Lecture 4-5 Dual Method

Primal

Key point ① ▨ Solving $\min f(\underset{\sim}{x})$

subject to $g_i(x) \le 0$ and $h_j = 0$

$\Updownarrow$ equivalent to

"Dual"

▨ Solving $\max\limits_{\mu \ge 0, \lambda} \phi(\lambda, \mu)$

where $\phi(\lambda, \mu) = \min\limits_{\underset{\sim}{x}} \left[f(\underset{\sim}{x}) + \sum \lambda_j h_j + \sum \mu_i g_i \right]$

[under some conditions]

Key point ② : In some cases or with some approximations, it is much easier to solve dual problems than primal problems.

Lectures on this subject?

i) Proof of Duality of the problem

ii) examples to understand dual problem

iii) Special cases where primal problems are extremely simple to solve → "separable problems"

iv) Alogorithm consisting of Convex approximation ⊕ seprable assumption (MMA)

<Primal problem>

$$\min f(x)$$

$$\text{s.t.} \quad g_i(\underline{x}) \le 0, \quad i=1,\cdots,m; \quad h_j(\underline{x}) = 0, \quad j=1,\cdots,l$$

KKT : if $\underline{x}^*$ is a Local min and a regular pt (CQ)

then $\exists$ $\underline{\mu}^* \geq 0$ and λ^* such that

$$\begin{cases} \nabla f(x^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\underline{x}^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(\underline{x}^*) = 0 \\ \sum_{i=1}^{m} \mu_i \, g_i(x^*) = 0 \end{cases}$$

< Dual Problem>

$$\max_{\underline{\mu} \geq 0, \underline{\lambda}} \phi(\underline{\lambda}, \underline{\mu})$$

$$\phi(\underline{\lambda}, \underline{\mu}) = \min_{\underset{\sim}{x}} \left[f(\underline{x}) + \sum \lambda_j h_j(\underline{x}) + \sum \mu_i g_i(\underline{x}) \right]$$

Assumption: Local Convexity that the Hessian of the Lagrangian is positive-definite on the whole Space (not just on the tangent plane M)

$$\Big[H_L(\underline{x}^*, \underset{\sim}{\lambda}^*, \mu_i^*) = H_f(\underline{x}^*) + \sum \lambda_j H_{h_j}(x^*) + \sum \mu_i H_{g_i}(x^*) \Big]$$

Then

Local Max of ϕ = Local Min of f

i.e; ① $(\underline{\lambda}^*, \underline{\mu}^*)$ of DP $= (\lambda^*, \mu^*)$ of PP

Dual problem ↗ primal problem

(*) $\underline{x}_{min}(\underline{\lambda}^*, \underline{\mu}^*)$ of DP $= \underline{x}^*$ of PP

② Local max of ϕ

= Local Min of f

(at $\underline{\lambda}=\underline{\lambda}^*$, $\underline{\mu}=\underline{\mu}^*$, $\underline{x}=\underline{x}^*$)

Let us proove a simpler case;

(P) : $\min f(\underline{x})$, st $h(\underline{x})=0$

(D) : $$\max_{\lambda} \phi(\lambda) = \max_{\lambda}\left\{ \min_{\underset{\sim}{x}} \left[\underbrace{f(\underline{x})+\lambda h(\underline{x})}_{L(\underline{x},\lambda)} \right] \right\}$$

< Primal problem >

$$L(\underline{x},\lambda) \triangleq f(x)+\lambda h(x)$$

$$\rightarrow \left[\begin{array}{ll} \nabla f(\underline{x}^*) + \lambda^* \underline{\nabla} h(\underline{x}^*) = 0 & ① \\ h(\underline{x}^*) = 0 & ② \end{array} \right]$$

← optimality condition for P

Assumption: $L(\underline{x},\lambda)$ is assumed to be locally convex at $\underline{x}=\underline{x}^*$ for $\underline{x}\in \mathbb{R}^n$ (not just for $\underline{x} \in M$ tangent plane).

< Dual Problem >

Analysis : Let's solve $\max_{\lambda} \phi(\lambda)$

$$\phi(\lambda) = \min_{\underset{\sim}{x}} \left[\underbrace{f(\underline{x}) + \lambda h(\underline{x})}_{L(\underline{x},\lambda)} \right]$$

step 1: For $L(\underline{x},\lambda)$ to be minimized,

- $\frac{\partial L}{\partial \underset{\sim}{x}} = 0$; $\quad \frac{\partial f}{\partial x_i} + \lambda \frac{\partial h}{\partial x_i} = 0 \quad (i=1,\cdots,n)$ (a)

- Let $\underline{x}(\lambda)$ be the min point of $L(\underline{x},\lambda)$ for a given (unknown) λ;

Define
$$\phi(\lambda) = \min_{\underline{x}} L(\underline{x},\lambda)$$
$$= f(\underline{x}(\lambda)) + \lambda h(\underline{x}(\lambda)) \qquad (b)$$

Step 2: For $\phi(\lambda)$ to be minimized

$$0 = \frac{d\phi(\lambda)}{d\lambda} \underset{(b)}{=} \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \lambda} + h(\underline{x}(\lambda))$$
$$+ \lambda \sum_{i=1}^{n} \frac{\partial h}{\partial x_i} \frac{\partial x_i}{\partial \lambda}$$
$$= \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} + \lambda \frac{\partial h}{\partial x_i} \right) \frac{\partial x_i}{\partial \lambda} + h(\underline{x}(\lambda)) \qquad (c)$$

Using (a) $\Leftrightarrow$ $\partial f/\partial x_i + \lambda\, \partial h/\partial x_i = 0$,

(c) become:
$$0 = \frac{d\phi(\lambda)}{d\lambda} = h(x(\lambda)) \qquad (c)'$$

Thus, the optimality condition for (D)

$$\begin{cases} \underline{\nabla} f(\underline{x}(\lambda)) + \lambda\, \underline{\nabla} h(\underline{x}(\lambda)) = 0 & (a) \\ h(\underline{x}(\lambda)) = 0 & (c)' \end{cases}$$

Because (①, ③) : equivalent to (a), (c)',
(x̲*, λ*) will be either max or min
of φ(λ).

Question? $(\underline{x}^*, \lambda^*)$ = max or min of $\phi(\lambda)$?

Response: check the Hessian of $\phi(\lambda)$ wrt λ

Hessian of $\phi(\lambda)$ wrt λ:

- $$\frac{\partial^2\phi}{\partial\lambda^2} = \frac{\partial}{\partial\lambda}\left(\frac{\partial\phi}{\partial\lambda}\right) \underset{(c)'}{=} \frac{\partial}{\partial\lambda}[h(\underline{x}(\lambda)))$$
$$= \sum_{j=1}^{n} \frac{\partial h}{\partial x_j}\frac{\partial x_j}{\partial\lambda} = (\underset{\sim}{\nabla} h)^T \frac{\partial \underline{x}}{\partial\lambda} \qquad (d)$$

- To determine $\partial\underline{x}/\partial\lambda$;

  use $\frac{\partial}{\partial\lambda}[(a)]$

$$\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\frac{\partial x_j}{\partial \lambda} + \frac{\partial h}{\partial x_i} + \lambda \sum_{j=1}^{n}\frac{\partial^2 h}{\partial x_i\partial x_j}\frac{\partial x_j}{\partial\lambda} = 0$$

$$\rightarrow \left[\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i\partial x_j} + \lambda\sum_{j=1}^{n} \frac{\partial^2 h}{\partial x_i \partial x_j}\right]\frac{\partial x_j}{\partial\lambda} = -\frac{\partial h}{\partial x_i} \quad (i=1,\dots,n)$$

i.e.

$$\left[H_f + \lambda H_h\right]\frac{\partial \underline{x}}{\partial\lambda} = -\underline{\nabla} h$$

$$\therefore \quad \boxed{\frac{\partial \underline{x}}{\partial\lambda} = -\left[H_f+\lambda H_h\right]^{-1}\underline{\nabla}h} \qquad (e)$$

(e) → (d)

$$\frac{\partial^2\phi}{\partial\lambda^2} = -\underline{\nabla}h^T \underbrace{\left[H_f+\lambda H_f\right]}_{\substack{\text{postive-definite} \\ \text{by local convexity assumption} \\ \text{of } L(\underline{x},\lambda)}}^{-1}\underline{\nabla}h$$

<u>i.e</u>

$$\frac{\partial^2\phi}{\partial\lambda^2} \Rightarrow \text{Negative-definite} \qquad (f)$$

Therefore $(\underline{x}^*, \underline{\lambda}^*)$: maximize $\phi(\lambda)$

Remarks:

① Usually difficult to solve

$$\min_{\underset{\sim}{x}} f(\underset{\sim}{x}) + \lambda h(\underset{\sim}{x})$$

because λ is undetermined in this stage.

② However, easy to solve for a special class of problems known as "seperable problems"

③ Convexity requirement was used to convert (P) to (D)

⇒ · Applicable for special objective/constraint

· Thus, if $f(\underset{\sim}{x})$ and constraints equations are "approximated" by convex functions at every iteration step, it is okay!!

⑧

Example: $\min f(x,y) = -xy$

$\text{s.t } h(x,y) = (x-3)^2 + y^2 - 5 = 0$

Use Both (P) and (D) forms to solve this

Sol:

< primal >

Let $L(x,y;\lambda) = f(x,y) + \lambda h(x,y)$

$= -xy + \lambda[(x-3)^2 + (y^2-5)]$ ①

FONC $\frac{\partial f}{\partial x} = 0$; $-y + (2x-6)\lambda = 0$ ③

$\frac{\partial f}{\partial y} = 0$; $-x + 2y\lambda = 0$ ③

$\frac{\partial f}{\partial \lambda} = 0$ $(x-3)^2 + (y^2-5) = 0$ ④

Solving for (x,y,λ): $x^* = 4,\ y^* = 2,\ \lambda^* = 1$ ⑤

Check $H(x^*,\lambda^*) = H_f + \lambda^* H_h$

$$= \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} + (1)\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

↳ positive-definite for all $(x,y) \in R^2$ at (x^*,λ^*)

⇒ ∴ i) (x^*,y^*) of ⑤ : Local min of P

ii) Dual Method can be also used.

<Dual>

$$\max_{\lambda} \phi(\lambda) = \max_{\lambda} \underbrace{\min_{\underline{x}} L(\underline{x},\lambda)} \quad \text{---(a)}$$

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial x} = 0 ; \quad -y + (2x-6)\lambda = 0 \\ \frac{\partial L}{\partial y} = 0 ; \quad -x + 2y\lambda = 0 \end{array} \right.$$

same as ②, ③

Solving for $x(\lambda)$, and $y(\lambda)$;

$$x(\lambda) = \frac{12\lambda^2}{4\lambda^2-1} ; \quad y(\lambda) = \frac{6\lambda}{4\lambda^2-1} \quad (b, c)$$

$(\text{if } \lambda^2 \neq 1/4)$

$(b, c) \rightarrow \phi(\lambda)$

$$\phi(\lambda) = \frac{-80\lambda^5 + 4\lambda^3 + 4\lambda}{(4\lambda^2-1)^2} \quad (d)$$

To maximize $\phi(\lambda)$

$$\frac{d\phi}{d\lambda} = 0 \rightarrow \lambda = 1 \quad \left(\left.\frac{d^2\phi}{d\lambda^2}\right|_{\lambda=1} < 0 \right)$$

$\therefore$ $\phi(\lambda)$ is max at $\lambda^* = 1$

and $x^*(\lambda^*) = 4$; $y^*(\lambda^*) = 2$

Seperable Problems ?

Definition :

$$\boxed{\min \sum_{i=1}^{n} f_k(x_k)} \leftarrow \text{look at the form!}$$

$$\text{s.t.} \begin{cases} g_j(\underline{x}) = \boxed{\sum_{k=}^{n} g_{jk}(x_k)} \leq 0 & j=1,\cdots,m \\ h_i(\underline{x}) = \boxed{\sum_{k=1}^{n} h_{ik}(x_k)} = 0 & i=1,\cdots,l \end{cases}$$

Why separable problems are suitable for Dual formulation?

$$\phi(\underline{\lambda},\underline{\mu}) = \min\left\{ \sum_{k=1}^{n} \left[f(x_k) + \sum_{j=1}^{m} \mu_j\, g_{jK}(x_k) + \sum_{i=1}^{l} \lambda_i h_{ik}(x_k) \right] \right\}$$

$$= \sum_{k=1}^{n} \underbrace{\min_{x_k} \left[f_k(x_k) + \sum_{j=1}^{m} \mu_j\, g_{jK}(x_k) + \sum_{i=1}^{l} \lambda_i h_{ik}(x_k) \right]}_{\text{involves only one variable!!}}$$

$\Rightarrow$ Dual problems become decomposed into small subproblems.

$\boxed{\text{Example of Separable Problems}}$ (see Haftka, p355)

$$\left.\begin{aligned} \text{Min} \quad & f(\underset{\sim}{x}) = x_1^2 + x_2^2 + x_3^2 \\ \text{s.t} \quad & g_1(x) = 10 - x_1 - x_2 \le 0 \quad \swarrow \mu_1 \\ & g_2(x) = 8 - x_2 - 2x_3 \le 0 \quad \swarrow \mu_2 \end{aligned}\right\}$$

all convex and all seprable

== To solve by Dual Formulation ==

▨ Let

$$L(\underset{\sim}{x}, \underset{\sim}{\mu}) = (x_1^2 + x_2^2 + x_3^2) + \mu_1(10 - x_1 - x_2) + \mu_2(8 - x_2 - 2x_3)$$

$$= \underbrace{(10\mu_1 + 8\mu_2)}_{L_0} + \underbrace{(x_1^2 - \mu_1 x_1)}_{L_1(x_1)} + \underbrace{\left[x_2^2 - (\mu_1 + \mu_2)x_2\right]}_{L_2(x_2)} + \underbrace{\left[x_3^2 - 2\mu_2 x_3\right]}_{L_3(x_3)}$$

Thus

$$\min_{\underset{\sim}{x}} L(x, \underset{\sim}{\mu}) \Rightarrow \sum_k \min L_k(x_k)$$

① $\min\limits_{x_1} L_1(x_1) \rightarrow \dfrac{\partial L}{\partial x_1} = 0 \qquad 2x_1 - \mu_1 = 0$

$\Rightarrow x_1 = \mu_1/2$

② $\min\limits_{x_2} L_2(x_2) \rightarrow \dfrac{\partial L}{\partial x_2} = 0; \qquad 2x_2 - (\mu_1 + \mu_2) = 0$

$\Rightarrow x_2 = (\mu_1 + \mu_2)/2$

③ $\min_{x_3} L_3(x_3) \rightarrow \frac{\partial L}{\partial x_3} = 0;$ $\quad 2x_3 - 2\mu_2 = 0$

$\Rightarrow x_3 = \mu_2$

$\therefore \phi(\mu_1, \mu_2) = \min_{x} L(\underline{x}, \underline{\mu})$

$= L\left(x_1 = \frac{\mu_1}{2}, x_2 = \frac{\mu_1+\mu_2}{2}, x_3 = \mu_2; \mu_1, \mu_2\right)$

$= 0.5\mu_1^2 - 1.25\mu_2^2 - 0.5\mu_1\mu_2 + 10\mu_1 + 8\mu_2$

▨ To find $\max_{\substack{\mu_1 > 0 \\ \mu_2 \ge 0}} \phi(\mu_1, \mu_2) \leftarrow$ unconstrained problem !! (except side constraints)

$0 = \frac{\partial \phi}{\partial \mu_1} = -\mu_1 - 0.5\mu_2 + 10 = 0$

$0 = \frac{\partial \phi}{\partial \mu_2} = -2.5\mu_1 - 0.5\mu_1 + 8 = 0$

$\therefore \mu_1^* = 28/3, \quad \mu_2^* = 4/3,$ ④

(satisfies $\mu_i \ge 0$)

$\phi(\mu_1^*, \mu_2^*) = 52$

$\boxed{4}$ Check the negative-definiteness for maximization

$$\left.\frac{\partial^2 \phi}{\partial \mu_i \partial \mu_j}\right|_{\text{at optimal}} = \begin{bmatrix} -1 & -0.5 \\ -0.5 & -2.5 \end{bmatrix}$$

$$\left(\text{Recall : } \quad (-1)^i \det A^{(i)} > 0 \text{ for all } i \quad \rightarrow \quad A \text{ is negative-definite} \right)$$

$$(-1)^1 (-1) > 0, \qquad (-1)^2 \begin{vmatrix} -1 & -0.5 \\ -0.5 & -2.5 \end{vmatrix} > 0$$

$\therefore$ $\underline{(\mu_1^*, \mu_2^*) \text{ of } \textcircled{4} \text{ maximizes}}$ $\phi(\mu_1, \mu_2)$

$$\therefore \left. \begin{array}{l} x_1^* \underset{\textcircled{1},\textcircled{4}}{=} 14/3, \quad x_2^* \underset{\textcircled{2},\textcircled{4}}{=} 16/3, \quad x_3^* \underset{\textcircled{3},\textcircled{4}}{=} 4/3 \\ f^{\min}(x^*) = 52 \end{array} \right] \rightarrow \begin{array}{c} \text{minimize} \\ f(\underline{x}) \end{array}$$

<u>Check</u> :

$$\underbrace{\min f(\underline{x})}_{\substack{\shortparallel \\ 52}} = \max_{\mu \geq 0} \underbrace{\phi(\mu_1, \mu_2)}_{52}$$

Remark: Convex Approximation in Separable form
$\Rightarrow$ Can use Dual Method

(An efficient agorithm : Method of Moving Asymptote by Svanberg (1987).)

Svanberg (1987) — Method of Moving Asymptote (MMA)

Given $f(\underline{x})$ (and $\underline{g}(\underline{x}), h(\underline{x})$)

$\swarrow$ Kth-iteration

$$\Rightarrow \text{Approximate } \underset{\substack{\uparrow \\ \text{convex ftn}}}{f_c^{(k)}}(\underline{x}) = r^k + \sum_{j=1}^{n} \left[\frac{p_j^{(k)}}{U_j^{(k)} - x_j} + \frac{q_j^{(k)}}{x_j - L_j^{(k)}} \right]$$

$\nearrow$ reciprocal Approx

with

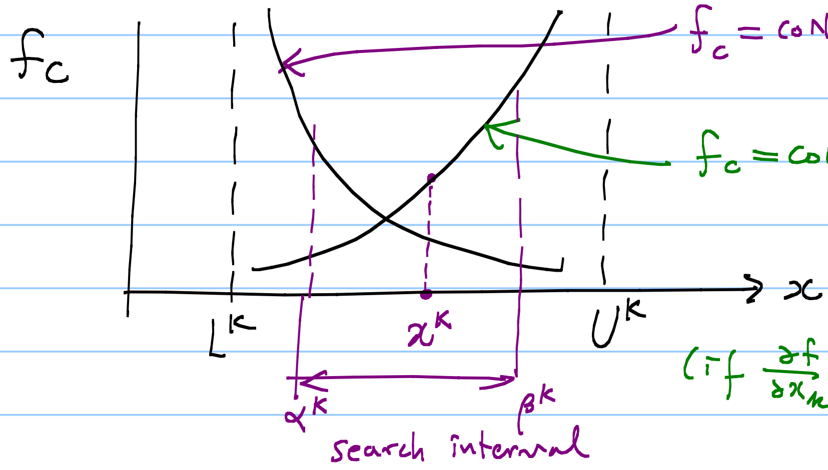
$$p_j^{(k)} = \begin{cases} \left(U_j^{(k)} - x_j^{(k)} \right)^2 \left. \frac{\partial f}{\partial x_j} \right|_{\underline{x}^k} & \text{if } \left. \frac{\partial f}{\partial x_j} \right|_{\underline{x}^k} > 0 \\ 0 & \text{else} \end{cases}$$

$$q_j^{(k)} = \begin{cases} 0 & \text{else} \\ -\left(x_j^{(k)} - L_j^{(k)} \right)^2 \left. \frac{\partial f}{\partial x_j} \right|_{\underline{x}^k} & \text{if } \left. \frac{\partial f}{\partial x_j} \right|_{\underline{x}^k} < 0 \end{cases}$$

$$r^{(k)} = f(\underline{x}^{(k)}) - \sum_{j=1}^{n}\left[\frac{p_j^{(k)}}{U_j^{(k)}-x_j^{(k)}} + \frac{q_j^{(k)}}{x_j^{(k)}-L_j^{(k)}}\right]$$

(Some rules to determine $U_j^{(k)}$ and $L_j^{(k)}$
↑ upper limit ↑ Lower limit)

Consider Approx. for Single-Variable Case



$f_c = \text{CONST} + \dfrac{\text{CONST}}{x-L^k}$ if $\dfrac{\partial f}{\partial x_k} < 0$

$f_c = \text{CONST} + \dfrac{\text{CONST}}{U^k-x}$ if $\dfrac{\partial f}{\partial x_k} < 0$

(if $\dfrac{\partial f}{\partial x_k} = 0$, $f = \text{CONST}$)

<Observation>

① $\dfrac{\partial^2 f_c}{\partial x^2} \geq 0$ (for any f_c approximated above)
⇒ Convex Approximation

② $\underbrace{L^k < \alpha^k < x^k < \beta^k}_{\text{"move limit"}} < U^k$

Typical Selections of L^k, U^k, α^k, β^k $(k \geq 0)$

- $$\begin{cases} L^{(k)} = x^{(k)} - S^{(k)} \Delta L^{(k-1)} \\ U^{(k)} = x^{(k)} + s^{(k)} \Delta U^{(k-1)} \end{cases}$$

with

$$\Delta L^{(k)} = x^{(k)} - L^{(k)} \; ; \; \Delta U^{(k)} = U^{(k)} - x^{(k)}$$

$$\left(\Delta L^{(-1)} = \Delta U^{(-1)} = x^{upper} - x^{Lower} = \Delta x \right)$$

- $$\begin{cases} \alpha^{(k)} = 0.9\, L^{(k)} + 0.1\, x^{(k)} \\ \beta^{(k)} = 0.9\, U^{(k)} + 0.1\, x^{(k)} \end{cases}$$

($s^{(k)}$ ~ selected by some rule , simplest: $s^{(k)} = 1$)