$\vdots$$

$$\frac{\partial L}{\partial \lambda_M} = g_M$$

Example of Equality Constraints

- See “Ex_Eq_Constraint.pdf”

Inequality Constraints

* $x = \underset{x}{\operatorname{argmin}} f(x) \; , \; x = [x_1 \; \cdots \; x_N]^T$

$h(x) \le 0$

Note

$x + s^2 = 0$ for real numbers x and s

$\Leftrightarrow \; x \le 0$

$L(x, \lambda, s) = f(x) + \lambda (h(x) + s^2)$

$\Rightarrow \; x = \underset{x}{\operatorname{argmin}} f(x) \; , \; x = [x_1 \; \cdots \; x_N]^T$

slack variable

$h(x) + s^2 = 0$ where $\lambda \ge 0$

↳ smaller $h(x) + s^2$ is better

if $\lambda < 0$, opposite

$\Leftrightarrow \; h(x) \le 0$

Generalization of Equality and Inequality Constraints

$$x = \underset{x}{\operatorname{argmin}}\ f(x), \quad x = [x_1 \cdots x_N]^T$$

$$g_m(x) \leq 0 \quad \text{for } m = 1, \cdots, M$$

$$h_k(x) = 0 \quad \text{for } k = 1, \cdots, K$$

$$L(x, \lambda, \mu, s) = f(x) + \sum_{k=1}^{K} \lambda h_k(x) + \sum_{m=1}^{M} \mu_m \left(g_m(x) + s_m^2\right)$$

$$\text{where } \mu_m \geq 0$$

Mathematical understanding of Lagrange Multiplier for Equality Constraints

- See “Mathematical understanding of Lagrange Multiplier for Equality Constraints.pdf”

Other Methods for Constrained Optimization

- Reduced gradient
- KKT conditions
- Active set algorithm
- Barrier function
- Penalty function
- Sequential Quadratic Programming (SQP)

Reduced Gradient

- $\min_{x,y} f(x,y) = x^2 + 2y^2 - 4$
  subject to $h(x,y) = 2(x-1)^2 - 10y + 3 = 0$
- There seem to be 2 decision variables, but actually there is a single decision variable because an equality constraint for x and y is given
- Procedure
  - ✓ Divide (x, y) into decision and state variables
    - # of design variables = 2, # of state variables = 1, # of decision variables = 2-1
    - x and y are a decision variable (d) and a state variable (s), respectively
  - ✓ Calculate $\partial f/\partial s$, $\partial f/\partial d$, $\partial h/\partial s$,$\partial h/\partial d$
  - ✓ Calculate reduced gradient $\partial z/\partial d$ and solve for $\partial z/\partial d = 0$ where $f = z\big(d, s(d)\big)$
  - ✓ Solve for s^* from optimal d^* using h
- See the example
- Generalized Reduced Gradient (GRG) solves a problem iteratively
  - ✓ $d_{k+1} = d_k - \alpha_k \left(\frac{\partial z}{\partial d}\right)^T$
  - ✓ Calculate s_{k+1}^* using $\partial f/\partial s$, $\partial f/\partial d$, $\partial h/\partial s$,$\partial h/\partial d$
  - ✓ Correct s_{k+1} from s_{k+1}^* using the given constraints

Karush-Kuhn-Tucker Conditions (1)

$$L(x, \lambda, \mu) = f(x) + \sum_{k=1}^{K} \lambda_k h_k(x) + \sum_{m=1}^{M} \mu_m g_m(x)$$

$h_k(x) = 0$ for $k = 1, \dots, K$

$g_m(x) \leq 0$ for $m = 1, \dots, M$

no slack variables

1. Stationarity : $\nabla f(x) + \sum_{k=1}^{K} \lambda_k \nabla h_k(x) + \sum_{m=1}^{M} \mu_m \nabla g_m(x) = 0$

2. Primal constraints : $h_k(x) = 0$

   $g_m(x) \leq 0$

3. Dual constraints : $\mu_m \geq 0$, $\lambda_k \neq 0$

4. Complementary slackness : $\mu_m g_m(x) = 0$

$\Rightarrow$ If x^* satisfies KKT conditions, x^* is a minimum candidate.

(only $\nabla L = 0$)

Karush-Kuhn-Tucker Conditions (2)

$$L(x, \lambda, \mu) = f(x) + \sum_{k=1}^{K} \lambda_k h_k(x) + \sum_{m=1}^{M} \mu_m (g_m(x) + s_m^2)$$

$h_k(x) = 0$ for $k = 1, \dots, K$

$g_m(x) \leqq 0$ for $m = 1, \dots, M$

1. $\partial L / \partial x$: $\nabla f(x) + \sum \lambda_k \nabla h_k(x) + \sum \mu_m \nabla g_m(x) = 0$

2. $\partial L / \partial \lambda_k$: $h_k(x) = 0$

3. $\partial L / \partial \mu_m$: $g_m(x) + s_k^2 = 0$

4. $\partial L / \partial s_k$: $\mu_m s_m = 0$

5. $\lambda_k \neq 0, \ \mu_m \geq 0$

Active-Set Algorithm

- An algorithm to solve inequality constrained optimization problems
- Procedure
  1) Start with an initial working inequality constraint set
     - ✓ Which consists of inequality constraints satisfying an initial solution
  2) Solve the problem for only the working set using KKT
  3) Check the constraints and the Lagrange multipliers (LM)
     - ✓ A Lagrange multiplier for an inequality constraint ≥ 0
  4) If all the constraints and Lagrange multipliers are satisfied, terminate
  5) Remove an inequality constraint from the working set that has the most negative LM
  6) Add an inequality constraint to the working set that is the most violated
  7) Repeat from Step 2

Barrier Function

- Add barrier functions to the Lagrange function to consider inequality constraints

Interior-point method (or barrier method)

$$\phi(x) = -\sum_{m=1}^{M} \log(-g_m(x)) \quad : \text{barrier function}$$

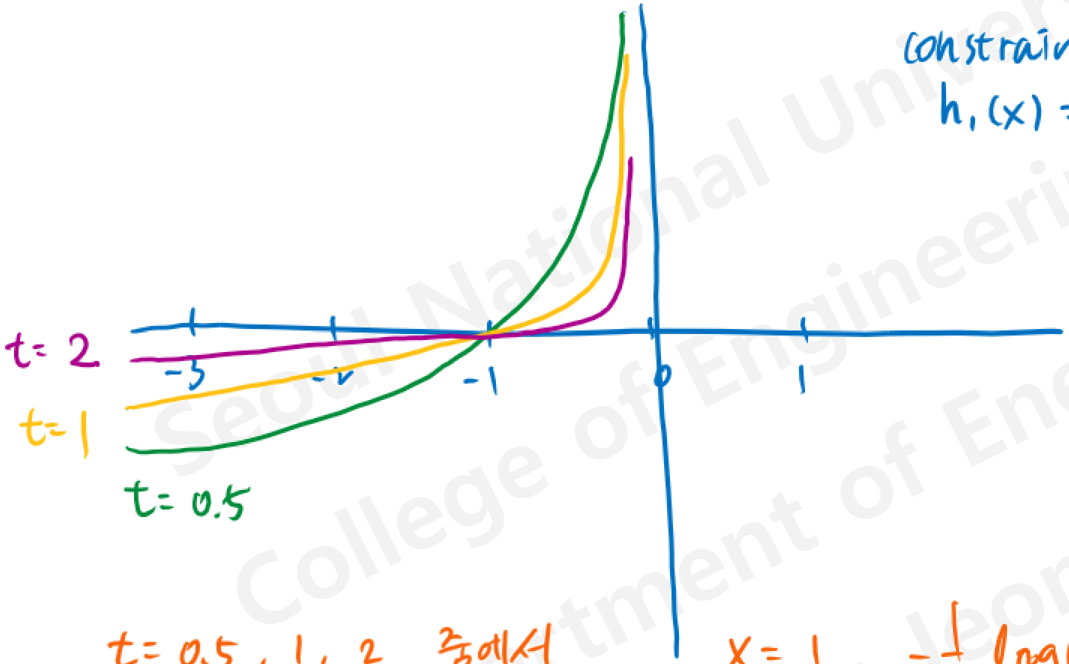
$$L(x, \lambda) = f(x) + \sum_{k=1}^{K} \lambda_k h_k + \frac{1}{t}\phi(x)$$

ex) $h_1(x) = x$

$$\phi(x) = -\log(-x)$$

$$\frac{1}{t}\phi(x) = -\frac{1}{t}\log(-x)$$

Example of Barrier Function



constraint

$h_1(x) = x \le 0$

X가 0보다 작을수록 $-1/2 \log(-x)$ 더 작아짐.
$X \le 0$ 을 모사하기에 적절함.
Inequality constraint를 함수 형태로 표현함으로써 미분가능해짐. gradient 기반 방법 적용 가능해짐.

t는 step마다 점점 커짐. (초반에는 느슨하게)

$t_{k+1} = \mu t_k \ , \ \mu > 1$

t= 0.5, 1, 2 중에서
t=2 가 가장 잘 모사

$\frac{1}{t}\phi(x) = -\frac{1}{2}\log(-x)$

$x = 1 \ , \ -\frac{1}{2}\log(-1) = \text{inf}$

$x = 0 \ , \ -\frac{1}{2}\log(0) = \text{inf}$

$x = -1 \ , \ -\frac{1}{2}\log(1) = 0$

$x = -2 \ , \ -\frac{1}{2}\log(2) < 0$

Penalty Function

$$\min_{x} T(x) = f(x) + \frac{1}{r} P(x) \quad , r>0 \longrightarrow$$ smaller r, penalty function values are more weighted

penalty function

inequality $P(x) = \sum_{m=1}^{M} [\max(0, g_m(x))]^2$

equality $P(x) = \sum_{k=1}^{K} h_k(x)^2$

Sequential Quadratic Programming (SQP)

- Approximate the objective function using the 2^{nd} order Taylor series (a quadratic function)
- Linerize constraints using the 1^{st} order Taylor series
- See the example