LECTURE 10: CONSTRAINED OPTIMIZATION – LAGRANGIAN DUAL PROBLEM

1. Lagrangian dual problem
2. Duality gap
3. Saddle point solution
4. Lagrangian dual vs. conjugate dual

Lagrangian dual problem

Primal Problem:

$$(P)\quad \begin{array}{lll} \text{minimize} & f(x) & \\ \text{s.t.} & h(x) = 0 & \leftarrow \quad \lambda \in E^m \\ & g(x) \leq 0 & \leftarrow \quad \mu \in E^p_+ \\ & x \in X & \end{array}$$

Lagrangian Dual Problem:

$$(LD)\quad \begin{array}{ll} \text{maximize} & \phi(\lambda, \mu) \\ \text{s.t.} & \mu \geq 0 \end{array}$$

where $\phi(\lambda, \mu) = \inf_{x \in X} [f(x) + \lambda^T h(x) + \mu^T g(x)]$

Property 1 – weak duality

Let $\bar{x}$ be a primal feasible solution and $(\bar{\lambda}, \bar{\mu})$ be a dual feasible solution.

Then

$$\begin{aligned}
\phi(\bar{\lambda}, \bar{\mu}) &= \inf_{x \in X} \{ f(x) + \bar{\lambda}^T h(x) + \bar{\mu}^T g(x) \} \\
&\leq f(\bar{x}) + \underbrace{\bar{\lambda}^T h(\bar{x})}_{=0} + \underbrace{\bar{\mu}^T g(\bar{x})}_{\leq 0} \\
&\leq f(\bar{x}).
\end{aligned}$$

Weak duality theorem

Theorem(Weak Duality Theorem):

Let $\bar{x}$ be primal feasible and $(\bar{\lambda}, \bar{\mu})$ be dual feasible. Then,

$$\phi(\bar{\lambda}, \bar{\mu}) \leq f(\bar{x}).$$

Corollary 1:

$$\inf_{x \in \mathscr{F}} f(x) \geq \sup_{(\lambda,\mu) \in \mathscr{D}} \phi(\lambda, \mu)$$

where $\mathscr{F} = \{x \in X \mid g(x) \leq 0, \text{ and } h(x) = 0\}$,

$$\mathscr{D} = \{(\lambda, \mu) \mid \lambda \in E^m, \mu \in E_+^p\}.$$

Corollary 2:

Let $\bar{x}$ be primal feasible and $(\bar{\lambda}, \bar{\mu})$ be dual feasible. If $f(\bar{x}) = \phi(\bar{\lambda}, \bar{\mu})$, then $\bar{x}$ solves (P) and $(\bar{\lambda}, \bar{\mu})$ solves (LD).

Corollary 3:

If $\sup_{(\lambda,\mu) \in \mathscr{D}} \phi(\lambda, \mu) = +\infty$, then (P) is infeasible.

Corollary 4:

If $\inf_{x \in \mathscr{F}} f(x) = -\infty$, then $\phi(\lambda, \mu) = -\infty$ for any $\mu \geq 0$.

Property 2 – concavity and subgradient

Let $X \in E^n$ be nonempty and compact, f, g, h be continuous. Then,

(a) $\phi(\lambda, \mu) = \inf_{x \in X} \{f(x) + \lambda^T h(x) + \mu^T g(x)\}$ is well defined on $E^m \times E^p_+$.

(b) $\phi(\lambda, \mu)$ is concave over $E^m \times E^p_+$.

Proof: Given any $\omega \in (0,1)$,

$$\phi(\omega\bar{\lambda} + (1-\omega)\bar{\bar{\lambda}}, \omega\bar{\mu} + (1-\omega)\bar{\bar{\mu}})$$

$$\geq \omega\phi(\bar{\lambda}, \bar{\mu}) + (1-\omega)\phi(\bar{\bar{\lambda}}, \bar{\bar{\mu}}).$$

(c) Given any $(\bar{\lambda}, \bar{\mu}) \in E^m \times E^p_+$, define

$$X(\bar{\lambda}, \bar{\mu}) \triangleq \{\bar{x} \in X \mid \bar{x} \text{ minimizes } f(x) + \bar{\lambda}^T h(x) + \bar{\mu}^T g(x) \text{ over } X\}.$$

Then $X(\bar{\lambda}, \bar{\mu}) \neq \phi$ in our setting.

(d) For any $\bar{x} \in X(\bar{\lambda}, \bar{\mu})$,

$$\begin{aligned}
\phi(\lambda, \mu) &= \inf_{x \in X} \{f(x) + \lambda^T h(x) + \mu^T g(x)\} \\
&\leq f(\bar{x}) + \lambda^T h(\bar{x}) + \mu^T g(\bar{x}) \\
&= \underline{f(\bar{x})} + (\lambda - \bar{\lambda})^T h(\bar{x}) + (\mu - \bar{\mu})^T g(\bar{x}) \\
&\quad \underline{+\bar{\lambda}^T h(\bar{x}) + \bar{\mu}^T g(\bar{x})} \\
&= \underline{\phi(\bar{\lambda}, \bar{\mu})} + (\lambda - \bar{\lambda})^T h(\bar{x}) + (\mu - \bar{\mu})^T g(\bar{x}).
\end{aligned}$$

$\Rightarrow \begin{pmatrix} h(\bar{x}) \\ g(\bar{x}) \end{pmatrix}$ is a subgradient of ϕ at $(\bar{\lambda}, \bar{\mu})$.

(e) If $X(\bar{\lambda}, \bar{\mu})$ is singleton, and $\bar{x} \in X(\bar{\lambda}, \bar{\mu})$, then ϕ is differentiable at $(\bar{\lambda}, \bar{\mu})$ and

$$\nabla\phi(\bar{\lambda}, \bar{\mu}) = \begin{pmatrix} h(\bar{x}) \\ g(\bar{x}) \end{pmatrix}.$$

Property 3 – duality gap

- Duality gap may exist

Example 1:

$$\begin{aligned} \text{Minimize} \quad & f(x) = x^3 \\ \text{s. t.} \quad & h(x) = x - 1 = 0 \\ & x \in E^1. \end{aligned}$$

(a) f is not convex.

(b) $x^* = 1$ and $v^* = f(x^*) = 1$.

(c) $\phi(\lambda) = \inf_{x \in R} \{x^3 + \lambda(x-1)\}$

$$= \inf_{x \in R} \{x^3 + \lambda x - \lambda\}$$

$$= \begin{cases} -\infty\,, & \lambda > 0 \\ -\infty\,, & \lambda = 0 \\ -\infty\,, & \lambda < 0\,. \end{cases}$$

(d) $\phi(\lambda^*) = -\infty \neq f(x^*) = 1$.

(e) Can you check the local behavior of $\phi(\lambda)$ around $x^* = 1$ and $\lambda^* = -3$?

Example of duality gap

Example 2 (Bazaraa/Sherali/Shetty p. 205-206)

$$(P)\quad \begin{array}{ll} \text{Minimize} & f(x) = -2x_1 + x_2 \\ \text{s. t.} & h(x) = x_1 + x_2 - 3 = 0 \end{array}$$

$$x \in X = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$$

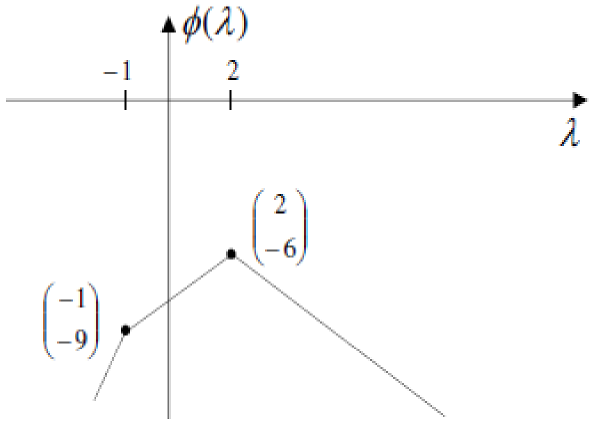
(a) X is compact, but not convex.

(b) Only $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ are feasible.

(c) $x^* = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ with $v^* = f(x^*) = -3$.

(d) $\phi(\lambda) = \min_{x \in X} \{-2x_1 + x_2 + \lambda(x_1 + x_2 - 3)\}$

$$= \begin{cases} -4 + 5\lambda\,, & \text{if} \quad \lambda \leq -1 \\ -8 + \lambda\,, & \text{if} \quad -1 \leq \lambda \leq 2 \\ -3\lambda\,, & \text{if} \quad \lambda \geq 2. \end{cases}$$



(e) $\lambda^* = 2$ with $\phi(\lambda^*) = -6 \neq -3 = f(x^*)$!!

Property 4 – strong duality

Duality gap vanishes only under proper conditions — Strong Duality Theorem

- <u>Theorem</u>: (Bazaraa/Sherali/Shetty p.208)

  Assume that

  (i) $X \neq \emptyset$ and is convex;

  (ii) f, g are convex and h is affine;

  (iii) (CQ) There exists $\bar{x} \in X$ such that

  (a) $g(\bar{x}) < 0$,

  (b) $h(\bar{x}) = 0$,

  (c) $0 \in \text{int}\left[h(X) \triangleq \{h(x) | x \in X\}\right]$.

Then,

$$\inf_{x \in \mathscr{F}} f(x) = \sup_{(\lambda,\mu) \in \mathscr{D}} \phi(\lambda, \mu).$$

Moreover, if the inf is finite, then $\sup_{(\lambda,\mu) \in \mathscr{D}} \phi(\lambda, \mu)$ is achieved at an $(\bar{\lambda}, \bar{\mu})$ with $\bar{\mu} \geq 0$.

If the inf is achieved at $\bar{x}$, then $\bar{\mu}^T g(\bar{x}) = 0$.

Geometric interpretation of LD

Consider a case with only one inequality constraint:

$$(P)\quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_1(x) \le 0 \\ & x \in X \end{array} \qquad \begin{array}{ll} \max & \phi(\mu) \\ \text{s.t.} & \mu \ge 0 \\ & \phi(\mu) = \inf_{x \in X} \{f(x) + \mu g_1(x)\} \end{array} \quad (LD)$$

Let

$$G \triangleq \{(y, z) | y = g_1(x), z = f(x) \text{ for some } x \in X\}.$$

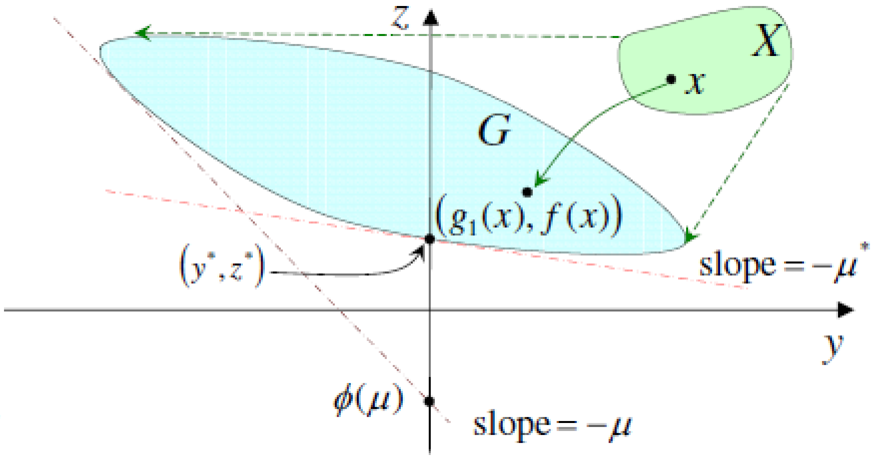
1. (P) says that "on the (y, z) plane, we are looking for a point in G with $y \le 0$ and a minimum ordinate."

2. $\phi(\mu) = \inf_{x \in X} \{\underbrace{f(x) + \mu g_1(x)}_{z + \mu y}\}$

   Note that the contour of

   $$\alpha = z + \mu y$$

   is a line in the (y, z) plane with slope $= -\mu$ (≤ 0) and intercept $= \alpha$ on the z axis.

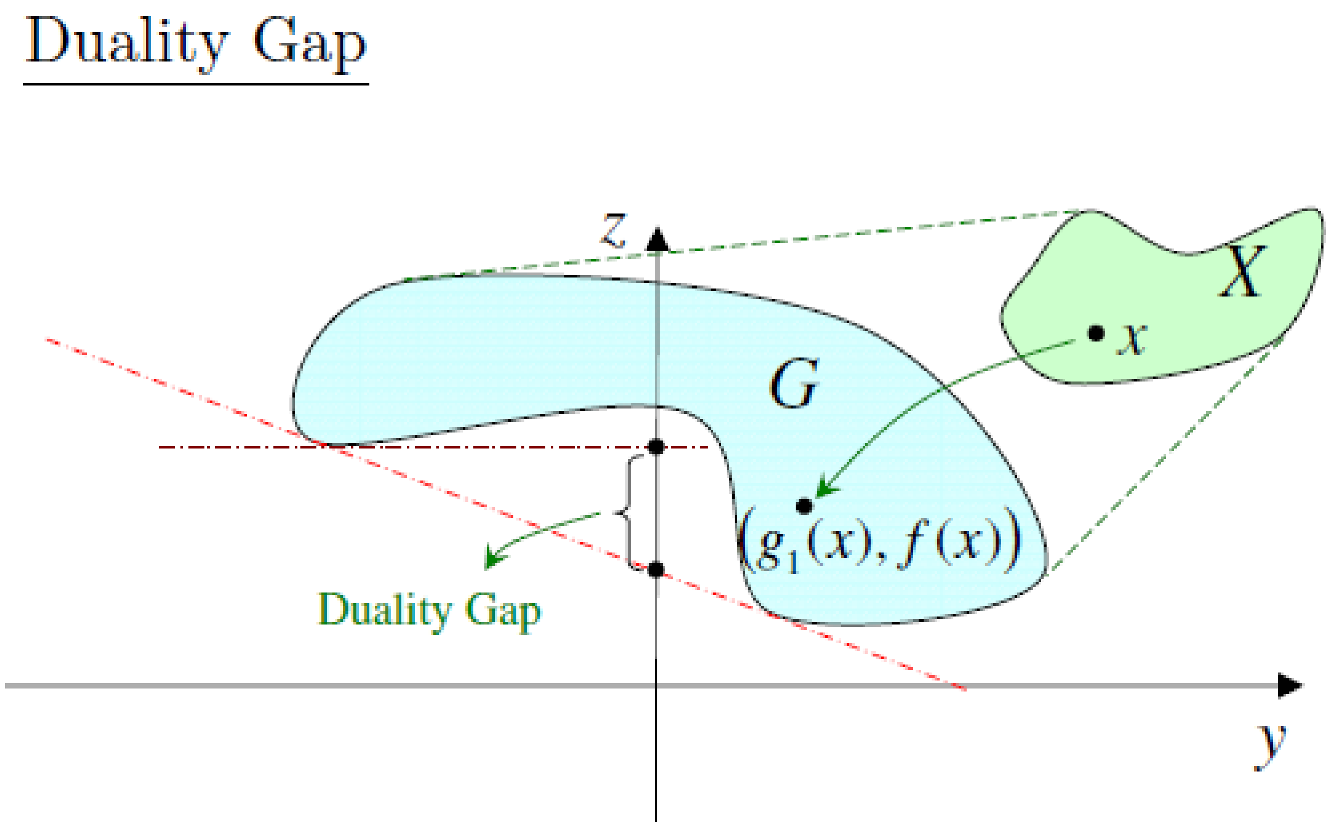


3. (LD) says that we should find the slope of the supporting hyperplane such that its intercept on the z axis is maximum.

4. When X is convex and f, g are convex, G must be convex. Its supporting hyperplane satisfies that

$$\begin{aligned} \phi(\mu^*) &= z^* + \underbrace{\mu^* y^*}_{=0} \\ &= z^* = f(x^*). \end{aligned}$$

Picture of duality gap



Full Lagrangian dual

$$\text{Minimize } f(x) = x^3$$
$$\text{s.t.} \quad -1 \le x \le 1$$
$$x \in E^1$$

Easy to observe $x^* = -1, f(x^*) = -1$.

Full Lagrangian dual

- Let $X = \{x \in E^1\}$.
- $\phi(\mu) = \inf_{x \in E^1}[x^3 + \mu_1(x - 1) + \mu_2(-x - 1)]$ for $\mu_1, \mu_2 \ge 0$.
- $\phi(\mu) = -\infty$ because $x^3 + \mu_1(x - 1) + \mu_2(-x - 1) \to -\infty$ as $x \to -\infty$.

Partial Lagrangian dual (1)

$$\text{Minimize } f(x) = x^3$$
$$\text{s.t.} \quad -1 \le x \le 1$$
$$x \in E^1$$

We know $x^* = -1, f(x^*) = -1.$

Partial Lagrangian dual (1):

- Let $X = \{x \in E^1 | \, x \ge -1\}$.
- $\phi(\mu) = \inf_{x \ge -1}[x^3 + \mu(x-1)]$ for $\mu \ge 0.$
- $x^* = -1$ because $x^3 + \mu(x-1)$ is increasing w.r.t. x.
- $\phi(\mu) = -1 - 2\mu.$

Dual: $$\text{Maximize } \phi(\mu) = -1 - 2\mu$$
$$\text{s.t.} \quad \mu \ge 0$$

$\mu^* = 0, \, \phi(\mu^*) = -1.$

Partial Lagrangian dual (2)

$$\text{Minimize } f(x) = x^3$$
$$\text{s.t.} \quad -1 \le x \le 1$$
$$x \in E^1$$

We know $x^* = -1, f(x^*) = -1.$

Partial Lagrangian dual (2):

- Let $X = \{x \in E^1 | \, x \le 1\}$.
- $\phi(\mu) = \inf_{x \le 1}[x^3 + \mu(-x - 1)]$ for $\mu \ge 0.$
- $\phi(\mu) = -\infty$ because $x^3 + \mu(-x - 1) \to -\infty$ as $x \to -\infty$.

Lagrangian dual of LP

Example 1 (Linear Programming)

$$
\text{(P)} \quad \begin{array}{ll} \text{minimize} & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array}
$$

Let $X = \{x \in E^n \mid x \geq 0\}$.

$$
\begin{aligned}
\phi(\lambda) &\triangleq \inf_{x \geq 0} \{c^T x + \lambda^T (b - Ax)\} \\
&= \lambda^T b + \inf_{x \geq 0} \{(c^T - \lambda^T A)x\} \\
&= \begin{cases} \lambda^T b, & \text{if } \mathrm{c}^{\mathrm{T}} - \lambda^{\mathrm{T}} \mathrm{A} \geq 0, \\ -\infty, & \text{otherwise.} \end{cases}
\end{aligned}
$$

$$
\text{(LD)} \quad \begin{array}{ll} \text{maximize} & \phi(\lambda) = b^T \lambda \\ \text{s.t.} & A^T \lambda \leq c \\ & \lambda : \text{unrestricted} \end{array}
$$

Lagrangian dual of QP

Example 2 (**Quadratic Programming**)

$$\text{(QP)} \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}x^T Q x + c^T x \\ \text{s.t.} & Ax \le b \end{array}$$

where Q is positive semi-definite.

Let $X = E^n$.

$$\phi(\mu) \triangleq \inf_{x \in E^n} \underbrace{\{\frac{1}{2}x^T Q x + c^T x + \mu^T (Ax - b)\}}_{\text{convex for any given } \mu}$$

The necessary and sufficient conditions for a minimum is that

$$Qx + A^T \mu + c = 0.$$

$$\text{(LD)} \quad \begin{array}{ll} \text{maximize} & \frac{1}{2}x^T Q x + c^T x + \mu^T (Ax - b) \\ \text{s.t.} & Qx + A^T \mu + c = 0 \\ & \mu \ge 0. \end{array}$$

Lagrangian dual of QP

Since $c^T x + \mu^T A x = -x^T Q x$, we have

$$\begin{array}{lll} & \text{maximize} & -\frac{1}{2}x^T Q x - b^T \mu \\ \text{(Dorn's Dual)} & \text{s.t.} & Qx + A^T \mu = -c \\ & & \mu \geq 0. \end{array}$$

When Q is positive definite, then

$$x^* = -Q^{-1}(c + A^T \mu)$$

and

$$\begin{aligned} \phi(\mu) = & \tfrac{1}{2}[Q^{-1}(c + A^T \mu)]^T Q [Q^{-1}(c + A^T \mu)] \\ & -c^T Q^{-1}(c + A^T \mu) \\ & +\mu^T(-AQ^{-1}(c + A^T \mu) - b) \\ = & \tfrac{1}{2}\mu^T \underbrace{(-AQ^{-1}A^T)}_{D:\ \text{negative definite}} \mu + \mu^T \underbrace{(-b - AQ^{-1}c)}_{d} \\ & -\tfrac{1}{2}c^T Q^{-1} c \end{aligned}$$

$$\begin{array}{lll} & \text{maximize} & \frac{1}{2}\mu^T D \mu + \mu^T d - \frac{1}{2}c^T Q^{-1} c \\ \text{(LD)} & \text{s.t.} & \mu \geq 0. \end{array}$$

Saddle point solution

$$\text{(NLP)} \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{s.t.} & \\ & \left.\begin{array}{l} g(x) \leq 0 \\ h(x) = 0 \\ x \in X \end{array}\right\} \mathscr{F} \end{array}$$

Lagrangian function

$$\ell(x, \mu, \lambda) \triangleq f(x) + \mu^T g(x) + \lambda^T h(x).$$

- Definition

$(\bar{x}, \bar{\mu}, \bar{\lambda}) \in E^{n+m+p}$ is called a <u>saddle point</u> (<u>solution</u>) of $\ell(x, \mu, \lambda)$ if

(i) $\bar{x} \in X$,

(ii) $\bar{\mu} \geq 0$,

(iii) $\ell(\bar{x}, \lambda, \mu) \leq \ell(\bar{x}, \bar{\mu}, \bar{\lambda}) \leq \ell(x, \bar{\mu}, \bar{\lambda})$,

$\forall\ x \in X,\ \mu \in E^p_+,\ \lambda \in E^m.$

Saddle point and duality gap

- Basic idea : The existence of a saddle point solution to the Lagrangian function is a necessary and sufficient condition for the absence of a duality gap!

Theorem 1:

Let $\bar{x} \in X$ and $\bar{\mu} \geq 0$. Then, $(\bar{x}, \bar{\mu}, \bar{\lambda})$ is a saddle point solution to $\ell(x, \mu, \lambda)$ if and only if

(a) $\ell(\bar{x}, \bar{\mu}, \bar{\lambda}) = \underset{x \in X}{\text{mimimize}} \; \ell(x, \bar{\mu}, \bar{\lambda}),$

(b) $g(\bar{x}) \leq 0$ and $h(\bar{x}) = 0,$

(c) $\bar{\mu}^T g(\bar{x}) = 0.$

Proof

Proof: (Part 1)

Let $(\bar{x}, \bar{\mu}, \bar{\lambda})$ be a saddle point solution.

By definition, we know (a) holds.

Moreover,

$$f(\bar{x}) + \bar{\mu}^T g(\bar{x}) + \bar{\lambda}^T h(\bar{x}) \geq f(\bar{x}) + \mu^T g(\bar{x}) + \lambda^T h(\bar{x}),$$
$$\forall\ \mu \in E^p_+,\ \ \lambda \in E^m.$$

This implies that $g(\bar{x}) \leq 0$ and $h(\bar{x}) = 0$, otherwise the right-hand-side may go unbounded above. This proves (b).

Now, let $\mu = 0$, the above inequality becomes

$$\bar{\mu}^T g(\bar{x}) \geq 0.$$

However, $\bar{\mu} \geq 0$ and $g(\bar{x}) \leq 0$ imply that

$$\bar{\mu}^T g(\bar{x}) \leq 0.$$

Hence $\bar{\mu}^T g(\bar{x}) = 0$. This proves (c).

(Part 2)

Suppose that $(\bar{x}, \bar{\mu}, \bar{\lambda})$ with $\bar{x} \in X$ and $\bar{\mu} \geq 0$ such that (a),(b),(c) hold. Then, by (a)

$$\ell(\bar{x}, \bar{\mu}, \bar{\lambda}) \leq \ell(x, \bar{\mu}, \bar{\lambda}),\ \forall\ x \in X.$$

By (b)and (c)

$$\begin{aligned} \ell(\bar{x}, \bar{\mu}, \bar{\lambda}) &= f(\bar{x}) + \bar{\mu}^T g(\bar{x}) + \bar{\lambda}^T h(\bar{x}) = f(\bar{x}) \\ &\geq f(\bar{x}) + \mu^T g(\bar{x}) + \lambda^T h(\bar{x}) \\ &= \ell(\bar{x}, \mu, \lambda) \end{aligned}$$

with $\mu \in E^p_+$ and $\lambda \in E^m$.

Hence $(\bar{x}, \bar{\mu}, \bar{\lambda})$ is a saddle point solution.

Saddle point theorem

Theorem 2:

$(\bar{x}, \bar{\mu}, \bar{\lambda})$ is a saddle point solution of $\ell(x, \mu, \lambda)$ if and only if $\bar{x}$ is a primal optimal solution, $(\bar{\mu}, \bar{\lambda})$ is a dual optimal solution and $f(\bar{x}) = \phi(\bar{\mu}, \bar{\lambda})$.

Proof: (Part 1)

Let $(\bar{x}, \bar{\mu}, \bar{\lambda})$ be a saddle point solution of $\ell(x, \mu, \lambda)$.

By (b) of Theorem 1, $\bar{x}$ is primal feasible. Since $\bar{\mu} \geq 0$, $(\bar{\mu}, \bar{\lambda})$ is dual feasible. Combing (a), (b), and (c), we have

$$\begin{aligned}\phi(\bar{\mu}, \bar{\lambda}) &= \ell(\bar{x}, \bar{\mu}, \bar{\lambda}) \\ &= f(\bar{x}) + \bar{\mu}^T g(\bar{x}) + \bar{\lambda}^T h(\bar{x}) \\ &= f(\bar{x}).\end{aligned}$$

By the Weak Duality Theorem, we know that $\bar{x}$ is primal optimal and $(\bar{\mu}, \bar{\lambda})$ is dual optimal.

(Part 2)

Let $\bar{x}$ and $(\bar{\mu}, \bar{\lambda})$ be optimal solutions to (P) and (D), respectively, with

$$f(\bar{x}) = \phi(\bar{\mu}, \bar{\lambda}).$$

Hence, we have $\bar{x} \in X$, $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $\bar{\mu} \geq 0$. Moreover,

$$\begin{aligned}\phi(\bar{\mu}, \bar{\lambda}) &\triangleq \inf_{x \in X} \{f(x) + \bar{\mu}^T g(x) + \bar{\lambda}^T h(x)\} \\ &\leq f(\bar{x}) + \bar{\mu}^T g(\bar{x}) + \bar{\lambda}^T h(\bar{x}) \\ &= f(\bar{x}) + \bar{\mu}^T g(\bar{x}) \\ &\leq f(\bar{x})\end{aligned}$$

But $\phi(\bar{\mu}, \bar{\lambda}) = f(\bar{x})$ is given, the inequalities become equalities. Hence $\mu^T g(\bar{x}) = 0$ and

$$\begin{aligned}\ell(\bar{x}, \bar{\mu}, \bar{\lambda}) &= f(\bar{x}) = \phi(\bar{\mu}, \bar{\lambda}) \\ &= \underset{x \in X}{\text{minimum}} \ \ell(x, \bar{\mu}, \bar{\lambda}).\end{aligned}$$

By Theorem 1, we know $(\bar{x}, \bar{\mu}, \bar{\lambda})$ is a saddle point solution to $\ell(x, \mu, \lambda)$.

Saddle point and KKT conditions

Question:

How does saddle point optimality relate to the K-K-T conditions?

Theorem 3:

Let $\bar{x} \in \mathscr{F}$ satisfies the K-K-T conditions with $\bar{\mu} \in E_+^p$ and $\bar{\lambda} \in E^m$.

Suppose that f, g_i $(i \in I(\bar{x}))$ are convex at $\bar{x}$, and that h_j is affine for those with $\bar{\lambda}_j \neq 0$.

Then,$(\bar{x}, \bar{\mu}, \bar{\lambda})$ is a saddle point of $\ell(x, \mu, \lambda)$.

Conversely, let $(\bar{x}, \bar{\mu}, \bar{\lambda})$ be a saddle point solution of $\ell(x, \mu, \lambda)$ with $\bar{x} \in \text{int } X$. Then $\bar{x}$ is primal feasible and $(\bar{x}, \bar{\mu}, \bar{\lambda})$ satisfies the K-K-T conditions.

Proof

(Part 1)

Let $\bar{x} \in \mathscr{F}$, $\bar{\mu} \in E^p_+$, $\bar{\lambda} \in E^m$ and $(\bar{x}, \bar{\mu}, \bar{\lambda})$ satisfies the K-K-T conditions, i.e.,

$$\left.\begin{aligned} \nabla f(\bar{x}) + \bar{\mu}^T \nabla g(\bar{x}) + \bar{\lambda}^T \nabla h(\bar{x}) = 0 \\ \bar{\mu}^T g(\bar{x}) = 0. \end{aligned}\right\} (*)$$

By convexity and linearity of f, g_i and h_j, we have

$$\begin{aligned} f(x) &\geq f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x}), \\ g_i(x) &\geq g_i(\bar{x}) + \nabla g_i(\bar{x})(x - \bar{x}), \quad i \in I(\bar{x}), \\ h_j(x) &= h_j(\bar{x}) + \nabla h_j(\bar{x})(x - \bar{x}), \quad j = 1, \cdots, m, \ \bar{\lambda}_j \neq 0, \end{aligned}$$

for $x \in X$.

Multiplying the second inequality by $\bar{\mu}_i$ and the third inequality by $\bar{\lambda}_j$, adding to the first inequality, and noting $(*)$, it follows from the definition of ℓ that

$$\ell(x, \bar{\mu}, \bar{\lambda}) \geq \ell(\bar{x}, \bar{\mu}, \bar{\lambda}), \ \forall \, x \in X.$$

Moreover, since $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $\bar{\mu}^T g(\bar{x}) = 0$, we have

$\ell(\bar{x}, \mu, \lambda) \leq \ell(\bar{x}, \bar{\mu}, \bar{\lambda})$ for $\mu \in E^p_+$ and $\lambda \in E^m$.

Hence $(\bar{x}, \bar{\mu}, \bar{\lambda})$ is a saddle point solution.

(Part 2)

Suppose that $(\bar{x}, \bar{\mu}, \bar{\lambda})$ with $\bar{x} \in \text{int } X$ and $\bar{\mu} \geq 0$ is a saddle point solution. Since $\ell(\bar{x}, \mu, \lambda) \leq \ell(\bar{x}, \bar{\mu}, \bar{\lambda})$ for $\mu \in E^p_+$ and $\lambda \in E^m$. Like in Theorem 1 (Part 1), we have $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $\bar{\mu}^T g(\bar{x}) = 0$.

Hence $\bar{x}$ is primal feasible. Moreover $\bar{x}$ is a primal optimal solution because $\ell(\bar{x}, \mu, \lambda) \leq \ell(\bar{x}, \bar{\mu}, \bar{\lambda})$ for $x \in X$.

Since $\bar{x} \in \text{int } X$, we have $\nabla_x \ell(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0$, i.e.,

$$\nabla f(\bar{x}) + \bar{\mu}^T \nabla g(\bar{x}) + \bar{\lambda}^T \nabla h(\bar{x}) = 0$$

This completes the proof.

Lagrangian dual vs. conjugate dual

- Any real number x can be uniquely expressed as

$$x = x^+ - x^-,$$

where $x^+ \geq 0$, $x^- \geq 0$ and $x^+ \cdot x^- = 0$.

- Consider a standard form NLP:

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ s.t. \quad & g_i(\boldsymbol{x}) \leq 0, i = 1, \dots, m \\ & h_j(\boldsymbol{x}) = 0, j = 1, \dots, p \\ & \boldsymbol{x} \in E_+^n \end{aligned}$$

Lagrangian dual vs. conjugate dual

- Conjugate dual (CD)

$$S \triangleq \{\boldsymbol{x} \in E^n | \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}, \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$$

$$\mathcal{X} \triangleq E^n_+$$

$$\hbar(\boldsymbol{y}) \triangleq \sup_{\boldsymbol{x} \in S}\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} - f(\boldsymbol{x})\}, \forall \boldsymbol{y} \in \Omega$$

$$\Omega \triangleq \left\{\boldsymbol{y} \in E^n \middle| \sup_{\boldsymbol{x} \in S}\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} - f(\boldsymbol{x})\}, < \infty\right\}$$

$$\mathcal{Y} \triangleq dual(\mathcal{X}) = E^n_+$$

(CD)
$$\begin{aligned} \min \quad & \hbar(\boldsymbol{y}) \\ s.t. \quad & \boldsymbol{y} \in \Omega \\ & \boldsymbol{y} \in E^n_+ \end{aligned}$$

Lagrangian dual vs. conjugate dual

- Lagrangian dual (LD)

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ s.t. \quad & \boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0} \\ & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \\ & -x_i \le 0, i = 1, \dots, n \end{aligned}$$

Let $S \triangleq \{\boldsymbol{x} \in E^n | \boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}, \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ for partial Lagrangian:

(Partial LD)

$$\begin{aligned} \max \quad & \phi(\boldsymbol{\lambda}) \triangleq \inf_{x \in S} \{f(\boldsymbol{x}) + \boldsymbol{\lambda}^{\mathrm{T}}(-\boldsymbol{x})\} \\ s.t. \quad & \boldsymbol{\lambda} \ge \boldsymbol{0} \end{aligned}$$

when $\inf_{x \in S} \{f(\boldsymbol{x}) + \boldsymbol{\lambda}^{\mathrm{T}}(-\boldsymbol{x})\}$ exists.

Lagrangian dual vs. conjugate dual

- Lagrangian dual (LD)

(Partial LD)
$$\max \quad \phi(\boldsymbol{\lambda}) \triangleq \inf_{\boldsymbol{x}\in S} \{f(\boldsymbol{x}) + \boldsymbol{\lambda}^{\mathrm{T}}(-\boldsymbol{x})\}$$
$$s.t. \quad \boldsymbol{\lambda} \geq \boldsymbol{0}$$

Notice that

$$\phi(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}\in S} \{f(\boldsymbol{x}) + \boldsymbol{\lambda}^{\mathrm{T}}(-\boldsymbol{x})\} > -\infty$$
$$= (-)\left\{\sup_{\boldsymbol{x}\in S} \{\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{x} - f(\boldsymbol{x})\} < \infty\right\}$$
$$= (-)h(\boldsymbol{\lambda})$$

if and only if $\boldsymbol{\lambda} \in \Omega$.

- (Partial LD) $\begin{array}{ll} \max & (-)h(\boldsymbol{\lambda}) \\ s.t. & \boldsymbol{\lambda} \in \Omega \\ & \boldsymbol{\lambda} \in E_+^n \end{array}$ = $\begin{array}{ll} (-)\min & h(\boldsymbol{y}) \\ s.t. & \boldsymbol{y} \in \Omega \\ & \boldsymbol{y} \in E_+^n \end{array}$ (–)(CD)