Chapter 22　Lagrangian Duality

1. Saddle Point Problem

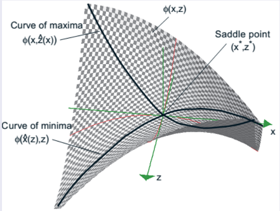
2. Duality

Saddle Point Problem

Let $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{Z} \subset \mathbb{R}^m$, and $\phi : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$.

$(x^*, z^*) \in \mathcal{X} \times \mathcal{Z}$ is a saddle point of ϕ if
$\phi(x^*, z) \leq \phi(x^*, z^*) \leq \phi(x, z^*)$, $\forall x \in \mathcal{X}, z \in \mathcal{Z}$.
Or equivalently,

$$\begin{cases} x^* = \arg\min_{x \in \mathcal{X}} \phi(x, z^*), \\ z^* = \arg\max_{z \in \mathcal{Z}} \phi(x^*, z). \end{cases}$$



Definition (minimax problem)

(P) $w^* = \min_{x \in \mathcal{X}} \max_{z \in \mathcal{Z}} \phi(x, z)$ (D) $q^* = \max_{z \in \mathcal{Z}} \min_{x \in \mathcal{X}} \phi(x, z)$

★ worst-case design, zero sum game theory, equilibrium, ...

Theorem (minimax inequality)

Let $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{Z} \subset \mathbb{R}^m$, and $\phi : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$. Then, $w^* \geq q^*$.

Saddle Point Problem

minimax inequality: $w^* = \min_{x\in\mathcal{X}} \max_{z\in\mathcal{Z}} \phi(x,z) \geq \max_{z\in\mathcal{Z}} \min_{x\in\mathcal{X}} \phi(x,z) = q^*$

$$\begin{array}{c|cccccc} x & & & & & & \\ & 3 & 6 & 7 & 2 & \boxed{7} & \\ & 8 & 3 & 6 & 5 & \boxed{8} & \\ & 5 & 13 & 28 & 4 & \boxed{28} & \\ & 7 & 0 & 2 & 1 & \boxed{7} & z \\ \hline \end{array}$$

$w^* = \min_{x\in\mathcal{X}} \max_{z\in\mathcal{Z}} \phi(x,z) = 7$

$$\begin{array}{c|ccccc} x & & & & & \\ & 3 & 6 & 7 & 2 & \\ & 8 & 3 & 6 & 5 & \\ & 5 & 13 & 28 & 4 & \\ & 7 & 0 & 2 & 1 & z \\ \hline & \boxed{3} & \boxed{0} & \boxed{2} & \boxed{1} & \end{array}$$

$q^* = \max_{z\in\mathcal{Z}} \min_{x\in\mathcal{X}} \phi(x,z) = 3$

proof. let $p(x) = \max_{z\in\mathcal{Z}} \phi(x,z)$

$\therefore\ p(x) \geq \phi(x,z),\ \forall z \in \mathcal{Z}.$

$\therefore\ \min_{x\in\mathcal{X}} p(x) \geq \min_{x\in\mathcal{X}} \phi(x,z),\ \forall z \in \mathcal{Z},$

i.e., $w^* \geq \min_{x\in\mathcal{X}} \phi(x,z),\ \forall z \in \mathcal{Z}.$

$\because z \in \mathcal{Z}$ is arbitrary,

$\therefore w^* \geq \max_{z\in\mathcal{Z}} \min_{x\in\mathcal{X}} \phi(x,z) = q^*$

proof. let $q(z) = \min_{x\in\mathcal{X}} \phi(x,z)$

$\therefore\ q(z) \leq \phi(x,z),\ \forall x \in \mathcal{X}.$

$\therefore\ \max_{z\in\mathcal{Z}} q(z) \leq \max_{z\in\mathcal{Z}} \phi(x,z),\ \forall x \in \mathcal{X},$

i.e., $q^* \leq \max_{z\in\mathcal{Z}} \phi(x,z),\ \forall x \in \mathcal{X}.$

$\because x \in \mathcal{X}$ is arbitrary,

$\therefore q^* \leq \min_{x\in\mathcal{X}} \max_{z\in\mathcal{Z}} \phi(x,z) = w^*$

Saddle Point Problem

Theorem

$(x^*, z^*) \in \mathcal{X} \times \mathcal{Z}$ is a saddle point iff the minimax equality holds and $x^* = \arg\min_{x\in\mathcal{X}}\max_{z\in\mathcal{Z}} \phi(x,z)$, $z^* = \arg\max_{z\in\mathcal{Z}}\min_{x\in\mathcal{X}} \phi(x,z)$.

optimization problem: $\min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x})$

$$\begin{aligned}\Omega &= \{\boldsymbol{x} \in \mathbb{R}^n \mid h_i(\boldsymbol{x}) = 0, i = 1, \ldots, m; g_j(\boldsymbol{x}) \le 0, j = 1, \ldots, l\}\\ &= \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}(\boldsymbol{x}) = 0, \boldsymbol{g}(\boldsymbol{x}) \le 0\},\end{aligned}$$

Lagrangian function: $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l_+ \to \mathbb{R}$,

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{h}(\boldsymbol{x}) + \boldsymbol{\mu}^\top \boldsymbol{g}(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^{l} \mu_j g_j(\boldsymbol{x}).$$

By saddle point theory, what are the

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \underbrace{\max_{\boldsymbol{\lambda}\in\mathbb{R}^m, \boldsymbol{\mu}\in\mathbb{R}^l_+} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})}_{p(\boldsymbol{x})} \quad \text{and} \quad \max_{\boldsymbol{\lambda}\in\mathbb{R}^m, \boldsymbol{\mu}\in\mathbb{R}^l_+} \underbrace{\min_{\boldsymbol{x}\in\mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})}_{q(\boldsymbol{\lambda},\boldsymbol{\mu})}$$

Duality

- $p(\boldsymbol{x}) = \max\limits_{\boldsymbol{\lambda}\in\mathbb{R}^m, \boldsymbol{\mu}\in\mathbb{R}^l_+} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \xRightarrow{\text{why?}} p(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}), & \text{if } \boldsymbol{h}(\boldsymbol{x}) = 0, \boldsymbol{g}(\boldsymbol{x}) \le 0 \\ +\infty, & \text{otherwise} \end{cases}$

  $\therefore \min\limits_{\boldsymbol{x}\in\mathbb{R}^n} \max\limits_{\boldsymbol{\lambda}\in\mathbb{R}^m, \boldsymbol{\mu}\in\mathbb{R}^l_+} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \iff \min\limits_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x})$ [primal problem]
- $q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min\limits_{\boldsymbol{x}\in\mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$

  $\therefore \max\limits_{\boldsymbol{\lambda}\in\mathbb{R}^m, \boldsymbol{\mu}\in\mathbb{R}^l_+} \min\limits_{\boldsymbol{x}\in\mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \iff \max\limits_{\boldsymbol{\lambda}\in\mathbb{R}^m, \boldsymbol{\mu}\in\mathbb{R}^l_+} q(\boldsymbol{\lambda}, \boldsymbol{\mu})$ [Dual problem]

(P) $\min_{\boldsymbol{x}\in\mathbb{R}^2} f(\boldsymbol{x}) = \frac{1}{2}(x_1^2 + x_2^2)$, s.t. $g(\boldsymbol{x}) = 1 - x_1 \le 0$

$\because$ Lagrangian function $L : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ is defined by
$L(\boldsymbol{x}, \mu) = f(\boldsymbol{x}) + \mu g(\boldsymbol{x}) = \frac{1}{2}(x_1^2 + x_2^2) + \mu(1 - x_1)$.

$\therefore q(\mu) = \min\limits_{\boldsymbol{x}\in\mathbb{R}^2} L(\boldsymbol{x}, \mu) = \min\limits_{\boldsymbol{x}\in\mathbb{R}^2} \left\{\frac{x_1^2+x_2^2}{2} + \mu(1 - x_1)\right\} \xRightarrow[rule]{Fermat} (x_1^*, x_2^*) = (\mu, 0)$.

$\therefore q(\mu) = -\frac{\mu^2}{2} + \mu$ with $\mu \in \mathbb{R}_+$.

(D) problem: $\max\limits_{\mu\in\mathbb{R}_+} q(\mu) = -\frac{\mu^2}{2} + \mu \xRightarrow{KKT} \mu^* = 1 \Rightarrow q^* = q(\mu^*) = \frac{1}{2}$.

$\therefore \boldsymbol{x}^* = \arg\min\limits_{\boldsymbol{x}\in\mathbb{R}^2} L(\boldsymbol{x}, \mu^*) \Rightarrow \boldsymbol{x}^* = (1, 0) \Rightarrow w^* = f(\boldsymbol{x}^*) = \frac{1}{2} \Longrightarrow w^* = q^*$.

Duality

Assmue that $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{Q} \succ 0$, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $\boldsymbol{c} \in \mathbb{R}^n$.

(P) $\min f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x}$, s.t. $\boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}$

$\because$ Lagrangian function $L : \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}$ is defined by
$L(\boldsymbol{x}, \boldsymbol{\mu}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\mu}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$.

$\therefore q(\boldsymbol{\mu}) = \min_{\boldsymbol{x}\in\mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{\mu}) = \min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{x}\right\} - \boldsymbol{\mu}^\top \boldsymbol{b}$.

By optimal condition, $\boldsymbol{x}^* = -\boldsymbol{Q}^{-1}(\boldsymbol{c} + \boldsymbol{A}^\top \boldsymbol{\mu})$.

$\therefore\ q(\boldsymbol{\mu}) = -\frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top \boldsymbol{\mu} - \boldsymbol{\mu}^\top(\boldsymbol{b} + \boldsymbol{A}\boldsymbol{Q}^{-1}c) - \frac{1}{2}\boldsymbol{c}^\top \boldsymbol{Q}^{-1}\boldsymbol{c}$.

(D) problem: $\max_{\boldsymbol{\mu}\in\mathbb{R}^m_+} q(\boldsymbol{\mu}) = \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{P}\boldsymbol{\mu} + \boldsymbol{d}^\top \boldsymbol{\mu}$,

where $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top$ and $\boldsymbol{d} = \boldsymbol{b} + \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{c}$.

Duality

Example ((P) $\min f(\boldsymbol{x}) = \boldsymbol{c}^\top \boldsymbol{x}$, s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, $\boldsymbol{x} \geq 0$)

$\because$ Lagrangian function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n_+ \to \mathbb{R}$ is defined by
$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{c}^\top \boldsymbol{x} - \boldsymbol{\lambda}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{\mu}^\top \boldsymbol{x}$.

$$\therefore q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ (\boldsymbol{c} - \boldsymbol{A}^\top \boldsymbol{\lambda} - \boldsymbol{\mu})^\top \boldsymbol{x} \right\} + \boldsymbol{\lambda}^\top \boldsymbol{b}.$$

$$\therefore \; q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} \boldsymbol{\lambda}^\top \boldsymbol{b}, & \text{if } \boldsymbol{c} - \boldsymbol{A}^\top \boldsymbol{\lambda} - \boldsymbol{\mu} = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

(D) problem: $\displaystyle\max_{\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}^n_+} q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{\lambda}^\top \boldsymbol{b}$, s.t. $\boldsymbol{c} - \boldsymbol{A}^\top \boldsymbol{\lambda} = \boldsymbol{\mu}$

i.e., $\max \boldsymbol{\lambda}^\top \boldsymbol{b}$, s.t. $\boldsymbol{A}^\top \boldsymbol{\lambda} \leq \boldsymbol{c}$.

★ Duality can reformulate an optimization problem in another format, even an unconstrained one (why?). It provides another way to seek the minimizers.

★ Some dual methods can be developed by probing the dual problem.

Theorem (weak/strong duality)

For any feasible points $\bar{\boldsymbol{x}} \in \Omega$, $(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) \in \mathbb{R}^m \times \mathbb{R}^l_+$, then $f(\bar{\boldsymbol{x}}) \geq q(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}})$.

Duality

★ Dual gap: $\mathrm{gap}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}) = f(\bar{\boldsymbol{x}}) - q(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\mu}}))$. Strong duality: no dual gap.

Theorem (weak/strong duality)

If strong duality holds, then both (P) and (D) have the optima.

Theorem (when the strong duality holds?)

Lagrangian function admits the saddle points iff $\boldsymbol{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ are optima of (P) and (D), respectively, and stong duality holds.

convex optimization: $\min\limits_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x})$ with f as convex function and Ω as convex set
stationary point $\Leftrightarrow$ global optimum $\Leftrightarrow$ local optimum $\rightleftharpoons$ KKT point $\Leftrightarrow$ saddle point

★ "$\rightharpoondown$" in above relation satisfies with Slater condition.

Theorem (when Slater condition holds?)

- *Slater condition holds for all convex optimization with linear constraints.*
- *(P) is convex optimization, Ω has the relative interior point.*

★ strong duality may hold for nonconvex optimization (uneasy to analysis).

Homework: write the dual of the following problems

(1) $\min\limits_{\boldsymbol{x}\in\mathbb{R}^2} f(\boldsymbol{x}) = x_1^2 + x_2^2$, s.t. $x_1 + x_2 - 4 \geq 0,\ x_1 \geq 0,\ x_2 \geq 0.$

(2) $\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x}$, s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.$

(3) $\min\limits_{\boldsymbol{x}\in\mathbb{R}^2} f(\boldsymbol{x}) = |x_1| + x_2$, s.t. $x_1 \leq 0,\ x_2 \geq 0.$

(4) $\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \tau\|\boldsymbol{x}\|_1.$

(5) $\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \|\boldsymbol{x}\|_p$, s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.$