

10/8/13

1

DETOUR: ~~A~~ CONSTRAINED OPTIMIZATION / LAGRANGIAN DUALITY

[NOT GOING TO BE ON ANY EXAM]

① Method of Lagrangian Multipliers for Equality-Constrained Optimization

Canonical Problem: $\min_{w \in \mathbb{R}^d} f(w)$

$$h(w) = 0$$

Generalization: can have multiple constraints

$$\left. \begin{array}{l} h_1(w) = 0 \\ h_2(w) = 0 \\ \vdots \end{array} \right\} \vec{h}(w) = \vec{0}$$

Method of Lagrangian Mults. (MLM):

$$\begin{array}{l} \min_w f(w) \\ h_1(w) = 0 \\ \vdots \\ h_m(w) = 0 \end{array} \quad \left. \begin{array}{l} \beta_1 \\ \vdots \\ \beta_m \end{array} \right\} \leftarrow \text{Lagrangian multipliers}$$

Lagrangian: $L(w, \beta_1, \dots, \beta_n) = f(w) + \beta_1 h_1(w) + \beta_2 h_2(w) + \dots + \beta_n h_n(w)$

Now $\left. \begin{array}{l} \frac{\partial L}{\partial w} = 0 \\ \frac{\partial L}{\partial \beta_1} = 0 \\ \frac{\partial L}{\partial \beta_n} = 0 \end{array} \right\} \begin{array}{l} \text{System of } n+d \text{ equations} \\ \text{in } n+d \text{ unknowns} \\ \text{solve for } \vec{w}^*, \vec{\beta}^* \end{array}$

(may be many) ↓

Note 1: MLM gives critical points. Then need to check which one is global min (if any)

Note 2: $\frac{\partial L}{\partial \beta_i} = 0 \Rightarrow h_i(w) = 0$ } Original constraint

So we are basically solving
find w s.t. $\frac{\partial L}{\partial w} = 0$

$$h_i(w) = 0 \quad \forall i$$

Note 3: For 1 constraint

$$\frac{\partial L}{\partial w} = 0 \Rightarrow \frac{\partial}{\partial w} [f(w) + \beta h(w)] = 0$$

$$\Rightarrow \left[\frac{\partial f}{\partial w} = -\beta \frac{\partial h}{\partial w} \right]$$

This has a geometric intuition in terms of tangents. See figure on slides.

2

② MLM Example

$$\min_{W \in \mathbb{R}^d} W^T A W$$

$$W^T W = 1$$

$$L(W, \beta) = W^T A W + \beta(1 - W^T W)$$

$$\frac{\partial L}{\partial W} = W^T A - \beta W^T = 0$$

$$\Rightarrow A W = \beta W$$

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow$$

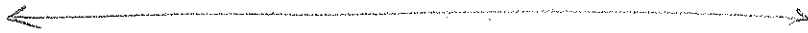
$$W^T W = 1$$

Eigensystem

→ every eigenvector, value is a critical point

$$f(W) = W^T A W = W^T (\beta W) = \beta$$

→ W^* = eigenvector with smallest eigenvalue.



③ Lagrangian Duality (for general inequality/equality constrained programs)

$$\min_{W \in \mathbb{R}^d} f(W)$$

Lagrangian Multiplier

$$\left\{ \begin{array}{l} \alpha \geq 0 \\ \beta \end{array} \right.$$

$$\left\{ \begin{array}{l} g(W) \leq 0 \\ h(W) = 0 \end{array} \right.$$

if you have a $q(W) \geq 0$ constraint, we can just define $g(W) = -q(W)$

Lagrangian $L(w, \alpha, \beta) = f(w) + \alpha g(w) + \beta h(w)$

Game-Theoretic Interpretation of Lagrangian Duality

Primal Player: Controls w

Dual Player: Controls α, β

L represents payoff from Primal player to Dual Player

Primal Player wants to	\min_w	$L(w, \alpha, \beta)$
Dual " " " "	$\max_{\alpha \geq 0, \beta}$	$L(w, \alpha, \beta)$

2 choices

$$\min_w \max_{\substack{\alpha \geq 0 \\ \beta}} L(w, \alpha, \beta)$$

Dual-Player Goes first
Primal-Player Goes Next

Primal Program

$$\min_w l_p(w)$$

where $l_p(w) = \max_{\substack{\alpha \geq 0 \\ \beta}} L(w, \alpha, \beta)$

$$\max_{\substack{\alpha \geq 0 \\ \beta}} \min_w L(w, \alpha, \beta)$$

Primal Player Goes first
Dual " " Next

Dual Program

$$\max_{\alpha, \beta} l_d(\alpha, \beta)$$

$\alpha \geq 0$

where $l_d(\alpha, \beta) = \min_w L(w, \alpha, \beta)$

(3)

Now
$$l_p(w) = \max_{\substack{\alpha \geq 0 \\ \beta}} L(w, \alpha, \beta)$$

[∴ dual-player will set $\alpha=0$] =
$$\begin{cases} f(w) & \text{if } g(w) \leq 0, h(w) = 0 \\ \infty & \text{else} \end{cases}$$

[∴ dual-player will set $\alpha = +\infty$ and/or $\beta = \pm\infty$]

So Primal - Program:

$$\min_w l_p(w) \equiv \min_w \begin{cases} f(w) \\ g(w) \leq 0 \\ h(w) = 0 \end{cases}$$

Basically, by going first the dual player extracts +∞ penalty from Primal player for violating constraints. Thus Primal player always satisfies constraints.

Dual Program:

$$\max_{\substack{\alpha \geq 0 \\ \beta}} l_D(\alpha, \beta)$$

$$l_D(\alpha, \beta) = \min_w [f(w) + \alpha g(w) + \beta h(w)]$$

Because Primal player goes first, he can violate constraints as long as the penalty for violation (α, β) is NOT ∞. Dual player may not set ∞ penalties because he may be able to maximize payoff at some NON-INF values.

④ Relationship between primal & dual values

Let $p^* = \min F(w)$

primal OPT value

$$g(w) \leq 0$$

$$h(w) = 0$$

$d^* = \max_b (\alpha, \beta)$

Dual OPT value

$$p^* \geq d^* \quad [\text{Weak Duality}]$$

Always true, without any assumptions on $f, g, h, \text{ etc.}$

Why?

$$p^* = \min \max L$$

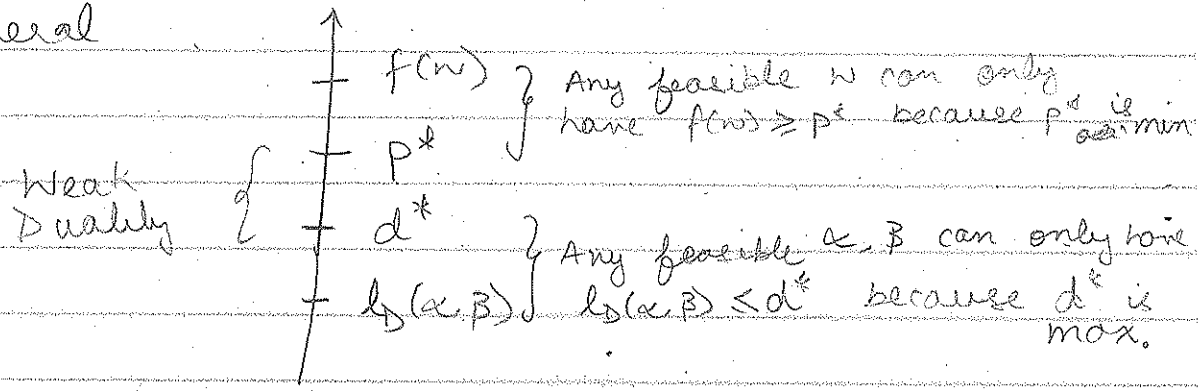
$$d^* = \max \min L$$

$$\min \max \text{ Something} \geq \max \min \text{ Something}$$

Example

	w	α	$\max \alpha$
	0	1	1
	1	0	1
$\min w$	[0, 0]		

In general



So $lb(\alpha, \beta)$ for ANY $\alpha \geq 0, \beta$ provides a lower-bound on p^* & $f(w)$ for ANY w .

④ KKT Conditions & Strong Duality

Let $w^* = \arg \min \begin{cases} f(w) \\ g(w) \leq 0 \\ h(w) = 0 \end{cases}$ } Technical point: $\arg \min$ may not exist [Ignore in ML class]

Let $\alpha^*, \beta^* = \arg \max \begin{cases} lb(\alpha, \beta) \\ \alpha \geq 0 \end{cases}$

Karush-Kuhn-Tucker (KKT) Conditions

