Numerical Optimization

Duality

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization

Two-player zero-sum game

A Game between two players P and D

- Game setting
  - $\mathcal{X}$: A set of strategies for P
  - $\mathcal{Y}$: A set of strategies for D
  - Payoff function, $\psi(x, y), \; x \in \mathcal{X}, \; y \in \mathcal{Y}$
- Example
  - Let $\mathcal{X} = \{1, 2\}, \; \mathcal{Y} = \{1, 2\}$
  - Payoff $\psi(x, y) = a_{x,y}$ where $A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$
- Game Rules :
  - P chooses a strategy $x \in \mathcal{X}$ and D chooses a strategy $y \in \mathcal{Y}$ independently
  - The referee reveals both the strategies simultaneously
  - Game Outcome : Depends on $\psi(x, y)$

Two-player zero-sum game

A Game between two players P and D

- Game Outcome:

$$\begin{aligned} \psi(x,y) > 0 \;&\Rightarrow\; P \text{ pays an amount } \psi(x,y) \text{ to } D \\ \psi(x,y) < 0 \;&\Rightarrow\; D \text{ pays an amount } -\psi(x,y) \text{ to } P \end{aligned}$$

- P wishes to minimize payoff to D, while D wishes to receive maximum payoff from P
- Assume that minimum and maximum exist

Example: Game 1

$\mathcal{X} = \{1, 2\},\ \mathcal{Y} = \{1, 2\},\ \psi(x, y) = a_{x,y},$ where

$$A = \begin{pmatrix} -2 & 1 \\ 2 & -3 \end{pmatrix}$$

Player P's strategy

$$\begin{aligned} & \min\{\max_y a_{1,y}, \max_y a_{2,y}\} \\ = & \min\{1, 2\} \\ = & 1 \end{aligned}$$

Choose $x = 1$

Player D's strategy

$$\begin{aligned} & \max\{\min_x a_{x,1}, \min_x a_{x,2}\} \\ = & \max\{-2, -3\} \\ = & -2 \end{aligned}$$

Choose $y = 1$

min-max $\geq$ **max-min**

Example: Game 2

$\mathcal{X} = \{1, 2\},\ \mathcal{Y} = \{1, 2\},\ \psi(x, y) = a_{x,y},$ where

$$A = \begin{pmatrix} -2 & 1 \\ 2 & 3 \end{pmatrix}$$

Player *P*'s strategy

$$\begin{aligned} & \min\{\max_y a_{1,y}, \max_y a_{2,y}\} \\ = & \min\{1, 3\} \\ = & 1 \end{aligned}$$

Choose $x = 1$

Player *D*'s strategy

$$\begin{aligned} & \max\{\min_x a_{x,1}, \min_x a_{x,2}\} \\ = & \max\{-2, 1\} \\ = & 1 \end{aligned}$$

Choose $y = 2$

min-max $=$ **max-min**

Primal Problem

$$\min_{x \in \mathcal{X}} \underbrace{\max_{y \in \mathcal{Y}} \psi(x, y)}_{\text{primal function}}$$

Dual Problem

$$\max_{y \in \mathcal{Y}} \underbrace{\min_{x \in \mathcal{X}} \psi(x, y)}_{\text{dual function}}$$

- The two problems are *dual* to each other
- For any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$

$$\min_{x \in \mathcal{X}} \psi(x, y) \leq \psi(x, y) \leq \max_{y \in \mathcal{Y}} \psi(x, y)$$

$$\therefore \min_{x \in \mathcal{X}} \psi(x, y) \leq \max_{y \in \mathcal{Y}} \psi(x, y)$$

$$\therefore \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \psi(x, y) \leq \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \psi(x, y)$$

Weak Duality

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \psi(x, y) \leq \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \psi(x, y)$$

Weak Duality

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \psi(x, y) \leq \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \psi(x, y)$$

- When does the equality hold?

Definition

Let $x^* \in \mathcal{X}$ and $y^* \in \mathcal{Y}$. A point (x^*, y^*) is a saddle point for $\psi(x, y)$ if

$$\psi(x^*, y) \leq \psi(x^*, y^*) \leq \psi(x, y^*) \;\; \forall\, x \in \mathcal{X}, y \in \mathcal{Y}$$

- $x^* = \text{argmin}_{x \in \mathcal{X}} \psi(x, y^*)$
- $y^* = \text{argmax}_{y \in \mathcal{Y}} \psi(x^*, y)$

Theorem

The following equality holds

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \psi(x, y) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \psi(x, y)$$

if and only if there exists a saddle point, (x^*, y^*), *for* $\psi(x, y)$.

Proof.

(a) Let (x^*, y^*) be a saddle point for $\psi(x, y)$.

$$\therefore \psi(x^*, y) \leq \psi(x^*, y^*) \leq \psi(x, y^*) \ \ \forall\, x \in \mathcal{X}, y \in \mathcal{Y}$$
$$\therefore \max_{y \in \mathcal{Y}} \psi(x^*, y) \leq \psi(x^*, y^*) \leq \min_{x \in \mathcal{X}} \psi(x, y^*)$$

Note that

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \psi(x, y) \leq \max_{y \in \mathcal{Y}} \psi(x^*, y)$$
$$\min_{x \in \mathcal{X}} \psi(x, y^*) \leq \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \psi(x, y^*).$$

Proof.(continued)

Therefore,

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \psi(x, y) \leq \psi(x^*, y^*) \leq \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \psi(x, y)$$

But, we know that

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \psi(x, y) \leq \psi(x^*, y^*) \leq \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \psi(x, y)$$

Therefore,

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \psi(x, y) = \psi(x^*, y^*) = \max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \psi(x, y)$$

Proof. (continued)

(b) Suppose the following equality holds for some $x^* \in \mathcal{X},\ y^* \in \mathcal{Y}$,

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \psi(x, y) = \psi(x^*, y^*) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \psi(x, y)$$

Now,

$$\max_{y \in \mathcal{Y}} \psi(x^*, y) = \psi(x^*, y^*) = \min_{x \in \mathcal{X}} \psi(x, y^*)$$

$$\therefore \psi(x^*, y) \leq \max_{y \in \mathcal{Y}} \psi(x^*, y) = \psi(x^*, y^*) = \min_{x \in \mathcal{X}} \psi(x, y^*) \leq \psi(x, y^*)$$

Therefore, (x^*, y^*) is a saddle point for $\psi(x, y)$. $\square$

Strong Duality

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \psi(x, y) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \psi(x, y)$$

Consider the problem (**NLP**):

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m \end{array}$$

- Can we define a game with a payoff function $\psi(\cdot)$ so that the solution to **NLP** is a solution to the *primal* problem, $\min_x \max_y \psi(x, y)$?
- What is the saddle point condition in terms of f, h_j's and e_i's?

Consider the problem(**P**):

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0, \; j = 1, \ldots, l \\ & x \in X \end{array}$$

Define a payoff function as the Lagrangian,

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x})$$

where $\boldsymbol{x} \in X$ and $\lambda_j \geq 0, \; j = 1, \ldots, l$

- $\boldsymbol{x}$: Primal Variables, $\boldsymbol{\lambda}$: Dual Variables
- $\mathcal{X} = X, \mathcal{Y} = \{\boldsymbol{\lambda} \in \mathbb{R}^l : \lambda_j \geq 0, \; j = 1, \ldots, l\}$

Duality : Define a **min max** problem *equivalent* to the primal problem **P**. Then, the corresponding dual **max min** problem is the dual problem **D**.

Assumption: Minimum and Maximum exist for the problems defined here (Use infimum or supremum appropriately).

$$
\begin{aligned}
\text{Primal Function} &= \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \\
&= \max_{\boldsymbol{\lambda} \geq \mathbf{0}} f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x}) \\
&= \begin{cases} f(\boldsymbol{x}) & \text{if } h_j(\boldsymbol{x}) \leq 0 \ \forall\, j \\ +\infty & \text{Otherwise.} \end{cases}
\end{aligned}
$$

Primal Problem:

$$\min_{\boldsymbol{x} \in X} \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$$

That is, (ignoring the possibility of $h_j(\boldsymbol{x}) > 0 \ \forall\, j$),

$$
\begin{array}{ll}
\min & f(\boldsymbol{x}) \\
\text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \\
& \boldsymbol{x} \in X
\end{array}
$$

For $\boldsymbol{\lambda} \geq \mathbf{0}$, define

$$\begin{aligned} \text{Dual Function} &= \theta(\boldsymbol{\lambda}) \\ &= \min_{\boldsymbol{x} \in X} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \\ &= \min_{\boldsymbol{x} \in X} f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x}) \end{aligned}$$

Dual Problem:

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}} \min_{\boldsymbol{x} \in X} f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x})$$

Consider the problem:

$$\begin{array}{ll} \min & x^2 \\ \text{s.t.} & x \geq 1 \end{array}$$

- Primal solution: $x^* = 1,\ f(x^*) = 1$.
  $\mathcal{L}(x, \lambda) = x^2 + \lambda(1 - x)$
- Dual function: $\theta(\lambda) = \min_x x^2 + \lambda(1 - x)$.
  At the minimum, $x^* = \frac{\lambda}{2}$.
  For $\lambda \geq 0$, $\theta(\lambda) = -\frac{1}{4}\lambda^2 + \lambda$.
  Therefore, the dual problem is

$$\max_{\lambda \geq 0} \ -\frac{1}{4}\lambda^2 + \lambda$$

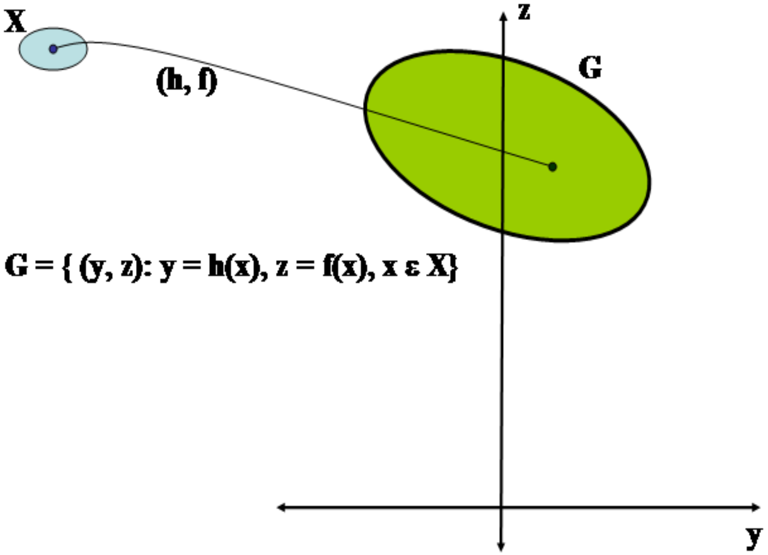
- $\lambda^* = 2,\ \theta(\lambda^*) = 1$
- $f(x^*) = 1 = \theta(\lambda^*)$

Geometric Interpretation

Consider the problem (**P1**):

$$\left.\begin{array}{cc} \min_{x\in X} & f(\boldsymbol{x}) \\ \text{s.t.} & h(\boldsymbol{x}) \le 0 \end{array}\right\} \equiv \min_{x\in X} \max_{\lambda\ge 0} f(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$$

Define $G = \{(y,z) : y = h(\boldsymbol{x}), z = f(\boldsymbol{x}), \boldsymbol{x} \in X\}$.

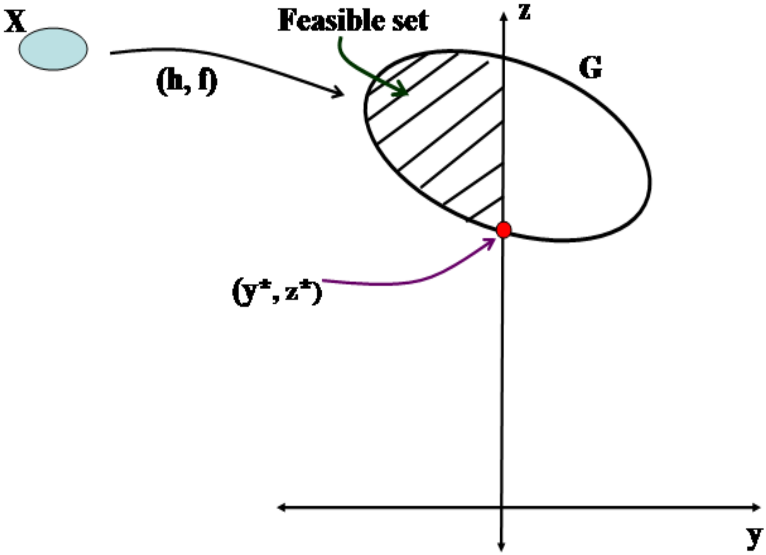

X
(h, f)
z
G
G = { (y, z): y = h(x), z = f(x), x ε X}
y

Geometric Interpretation

Consider the problem (**P1**):

$$\left.\begin{array}{ll} \min_{x\in X} & f(\boldsymbol{x}) \\ \text{s.t.} & h(\boldsymbol{x}) \leq 0 \end{array}\right\} \equiv \min_{x\in X} \max_{\lambda\geq 0} f(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$$

Define $G = \{(y,z) : y = h(\boldsymbol{x}), z = f(\boldsymbol{x}), \boldsymbol{x} \in X\}$.
A solution to the primal problem **P1** is a point in G with $y \leq 0$ and has minimum ordinate z.


X
(h, f)
Feasible set
z
G
(y*, z*)
y

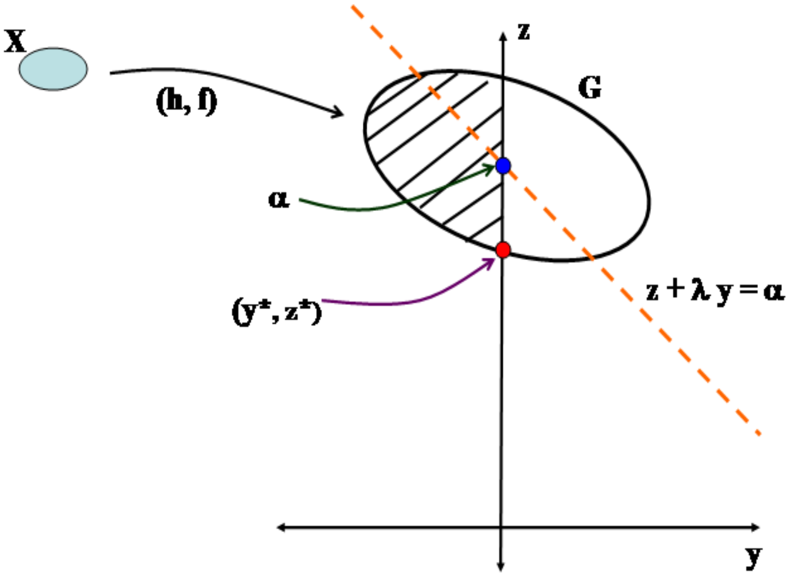
Geometric Interpretation

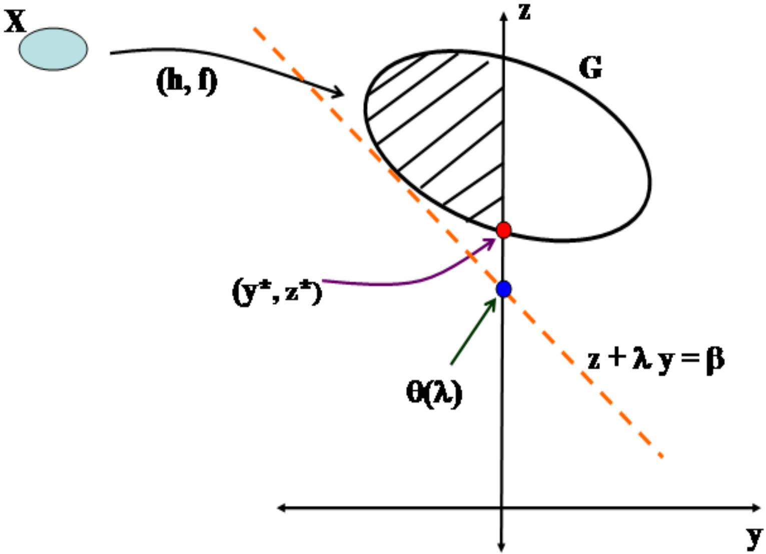
Consider the problem (**P1**):

$$\left.\begin{array}{cc} \min_{x \in X} & f(\boldsymbol{x}) \\ \text{s.t.} & h(\boldsymbol{x}) \leq 0 \end{array}\right\} \equiv \min_{x \in X} \max_{\lambda \geq 0} f(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$$

Define $G = \{(y, z) : y = h(\boldsymbol{x}), z = f(\boldsymbol{x}), \boldsymbol{x} \in X\}$.
A solution to the primal problem **P1** is a point in G with $y \leq 0$ and has minimum ordinate z.
Let (y^*, z^*) be this point in $y - z$ space.
For a given $\lambda \geq 0$,

- Define $\theta(\lambda) = \min_{x \in X} f(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$.
- $\theta(\lambda)$ is a minimum $z + \lambda y$ over feasible G in $y - z$ space.


X
(h, f)
z
G
α
(y*, z*)
z + λ y = α
y


X
(h, f)
z
G
(y^*, z^*)
$\theta(\lambda)$
$z + \lambda y = \beta$
y

Geometric Interpretation

Consider the problem (**P1**):

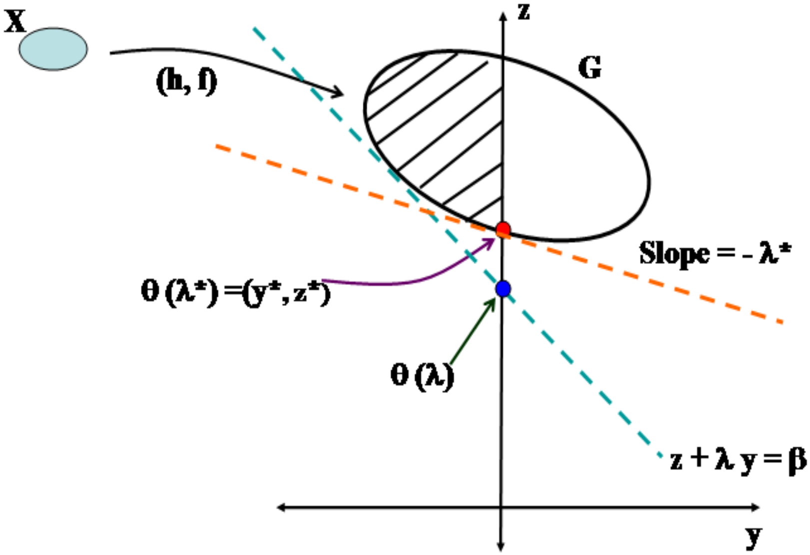
$$\left.\begin{array}{cc} \min_{x \in X} & f(\boldsymbol{x}) \\ \text{s.t.} & h(\boldsymbol{x}) \leq 0 \end{array}\right\} \equiv \min_{x \in X} \max_{\lambda \geq 0} f(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$$

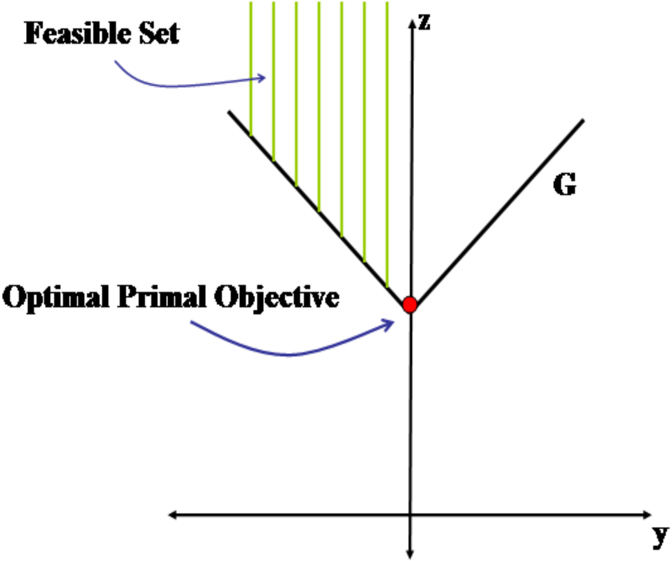
Define $G = \{(y, z) : y = h(\boldsymbol{x}), z = f(\boldsymbol{x}), \boldsymbol{x} \in X\}$.
A solution to the primal problem **P1** is a point in G with $y \leq 0$ and has minimum ordinate z.
Let (y^*, z^*) be this point in $y - z$ space.
For a given $\lambda \geq 0$,

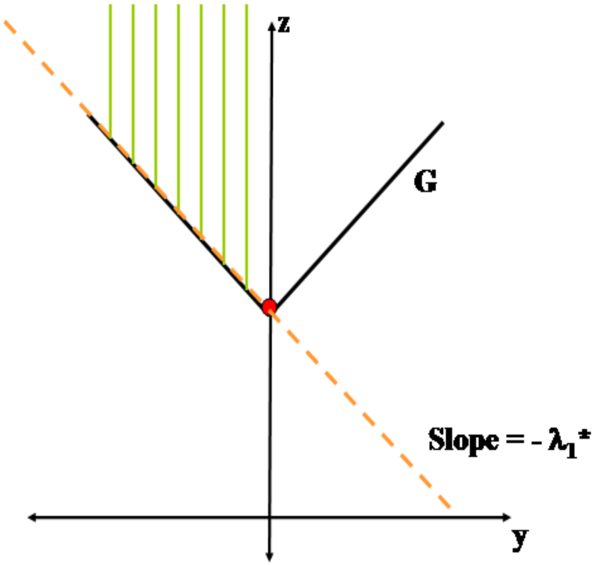
- Define $\theta(\lambda) = \min_{x \in X} f(\boldsymbol{x}) + \lambda h(\boldsymbol{x})$.
- $\theta(\lambda)$ is a minimum $z + \lambda y$ over feasible G in $y - z$ space.

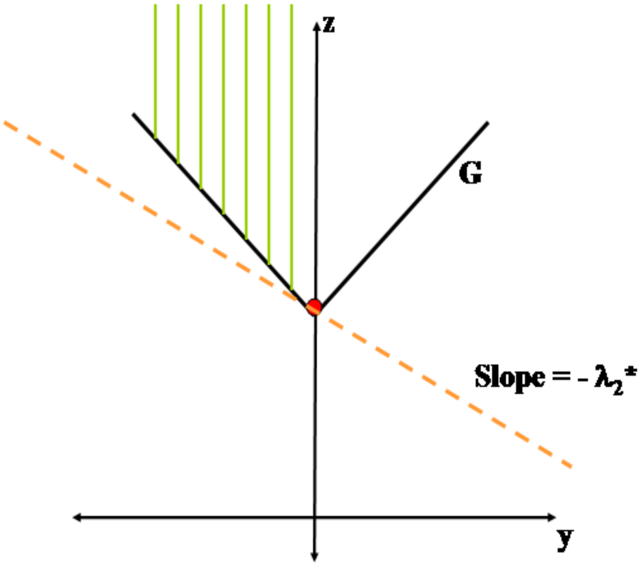
Lagrangian Dual Problem (**D1**):

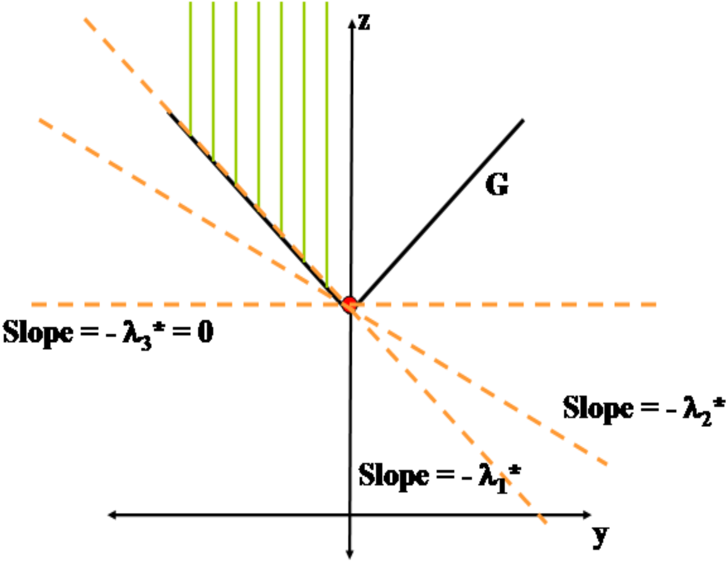
$$\max_{\lambda \geq 0} \theta(\lambda) \equiv \max_{\lambda \geq 0} \min_{x \in X} f(\boldsymbol{x}) + \lambda h(\boldsymbol{x}).$$

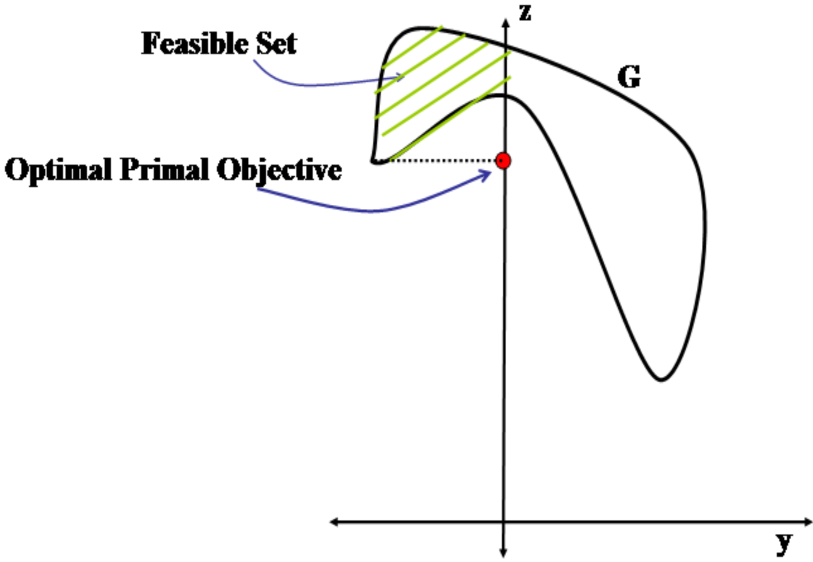

X
(h, f)
z
G
Slope = - λ*
θ (λ*) =(y*, z*)
θ (λ)
z + λ y = β
y

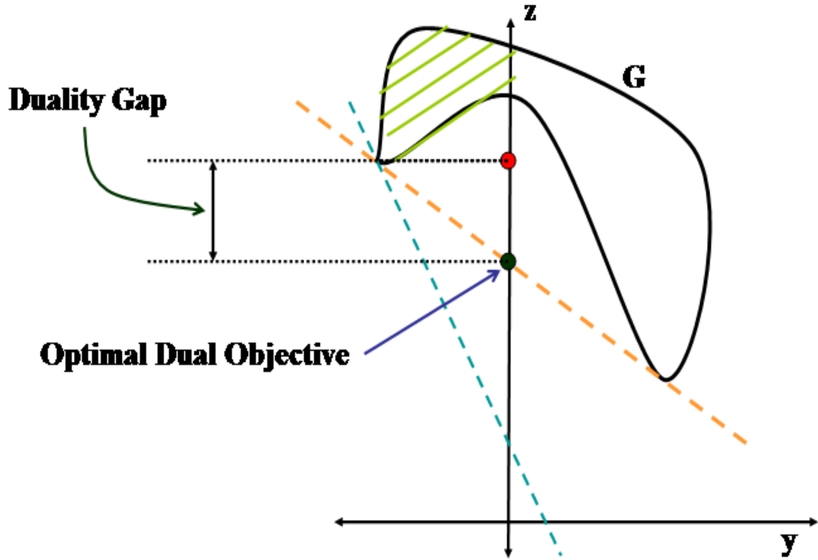

Feasible Set
z
G
Optimal Primal Objective
y


z
G
Slope = - λ_1*
y


z
G
Slope = - λ2*
y


z
G
Slope = - λ3* = 0
Slope = - λ2*
Slope = - λ1*
y


z
Feasible Set
G
Optimal Primal Objective
y



z
G
Duality Gap
Optimal Dual Objective
y

Primal Problem

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0, j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0, i = 1, \ldots, m \\ & \boldsymbol{x} \in X \end{array}$$

Dual Problem

$$\begin{array}{ll} \max & \theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} & \boldsymbol{\lambda} \geq \mathbf{0} \end{array}$$

where $\theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\boldsymbol{x} \in X} \ \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$.

Theorem

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

$$f(\boldsymbol{x}) \geq \theta(\boldsymbol{\lambda}, \boldsymbol{\mu}).$$

Primal Problem

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0, j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0, i = 1, \ldots, m \\ & \boldsymbol{x} \in X \end{array}$$

Dual Problem

$$\begin{array}{ll} \max & \theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} & \boldsymbol{\lambda} \geq \mathbf{0} \end{array}$$

where $\theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\boldsymbol{x} \in X} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$.

Proof.

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

$$\begin{aligned} \theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \min_{\boldsymbol{x} \in X} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &= \min_{\boldsymbol{x} \in X} f(\boldsymbol{x}) + \sum_{j=1}^{l} \underbrace{\lambda_j h_j(\boldsymbol{x})}_{\leq 0} + \sum_{i=1}^{m} \underbrace{\mu_i e_i(\boldsymbol{x})}_{=0} \\ &\leq f(\boldsymbol{x}) \end{aligned}$$

Primal Problem

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_j(\boldsymbol{x}) \le 0, j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0, i = 1, \ldots, m \\ & \boldsymbol{x} \in X \end{aligned}$$

Dual Problem

$$\begin{aligned} \max \quad & \theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} \quad & \boldsymbol{\lambda} \ge \mathbf{0} \end{aligned}$$

where $\theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\boldsymbol{x} \in X} \ \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$.

Weak Duality Theorem

Let p^* and d^* be optimal primal and dual objective function values respectively.

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

$$\begin{aligned} f(\boldsymbol{x}) \quad & \ge \quad \theta(\boldsymbol{\lambda}, \boldsymbol{\mu}). \\ \therefore \quad & \quad \min\{f(\boldsymbol{x}) : h_j(\boldsymbol{x}) \le 0 \ \forall \, j, e_i(\boldsymbol{x}) = 0 \ \forall \, i, \boldsymbol{x} \in X\} \\ & \ge \quad \max\{\theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\lambda} \ge \mathbf{0}\} \\ \therefore p^* \quad & \ge \quad d^*. \end{aligned}$$

Example:
Consider the problem:

$$\begin{array}{ll} \min & x^3 \\ \text{s.t.} & x = 1 \\ & x \in \mathbb{R} \end{array}$$

- $x^* = 1,\ f(x^*) = 1.$
- Dual function:

$$\begin{aligned} \theta(\mu) &= \min_{x \in \mathbb{R}} x^3 + \mu(x-1) \\ &= \min_{x \in \mathbb{R}} x^3 + \mu x - \mu \\ &= -\infty \ \forall \ \mu \in \mathbb{R} \end{aligned}$$

$$\therefore \theta(\mu^*) = -\infty \ < f(x^*) \ \Rightarrow \ d^* < p^*$$

$\Rightarrow$ There exists a duality gap.

Recall the example of two-player zero-sum game.

Example: Game 2

$\mathcal{X} = \{1, 2\},\ \mathcal{Y} = \{1, 2\},\ \psi(x, y) = a_{x,y},$ where

$$A = \begin{pmatrix} -2 & 1 \\ 2 & 3 \end{pmatrix}$$

Player P's strategy

$$\begin{aligned} & \min\{\max_y a_{1,y}, \max_y a_{2,y}\} \\ = & \min\{1, 3\} \\ = & 1 \end{aligned}$$

Choose $x = 1$

Player D's strategy

$$\begin{aligned} & \max\{\min_x a_{x,1}, \min_x a_{x,2}\} \\ = & \max\{-2, 1\} \\ = & 1 \end{aligned}$$

Choose $y = 2$

min-max $=$ **max-min**

Primal Problem

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0, j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0, i = 1, \ldots, m \\ & \boldsymbol{x} \in X \end{array}$$

Dual Problem

$$\begin{array}{ll} \max & \theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} & \boldsymbol{\lambda} \geq \mathbf{0} \end{array}$$

where $\theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\boldsymbol{x} \in X} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$.

Let $\boldsymbol{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ be optimal solutions to the primal and dual problems respectively. Let p^* and d^* be optimal primal and dual objective function values respectively.

$p^* = d^* \;\Rightarrow$ There is no duality gap.

Under what conditions is $p^* = d^*$?

Optimal primal and dual objective function values are same ($p^* = d^*$) if and only if $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a Lagrangian saddle point, that is, for $\boldsymbol{x}, \boldsymbol{x}^* \in X$ and $\boldsymbol{\lambda}, \boldsymbol{\lambda}^* \geq \mathbf{0}$,

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*).$$

Proof.

(a)

Let $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ be a Lagrangian saddle point where $\boldsymbol{x}^* \in X$ and $\boldsymbol{\lambda}^* \geq \mathbf{0}$. Let $\boldsymbol{\lambda} \geq \mathbf{0}$.

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

$$\therefore f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i e_i(\boldsymbol{x}^*)$$

$$\leq f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x}^*)$$

$$\therefore \left.\begin{array}{l} h_j(\boldsymbol{x}^*) \leq 0 \ \forall\, j \\ e_i(\boldsymbol{x}^*) = 0 \ \forall\, i \end{array}\right\} \text{ and } \boldsymbol{x}^* \in X \Rightarrow \ \boldsymbol{x}^* \text{ is primal feasible}$$

Proof.(continued)

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

$$\therefore \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i e_i(\boldsymbol{x}^*) \leq \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x}^*)$$

$$\therefore \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x}^*) \leq \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) \quad (\because e_i(\boldsymbol{x}^*) = 0 \ \forall\ i)$$

$$\therefore 0 \leq \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) \qquad (\text{Letting } \ \lambda_j = 0 \ \forall\ j)$$

$$\text{Also, } 0 \geq \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*). \quad (\because \lambda_j^* \geq 0, h_j(\boldsymbol{x}^*) \leq 0 \ \forall\ j)$$

$$\therefore \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) = 0 \ \Rightarrow \ \lambda_j^* h_j(\boldsymbol{x}^*) = 0 \ \forall\ j$$

Proof.(continued)

$(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a saddle point. $\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$.
Therefore, the dual function at $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$,

$$\begin{aligned}
\theta(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \min_{x \in X} f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x}) \\
&= \min_{x \in X} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \\
&= \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \\
&= f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x}^*) \\
&= f(\boldsymbol{x}^*)
\end{aligned}$$

$\therefore d^* = p^*$.

Proof.(continued)

(b)

Let $f(\boldsymbol{x}^*) = \theta(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. Note that $\boldsymbol{x}^*$ is primal feasible and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is dual feasible. Let $\boldsymbol{x}$ be primal feasible and $\lambda_j \geq 0 \ \forall\, j$.

$$
\begin{aligned}
\therefore\ \theta(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= \min_{x \in X} f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x}) \\
&\leq f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x}^*) \\
&= f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) \\
&\leq f(\boldsymbol{x}^*) \qquad (\because \lambda_j^* \geq 0, h_j(\boldsymbol{x}^*) \leq 0)
\end{aligned}
$$

But, $\theta(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}^*)$. Therefore, $\lambda_j^* h_j(\boldsymbol{x}^*) = 0 \ \forall\, j$.

Proof.(continued)

$$\begin{aligned}
\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x}^*) \\
&= f(\boldsymbol{x}^*) \\
&= \theta(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \\
&= \min_{x \in X} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \\
\therefore \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &\le \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \quad \ldots (1)
\end{aligned}$$

Also,

$$\begin{aligned}
\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &= f(\boldsymbol{x}^*) \\
&\ge f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i e_i(\boldsymbol{x}^*) \\
\therefore \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) &\ge \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \quad \ldots (2)
\end{aligned}$$

From (1) and (2), $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a Lagrangian saddle point. $\square$

How to find a saddle point if it exists?

Consider the problem (**NLP**):

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m \\ & \boldsymbol{x} \in X \end{array}$$

Theorem

Let f and h_j's be continuously differentiable convex functions, $e_i(\boldsymbol{x}) = a_i^T \boldsymbol{x} - b_i\ \forall\ i$ and X be a convex set. Assume that Slater's condition holds. Then,

$(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ *is a KKT point* $\Rightarrow$ $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ *is a Lagrangian saddle point.*

If $\boldsymbol{x}^*$ is primal feasible, $\boldsymbol{x}^* \in int(X)$, $\boldsymbol{\lambda}^*$ is dual feasible and $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a Lagrangian saddle point, then $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a KKT point.

Proof.

$\boldsymbol{x}^*$ is primal feasible. $\therefore h_j(\boldsymbol{x}^*) \leq 0 \ \forall\, j$ and $e_i(\boldsymbol{x}^*) = 0 \ \forall\, i$.
$(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a KKT point. Therefore,

$$\begin{aligned}
\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* \nabla e_i(\boldsymbol{x}^*) &= \boldsymbol{0} \\
\lambda_j^* h_j(\boldsymbol{x}^*) &= 0 \ \forall\, j \\
\lambda_j^* &\geq 0 \ \forall\, j
\end{aligned}$$

f is convex. Therefore, for all $\boldsymbol{x} \in X$,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^T (\boldsymbol{x} - \boldsymbol{x}^*). \qquad \ldots (3)$$

Similarly, since every h_j is convex,

$$h_j(\boldsymbol{x}) \geq h_j(\boldsymbol{x}^*) + \nabla h_j(\boldsymbol{x}^*)^T (\boldsymbol{x} - \boldsymbol{x}^*). \qquad \ldots (4)$$

Every e_i is an affine function. Therefore,

$$e_i(\boldsymbol{x}) = e_i(\boldsymbol{x}^*) + \nabla e_i(\boldsymbol{x}^*)^T (\boldsymbol{x} - \boldsymbol{x}^*). \qquad \ldots (5)$$

Proof. (continued)

Multiplying (4) by λ_j^* and (5) by μ_i^*, adding and using KKT conditions,

$$f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x})$$

$$\geq f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x}^*)$$

$$\therefore \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \geq \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \quad \ldots (6)$$

Also,

$$\begin{aligned} f(\boldsymbol{x}^*) &= f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x}^*) \\ &\geq f(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i e_i(\boldsymbol{x}^*) \end{aligned}$$

$$\therefore \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \geq \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \quad \ldots (7)$$

Therefore, $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a Lagrangian saddle point.

Proof.(continued)

(b) $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a Lagrangian saddle point, where $\boldsymbol{x}^*$ is primal feasible, $\boldsymbol{x}^* \in int(X)$ and $\boldsymbol{\lambda}^*$ is dual feasible. Therefore,

$$\begin{array}{l} h_j(\boldsymbol{x}^*) \leq 0 \;\forall\, j \\ e_i(\boldsymbol{x}^*) = 0 \;\forall\, i \end{array} \quad \text{and } \lambda_j^* \geq 0 \;\forall\, j \qquad \ldots (8)$$

and

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*).$$

$$\therefore \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i e_i(\boldsymbol{x}^*) \leq \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x}^*)$$

$$\therefore \lambda_j^* h_j(\boldsymbol{x}^*) = 0 \;\forall\, j \qquad \ldots (9)$$

Also, $\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$.

$$\therefore \boldsymbol{x}^* = \underset{x \in X}{\text{argmin}}\; \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

Proof.(continued)

$$\therefore \boldsymbol{x}^* = \underset{x \in X}{\operatorname{argmin}} \ \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

Note that,

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j^* h_j(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i^* e_i(\boldsymbol{x}).$$

$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a *convex* function of $\boldsymbol{x}$ (since f and h_j's are convex functions, e_i's are affine functions and $\lambda_j^* \geq 0$). Further, $\boldsymbol{x}^* \in int(X)$.

$$\therefore \nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0} \quad \ldots (10)$$

Therefore, from (8), (9) and (10), we see that $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a KKT point. $\square$

Consider the convex programming problem (**CP**):

$$
\begin{array}{ll}
\min & f(\boldsymbol{x}) \\
\text{s.t.} & h_j(\boldsymbol{x}) \leq 0, \; j = 1, \ldots, l \\
& e_i(\boldsymbol{x}) = 0, \; e_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{x} - b_i, \; i = 1, \ldots, m \\
& \boldsymbol{x} \in \mathbb{R}^n
\end{array}
$$

where f and h_j's are continuously differentiable convex functions. Assume that Slater's condition holds.

$$
\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i e_i(\boldsymbol{x})
$$

$$
\text{Dual Problem}: \quad \max_{\substack{\lambda \geq 0 \\ \mu}} \; \min_{x \in \mathbb{R}^n} \; \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})
$$

which is the **Wolfe Dual** of **CP**:

$$
\begin{array}{ll}
\max\limits_{\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}} & \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\
\text{s.t.} & \nabla_x \, \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0 \\
& \boldsymbol{\lambda} \geq \boldsymbol{0}
\end{array}
$$

Example:

$$\begin{array}{ll} \min & (x-2)^2 \\ \text{s.t.} & 2x+1 \leq 0 \\ & x \in [-1, 1] \end{array}$$

- Convex Programming Problem
- Slater's condition holds
- $x^* = -\frac{1}{2},\ p^* = f(x^*) = \frac{25}{4}$
- Dual function: $\theta(\lambda) = \min_{x \in [-1,1]} (x-2)^2 + \lambda(2x+1)$

The Wolfe dual problem is:

$$\begin{array}{ll} \max & -\lambda^2 + 5\lambda \\ \text{s.t.} & \lambda \in [1, 3] \end{array}$$

Solution: $\lambda^* = \frac{5}{2}$
Optimal Dual Objective Value, $d^* = \frac{25}{4} = p^*$

Example:
Consider the problem:

$$\begin{array}{ll} \min & x_1^2 + x_2^2 + \ldots + x_n^2 \\ \text{s.t.} & x_1 + x_2 + \ldots + x_n = 1 \end{array}$$

- Convex programming problem
- Slater's condition holds
- $\boldsymbol{x}^* = (\frac{1}{n}, \ldots, \frac{1}{n})^T$, $f(\boldsymbol{x}^*) = \frac{1}{n}$
- $\mathcal{L}(\boldsymbol{x}, \mu) = x_1^2 + \ldots + x_n^2 + \mu(x_1 + \ldots + x_n - 1)$
- $\nabla_x \mathcal{L}(\boldsymbol{x}, \mu) = \boldsymbol{0} \Rightarrow x_i = -\frac{\mu}{2} \ \forall \ i$

Wolfe dual problem:

$$\left.\begin{array}{c} \max \ \mathcal{L}(\boldsymbol{x}, \mu) \\ \text{s.t.} \ \nabla_x \mathcal{L}(\boldsymbol{x}, \mu) = 0 \end{array}\right\} \equiv \max_{\mu \in \mathbb{R}} \ -\frac{n}{4}\mu^2 - \mu$$

Solution to the dual problem: $\mu^* = -\dfrac{2}{n} \ \Rightarrow \ x_i^* = \dfrac{1}{n} \ \forall \ i$

Example: Consider the *Linear Program* (**LP**),

$$\begin{array}{ll} \min & \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{array}$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $rank(\boldsymbol{A}) = m < n$.

- Convex programming problem
- Slater's condition holds
- $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\mu}^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) - \boldsymbol{\lambda}^T\boldsymbol{x}$
- $\nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{0} \ \Rightarrow\ \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{\mu} - \boldsymbol{\lambda} = \boldsymbol{0}$

Wolfe dual problem(**Dual-LP**):

$$\left.\begin{array}{c} \max\ \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.}\ \nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0 \\ \boldsymbol{\lambda} \geq \boldsymbol{0} \end{array}\right\} \equiv \begin{array}{ll} \max & \boldsymbol{b}^T\boldsymbol{\mu} \\ \text{s.t.} & \boldsymbol{A}^T\boldsymbol{\mu} \leq \boldsymbol{c} \end{array}$$

The dual of **Dual-LP** is LP!

Example: Consider the *Quadratic Program*,

$$\begin{array}{ll} \min & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \end{array}$$

where $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ is a symmetric positive semi-definite matrix and $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\text{rank}(\boldsymbol{A}) = m$.
$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\lambda}^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})$
$\nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = 0 \;\Rightarrow\; \boldsymbol{H}\boldsymbol{x} + \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{\lambda} = \boldsymbol{0}$
Therefore, the **Wolfe dual problem** is,

$$\begin{array}{ll} \max & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\lambda}^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{H}\boldsymbol{x} - \boldsymbol{A}^T\boldsymbol{\lambda} = -\boldsymbol{c} \\ & \boldsymbol{\lambda} \geq \boldsymbol{0}. \end{array}$$

The dual problem cannot be given explicitly in terms of dual variables.

Example: Consider the *Quadratic Program*,

$$\begin{array}{ll} \min & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \end{array}$$

where $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix.
$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\lambda}^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})$
$\nabla_x \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = 0 \;\Rightarrow\; \boldsymbol{H}\boldsymbol{x} + \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{\lambda} = \boldsymbol{0}$
Therefore, the **Wolfe dual problem** is,

$$\begin{array}{ll} \max & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\lambda}^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{H}\boldsymbol{x} + \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{\lambda} = \boldsymbol{0} \\ & \boldsymbol{\lambda} \geq \boldsymbol{0}. \end{array}$$

Using $\boldsymbol{x} = \boldsymbol{H}^{-1}(\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{c})$, the dual problem is,

$$\max_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \; -\frac{1}{2}\boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{H}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda} + (\boldsymbol{A}\boldsymbol{H}^{-1}\boldsymbol{c} + \boldsymbol{b})^T\boldsymbol{\lambda}$$

Example:

$$\begin{aligned} \min \quad & \textstyle\sum_{i=1}^{n} x_i \log(\frac{x_i}{c_i}) \\ \text{s.t.} \quad & \boldsymbol{Ax} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

where $c_i > 0 \; \forall \, i, \; \boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $m \ll n$.

- Convex programming problem
- Slater's condition holds

The Wolfe dual problem is:

$$\max_{\boldsymbol{\mu} \in \mathbb{R}^m} \; -\sum_i c_i \exp\{(\boldsymbol{A}^T \boldsymbol{\mu})_i - 1\} + \boldsymbol{b}^T \boldsymbol{\mu}$$

Consider the problem (**NLP**):

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & h_j(\boldsymbol{x}) \leq 0,\ j = 1, \ldots, l \\ & e_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, m \\ & \boldsymbol{x} \in X \ \text{ where } X \text{ is a compact set.} \end{array}$$

- Dual Function:

$$\theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{x \in X} f(\boldsymbol{x}) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i e_i(\boldsymbol{x})$$

- Dual function is a pointwise minimum of a family of affine functions of $(\boldsymbol{\lambda}, \boldsymbol{\mu})$.
  $\therefore\ \theta(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is a *concave* function.
-

$$\begin{array}{ll} \max & \theta(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} & \boldsymbol{\lambda} \geq \boldsymbol{0} \end{array}$$

Therefore, the dual problem is a convex programming problem even if the primal problem is not!