Convex Optimization

Wha Sook Jeon
Mobile Computing & Communications Lab.
Seoul National University

Convex Optimization

- General Optimization Problem

$$\begin{aligned} &\text{minimize} && f_0(x) \\ &\text{subject to} && f_i(x) \le 0, \quad \mathrm{i}=1,2,\ldots,\mathrm{m} \\ & && h_i(x) = 0, \quad \mathrm{i}=1,2,\ldots,\mathrm{p} \end{aligned}$$

$$x = (x_1, x_2, \cdots, x_n)$$
$$f_0 : R^n \to R$$
$$f_i : R^n \to R$$
$$h_i : R^n \to R$$

- Example

$$\begin{aligned} &\text{minimize} && -\int_0^\infty B\log_2\left(1+P_t(\gamma)\gamma/\overline{P}\right)p(\gamma)d\gamma \\ &\text{subject to} && \int_0^\infty \left(P_t(\gamma)/\overline{P}\right)p(\gamma)d\gamma \le 1 \end{aligned}$$

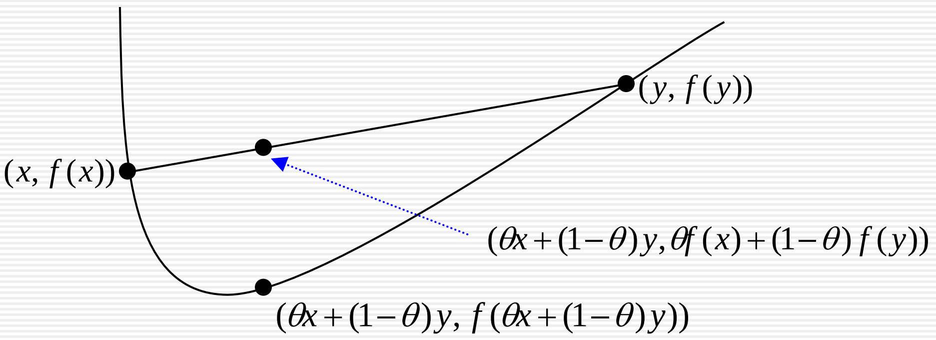
- A problem is a convex optimization problem if
  - Objective function & Inequality constraint functions: convex
  - Equality constraint functions: linear (or affine)
- Convexity guarantees some nice properties to solve it.

Convex function (1)

$f : \mathrm{R}^n \to \mathrm{R}$ is convex

if dom f is a convex set and for all $x, y \in$ dom f and θ with $0 \le \theta \le 1$,

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$$



$h : \mathrm{R}^n \to \mathrm{R}^m$ is affine

if it is a sum of a linear function and a constant, i.e.,

$h(x) = Ax + b$ where $A \in \mathrm{R}^{m \times n}$ and $b \in \mathrm{R}^m$

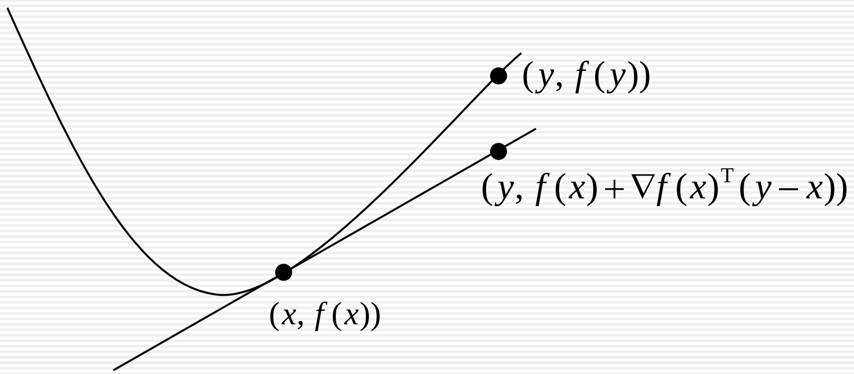
Convex function (2)

First-order Condition

When f is differentiable,

f is convex if $\mathrm{dom}\ f$ is a convex set,

$f(y) \geq f(x) + \nabla f(x)^{\mathrm{T}}(y - x)$



Second-order Condition

When f is twice differentiable,

f is convex if $\mathrm{dom}\ f$ is a convex set and

$\nabla^2 f(x) \geq 0$ for all $x \in \mathrm{dom}\ f$

Primal Problem & Dual Problem

- Primal Problem

$$\begin{aligned} &\text{minimize} \quad f_0(x) \\ &\text{subject to} \quad f_i(x) \le 0, \quad \mathrm{i} = 1,2,\ldots,\mathrm{m} \\ &\qquad\qquad\quad h_i(x) = 0, \quad \mathrm{i} = 1,2,\ldots,\mathrm{p} \end{aligned}$$

- Lagrangian $L : R^n \times R^m \times R^p \to R$

$$L(x,\lambda,\upsilon) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \upsilon_i h_i(x)$$

  - Lagrangian multipliers: $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_m) \quad \upsilon = (\upsilon_1, \upsilon_2, \cdots, \upsilon_p)$

- Lagrange Dual Function

$$g(\lambda,\upsilon) = \inf_{x \in D} L(x,\lambda,\upsilon)$$

  - D: a domain of a decision vector x

Lower Bound Property (1)

- Suppose $\tilde{x}$ is a feasible point for the primal problem:

  $f_i(\tilde{x}) \le 0$ and $h_i(\tilde{x}) = 0$

- Since $\lambda \ge 0$, $\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \upsilon_i h_i(\tilde{x}) \le 0$

  Then, $L(\tilde{x}, \lambda, \upsilon) = f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \upsilon_i h_i(\tilde{x}) \le f_0(\tilde{x})$

  Thus, $g(\lambda, \upsilon) = \inf_{x \in D} L(x, \lambda, \upsilon) \le L(\tilde{x}, \lambda, \upsilon) \le f_0(\tilde{x})$

  Since $g(\lambda, \upsilon) \le f_0(\tilde{x})$ holds for every feasible point $\tilde{x}$, $g(\lambda, \upsilon) \le f_0(x^*)$

$g(\lambda, \upsilon) \le p^*$ for any $\lambda \ge 0$ and any υ,

p^* : optimal solution of the primal problem

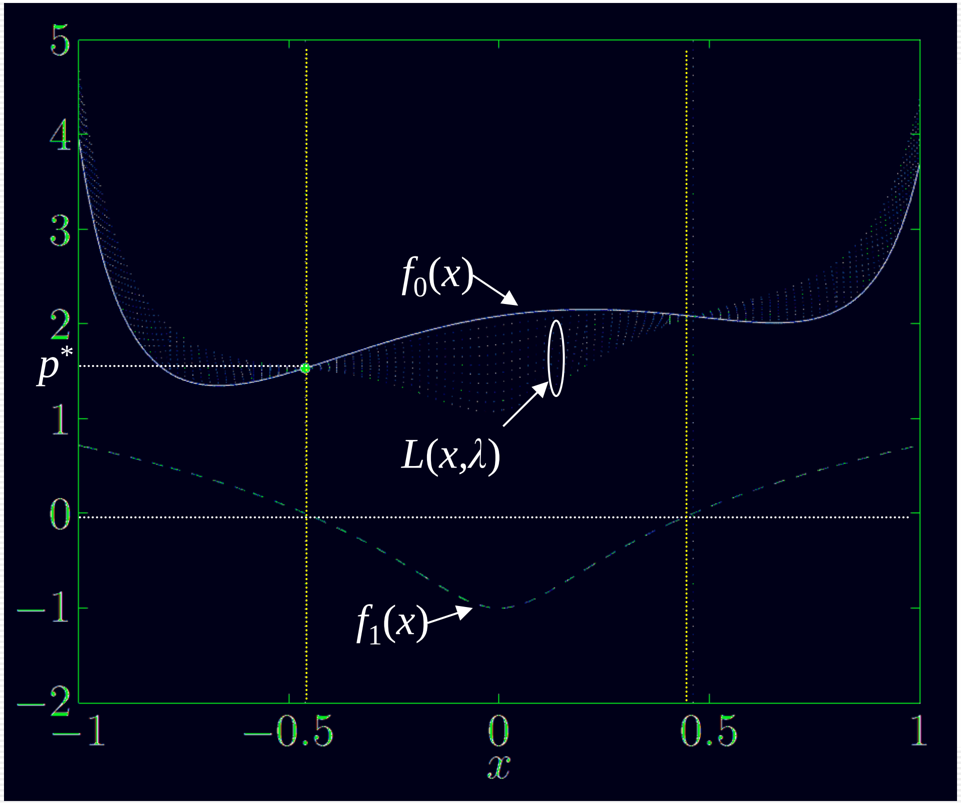
Lower Bound Property (2)

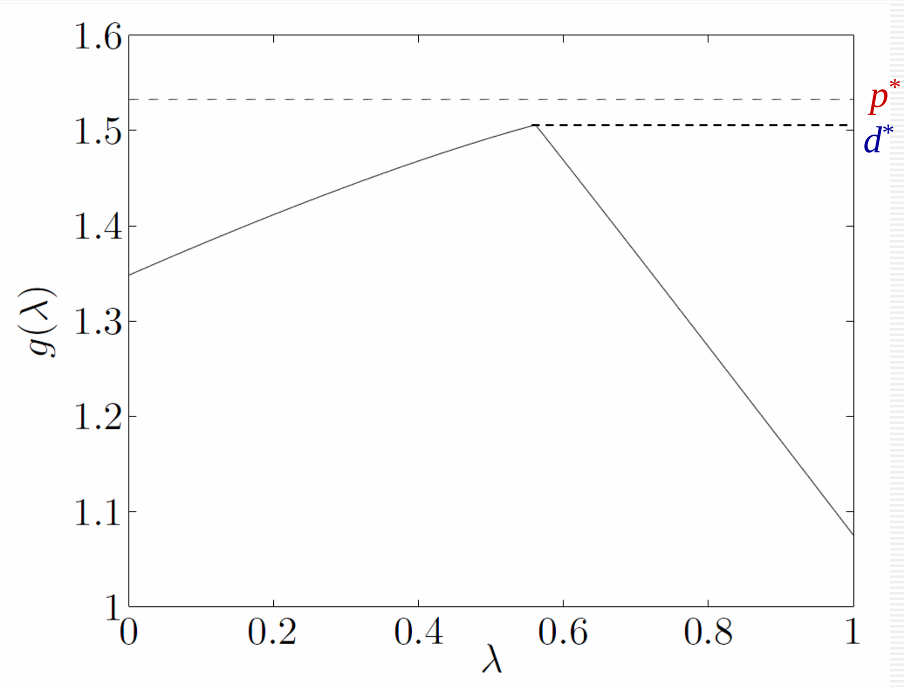
Example :

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \le 0 \end{array}$$

$\Rightarrow$

$$L(x,\lambda) = f_0(x) + \lambda f_1(x)$$
$$g(\lambda) = \inf_x L(x,\lambda) \quad \text{for } \lambda \ge 0$$

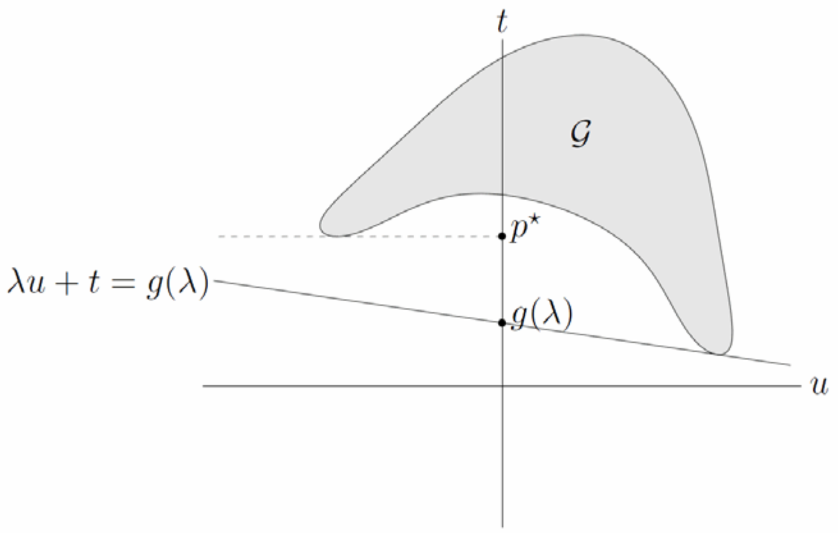


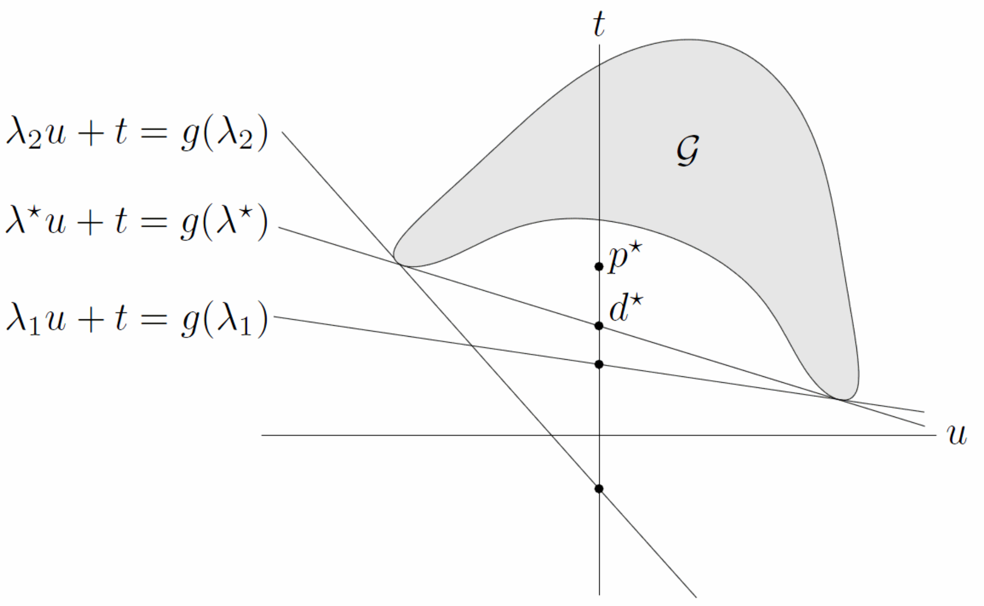


Geometric Representation

$$\Omega = \{(f_1(x), f_2(x), \cdots, f_m(x), h_1(x), h_2(x), \cdots, h_p(x)), f_0(x) \in R^m \times R^p \times R \mid x \in D\}$$

$$p^* = \inf\{t \mid (u, v, t) \in \Omega, u \preceq 0, v = 0\}$$





Lagrange Dual Problem

- What is the best lower bound from the Lagrange dual function?
- Lagrange Dual Problem

$$\begin{aligned} &\text{maximize } \; g(\lambda, \upsilon) \\ &\text{subject to } \; \lambda_i \geq 0, \quad i = 1,2,\ldots,m \end{aligned}$$

  - This dual problem is a convex optimization problem, although the primal problem is not convex
  - Thus, generally it is easier to solve than the primal problem.
- Duality
  - Weak Duality (always holds): $d^* \leq p^*$

    (d^*: Solution of Lagrange dual problem)
  - Strong Duality: $d^* = p^*$

Strong Duality Condition

- Convex Primal Problem

$$\begin{aligned} &\text{minimize} \quad f_0(x) \\ &\text{subject to} \quad f_i(x) \le 0, \quad \mathrm{i} = 1,2,\ldots,\mathrm{m} \\ &\qquad\qquad\quad Ax = b \end{aligned}$$

  - $f_0(x), f_1(x), \cdots, f_m(x)$: convex functions
- Slater's constraint qualification

$$\begin{aligned} &\text{There exists an } x \in \operatorname{relint} D \text{ such that} \\ &\qquad f_i(x) < 0, \quad i = 1, \cdots, m, \\ &\qquad Ax = b \end{aligned}$$

  - *x*: strictly feasible point
- When the primal problem is convex
  - if Slater's condition holds, strong duality holds

Strong Duality Condition

- Primal Problem

$$\begin{aligned} &\text{minimize} \quad f_0(x) \\ &\text{subject to} \quad f_i(x) \le 0, \quad \mathrm{i} = 1,2,\ldots,\mathrm{m} \\ &\qquad\qquad\quad h_i(x) = 0, \quad \mathrm{i} = 1,2,\ldots,\mathrm{p} \end{aligned}$$

- Lagrange Dual Problem

$$\begin{aligned} &\text{maximize} \quad g(\lambda,\upsilon) = \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \upsilon_i h_i(x) \right) \\ &\text{subject to} \quad \lambda_i \ge 0 \quad \text{for all } i \end{aligned}$$

- Strong Duality: $f_0(x^*) = g(\lambda^*,\upsilon^*)$
  - x^*: primal optimal point, (λ^*,υ^*): dual optimal point
  - $$\begin{aligned} \mathrm{g}(\lambda^*,\upsilon^*) &= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \upsilon_i^* h_i(x) \right) \\ &\le f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \upsilon_i^* h_i(x^*) \\ &\le f_0(x^*) \end{aligned}$$

=

=

KKT Optimality Conditions

- $f_0, f_1, \cdots, f_m, h_1, \cdots, h_p$ should be differentiable.
- $g(\lambda^*, \upsilon^*)$

$$= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \upsilon_i^* h_i(x) \right)$$

$$\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \upsilon_i^* \nabla h_i(x^*) = 0$$

$$= f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \upsilon_i^* h_i(x^*)$$

$$\sum_{i=1}^p \lambda_i^* f_i(x^*) = 0 \Rightarrow \lambda_i^* f_i(x^*) = 0$$

$$= f_0(x^*)$$

- KKT conditions

$$f_i(x^*) \le 0, \qquad i = 1, \cdots, m$$

$$h_i(x^*) = 0, \qquad i = 1, \cdots, p$$

$$\lambda_i^* \ge 0, \qquad i = 1, \cdots, m$$

$$\lambda_i^* f_i(x^*) = 0, \qquad i = 1, \cdots, m,$$

$$\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \upsilon_i^* \nabla h_i(x^*) = 0,$$

Example: Water Filling (1)

- Problem Description
  - The problem is to allocate power to a set of *n* subchannels so that the total data rate is maximized.
  - The total power allocated to *n* subchannels is fixed to P_T.
- Problem Formulation
  - P_i: the power of subchannel *i*,
  - x_i : the normalized power of subchannel *i*: P_i/P_T
  - γ_i : SNR of subchannel *i* when the transmission power is P_T
  - Total data rate: $\sum_{i=1}^{n} B\log_2\left(1+\frac{P_i}{P_T}\gamma_i\right) = \sum_{i=1}^{n} B\log_2(1+x_i\gamma_i)$
  -

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} B\log_2(1+x_i\gamma_i) \\ \text{subject to} \quad & x_i \geq 0, \quad i = 1,\cdots,n \\ & \sum_{i=1}^{n} x_i = 1 \end{aligned}$$

$$\rightarrow B\sum_{i=1}^{n}\log_2(1/\gamma_i + x_i) - B\sum_{i=1}^{n}\log_2(\gamma_i)$$

$$\rightarrow \text{minimize} \quad -\sum_{i=1}^{n}\ln(1/\gamma_i + x_i)$$

Example: Water Filling (2)

- Primal Problem (Convex Problem)

$$\begin{aligned} &\text{minimize} \quad -\sum_{i=1}^{n} \ln(1/\gamma_i + x_i) \\ &\text{subject to} \quad \mathbf{x} \geq 0, \quad \mathbf{1}^{\mathrm{T}} \mathbf{x} = 1 \end{aligned}$$

- Lagrangian

$$-\sum_{i=1}^{n} \ln(1/\gamma_i + x_i) - \sum_{i=1}^{n} \lambda_i x_i + \upsilon\left(\sum_{i=1}^{n} x_i - 1\right)$$

Example: Water Filling (3)

- KKT condition:
  - $\mathbf{x}^*$: primal optimal point, $(\boldsymbol{\lambda}^*, \upsilon^*)$: dual optimal point
  - $$\mathbf{x}^* \geq 0,\ \ \mathbf{1}^T\mathbf{x}^* = 1,\ \ \boldsymbol{\lambda}^* \geq 0,$$
    $$\lambda_i^* x_i^* = 0, \quad i = 1, \cdots, n,$$
    $$\frac{-1}{1/\gamma_i + x_i^*} - \lambda_i^* + \upsilon^* = 0,\ \ i = 1, \cdots, n.$$
    $\Rightarrow$
    $$\mathbf{x}^* \geq 0,\ \ \mathbf{1}^T\mathbf{x}^* = 1,$$
    $$\upsilon^* \geq \frac{1}{1/\gamma_i + x_i^*}, \quad i = 1, \cdots, n,$$
    $$x_i^*(\upsilon^* - 1/(1/\gamma_i + x_i^*)) = 0,\ \ i = 1, \cdots, n,$$

- Solution
  - if $\upsilon^* < \gamma_i,\ \ x_i^* > 0 \ \Rightarrow\ \upsilon^* = \dfrac{1}{1/\gamma_i + x_i^*} \Rightarrow x_i^* = \dfrac{1}{\upsilon^*} - \dfrac{1}{\gamma_i}$

    if $\upsilon^* \geq \gamma_i,\ \ x_i^* \leq 0 \ \Rightarrow\ x_i^* = 0$

$$x_i^* = \begin{cases} 1/\upsilon^* - 1/\gamma_i & \text{if } \gamma_i > \upsilon^* \\ 0 & \text{if } \gamma_i \leq \upsilon^* \end{cases}$$

Convex Optimization

1. Formulate a primal problem
2. Derive the Lagrangian
3. Construct the Lagrange dual function
   - Concave
   - Not necessarily differentiable
4. Solve the Lagrange dual problem (convex optimization)
   - There are various methods
   - Gradient (subgradient) method
     - differential: projected gradient method
     - not differential: projected subgradient method
5. Prove the strong duality

Gradient/Subgradient Method

- Consider the following general concave maximization:

$$\begin{aligned}&\text{maximize}\ g(\lambda)\\&\text{subject to}\ \ \lambda \in \chi\end{aligned}$$

(where χ is a convex set)

- The methods generate a sequence of feasible points $\{\lambda(k)\}$ as

$$\lambda(k+1) = \left[\lambda(k) + \alpha(k) \times s(k)\right]_{\chi}$$

  - $\boldsymbol{s}(k)$: a gradient(subgradient) of g at the point of $\lambda(k)$
    - It depends on whether g is differentiable or not.
  - $[\cdot]_{\chi}$: the projection onto the feasible set χ
  - $\alpha(k)$: a positive step size at the kth iteration

Subgradient Method

- Subgradient property

$$g(y) \le g(x) + s(x) \times (y - x)$$

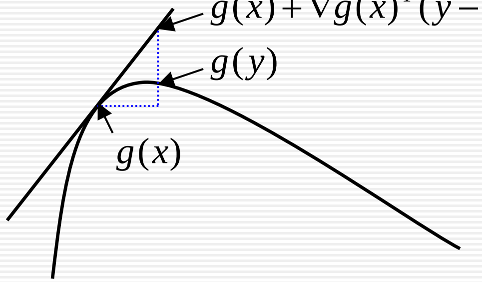
  - $s(x)$ is a subgradient of g at x

- Simple example

$$g(\lambda) = \min_{x \in X} \{ f_0(x) + \lambda f_1(x) \}$$

$$x(\lambda) = \arg\min_{x \in X} \{ f_0(x) + \lambda f_1(x) \}$$

Gradient property



$$\begin{aligned} g(\mu) &= \min_{x \in X} \{ f_0(x) + \mu f_1(x) \} \\ &\le f_0(x(\lambda)) + \mu f_1(x(\lambda)) \\ &= f_0(x(\lambda)) + \lambda f_1(x(\lambda)) + (\mu - \lambda) f_1(x(\lambda)) \\ &= g(\lambda) + (\mu - \lambda) f_1(x(\lambda)) \end{aligned}$$

$f_1(x(\lambda))$ is a subgradient of g at λ

Subgradient Method for Solving Lagrange Dual Problem

- For a given λ, υ

$$x(\lambda,\upsilon) = \arg\min_{x \in D} L(x, \lambda, \upsilon)$$
$$= \arg\min_{x \in D} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \upsilon_i h_i(x) \right\}$$

- Subgradients of g at (λ,υ): $f_i(x(\lambda,\upsilon)),\ h_i(x(\lambda,\upsilon))$

- For obtaining an optimal λ^*, υ^*

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} + \delta^{(k)} \times f_i(x(\lambda^{(k)}, \upsilon^{(k)}))$$
$$\upsilon_i^{(k+1)} = \upsilon_i^{(k)} + \delta^{(k)} \times h_i(x(\lambda^{(k)}, \upsilon^{(k)}))$$

⇨ converge to λ^*, υ^*

$\delta^{(k)}$: positive step size at the kth iteration

$$\delta^{(k)} > 0, \sum_{k=1}^{\infty} \delta^{(k)} = \infty, \sum_{k=1}^{\infty} \left(\delta^{(k)}\right)^2 < \infty \quad (\text{ex. } \delta^{(k)} = c/k)$$