CSE203B Convex Optimization: Chapter 5 Duality

CK Cheng
Dept. of Computer Science and Engineering
University of California, San Diego

Chapter 5 Duality

- Primal and Dual Problem Mechanism
  - Primal Problem
  - Lagrangian Function
  - Lagrange Dual Problem
- Examples (Primal Dual Conversion Procedure) Examples
  - Linear Programming
  - Quadratic Programming
  - Conjugate Functions (Duality)
  - Entropy Maximization
- Interpretation (Duality) Theory
  - Saddle-Point Interpretation
  - Geometric Interpretation
  - Slater's Condition
  - Shadow-Price Interpretation
- KKT Conditions (Optimality Conditions) Optimality
- Sensitivity (Shadow-Price)
- Generalized Inequalities

Duality

<u>Primal Problem (Feasible Solution)</u>

$$\left.\begin{array}{ll} \min\ f_0(x) & x \in R^n \\ s.t.\ f_i(x) \le 0 & i = 1, \dots, m \\ \quad h_i(x) = 0 & i = 1, \dots, p \end{array}\right] \text{domain } D = dom\ f_0 \cap_i\ dom\ f_i\ \cap_i dom\ h_i$$

feasible set $E = \{x \mid f_i(x) \le 0, h_i(x) = 0\ \forall i\ \forall x \in D\}$

$E \subset D$

Opt: $x^*, p^* = f_0(x^*)$

Lagrangian: $L: R^n \times R^m \times R^p \to R$

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p v_i h_i(x)$$

λ_i, v_i: *Lagrange multiplier*, $\lambda_i, \in R_+, v_i \in R$.

Lagrange dual function

$g(\lambda, v) = \inf_{x \in D} L(x, \lambda, v)$ (x may not be feasible)

$$g(\lambda, \nu) = \min_{x \in D} L(x, \lambda, \nu)$$

$$= -\max_{x \in D} -L(x, \lambda, \nu)$$

$$= -\max_{x \in D} \underbrace{\underbrace{\underbrace{-f_0(x) - \sum \lambda_i f_i(x) - \sum \nu_i h_i(x)}_{\text{convex w.r.t } \lambda_i, \nu_i}}_{\text{convex w.r.t } \lambda_i, \nu_i}}_{\text{concave w.r.t } \lambda_i, \nu_i}$$

Duality

<u>Dual Problem (Infeasible Solution)</u>

$max_{\lambda, v}\ g(\lambda, v)\quad s.t. \lambda \succcurlyeq 0$

1. $g(\lambda, v)$ *is concave*
2. $g(\lambda, v) \le p^*$ an optimal value where $\lambda \succcurlyeq 0$

Proof 1: By definition of $g(\lambda, v)$ and the convexity of pointwise max operation on convex functions.

Proof 2: For any feasible $\tilde{x}$ and $\lambda \succcurlyeq 0$

$f_0(\tilde{x}) \ge L(\tilde{x}, \lambda, v)$ (Because $\sum \lambda_i f_i(\tilde{x}) + \sum v_i h_i(\tilde{x}) \le 0$)

$L(\tilde{x}, \lambda, v) \ge g(\lambda, v)$ by definition of $g(\lambda, v)$

Thus $p^* = f_0(x^*) \ge g(\lambda, v)$

$$\min_{x} \max_{\lambda, \nu} L(x, \lambda, \nu) \ge \max_{\lambda, \nu} \min_{x} L(x, \lambda, \nu)$$

Example: Linear Programming

$\nabla f_0(x) = c$

Prime: $\min c^T x$

$$\text{subject to} \begin{cases} Ax \preccurlyeq b \\ x \succcurlyeq 0 \end{cases} \Rightarrow \begin{cases} Ax - b \preccurlyeq 0 \\ -x \preccurlyeq 0 \end{cases}$$

$K_1 = \{u \mid Au \leq 0\}$

$= K_1 = \{u \mid -Au \geq 0\}$

Lagrangian

$$L(x,\lambda) = c^T x + \lambda_I^T (Ax - b) - \lambda_{II}^T x$$
$$= -\lambda_I^T b + (A^T\lambda_I - \lambda_{II} + c)^T x, \quad \lambda_I, \lambda_{II} \succcurlyeq 0$$
$$g(\lambda) = \inf_x L(x,\lambda)$$

$K_1^* = \{-A^T\theta \mid \theta \geq 0\}$

$$g(\lambda) = \begin{cases} -b^T\lambda_I, & A^T\lambda_I + c \geq 0 \; (A^T\lambda_I - \lambda_{II} + c = 0) \rightarrow A^T\lambda_I + c = \lambda_{II} \geq 0 \\ -\infty, & otherwise \quad (A^T\lambda_I - \lambda_{II} + c \neq 0) \end{cases}$$

$c = -A^T\theta \quad \theta \geq 0$

Dual:

$C + A^T\theta = 0$

$$\max \; -\mathrm{b}^T\lambda_I \; (\min b^T\lambda_I)$$
$$\text{subject to } A^T\lambda_I + c \geq 0$$
$$\lambda_I \geq 0$$

$K_2 = \{u \mid +Iu \geq 0\}$

$K_2^* = \{-I\theta \mid \theta \geq 0\}$

Example: Linear Programming

Prime: $\min[-1 \; -1]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

subject to $\begin{bmatrix} 1 & 3 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \preccurlyeq \begin{bmatrix} 3 \\ 2 \end{bmatrix}$, $x_1, x_2 \geq 0$.

Dual: $max -[3 \; 2]\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$

subject to $\begin{bmatrix} 1 & 1 \\ 3 & 0 \end{bmatrix}\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix} \succcurlyeq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\lambda_1, \lambda_2 \geq 0$

Results: $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1/3 \end{bmatrix}$, $p^* = -\frac{7}{3}$

$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 2/3 \end{bmatrix}$, $d^* = -\frac{7}{3}$

Example: Linear Programming

$\nabla f_0(x) = c$
$-x \leq 0$
$Ax - b = 0$

$$\min c^T x$$
$$\text{subject to } Ax = b,\ x \succcurlyeq 0,\ (\text{or } -x \leq 0)$$

Lagrangian: $L(x, \lambda, v) = c^T x + \lambda^T(-x) + v^T(Ax - b)$
$$= -b^T v + (c + A^T v - \lambda)^T x$$

Lagrange Dual: $g(\lambda, v) = \inf_x L(x, \lambda, v) \Rightarrow \nabla_x L = 0 \Rightarrow (c + A^T v - \lambda) = 0$

1. If $A^T v - \lambda + c = 0 \quad \rightarrow \quad g(\lambda, v) = -b^T v$
2. Else $\quad \rightarrow \quad g(\lambda, v) = -\infty$

Properties:

1. g is linear on affine domain $\{(\lambda, v) | A^T v - \lambda + c = 0\}$, hence concave.
2. If $\lambda \succcurlyeq 0 \Rightarrow A^T v + c \succcurlyeq 0$
   $p^* \geq -b^T v \quad if\ A^T v + c \succcurlyeq 0$

$\max -b^T v$ $A^T v + c \succcurlyeq 0$	**or**	$\max b^T v$ $A^T v \preccurlyeq c$

$v \in R^m$

$v \rightarrow -v$
change sign of v

Example: Quadratic Programming

$A \in R^{m \times n}$
Rank$(A) = m$
(Assumption)

$$\min x^T x$$
$$\text{subject to } Ax = b \quad \Big|\ Ax - b = 0$$

Lagrangian:
$$L(x, v) = x^T x + v^T(Ax - b) = x^T x + v^T A x - v^T b$$

To minimize L over x, we set $\nabla_x L(x, v) = 2x + A^T v = 0$
$\Rightarrow x = -\frac{1}{2} A^T v$ (1)

Lagrange Dual Function:

$$= \left(-\tfrac{1}{2} A^T v\right)^T \left(-\tfrac{1}{2} A^T v\right) + v^T A \left(-\tfrac{1}{2} A^T v\right) - v^T b$$

$$g(v) = L\left(x = -\frac{1}{2} A^T v, v\right) = -\frac{1}{4} v^T A A^T v - b^T v$$

$$= \frac{1}{4} v^T A A^T v - \frac{1}{2} v^T A A^T v - v^T b$$
$$= -\frac{1}{4} v^T A A^T v - v^T b$$

(A concave function of v)

Lower Bound Property: $p^* \geq -\frac{1}{4} v^T A A^T v - b^T v, \quad \forall v$

$\nabla_v g(v) = -\frac{1}{2} A A^T v - b = 0$

To maximize $g(v)$, we set $v = -2(AA^T)^{-1} b$

$v = -2(AA^T)^{-1} b$

Thus, we have $g(v) = -\frac{1}{4} v^T A A^T v - b^T v = b^T (AA^T)^{-1} b$ (2)

(3) From (1) and (2), we have $\begin{cases} x^* = A^T (AA^T)^{-1} b \\ p^* = {x^*}^T x^* = b^T (AA^T)^{-1} b \end{cases}$

$g(v)\big|_{v = -2(AA^T)^{-1}b} = -\frac{1}{4}\left[(-2)(AA^T)^{-1} b\right]^T AA^T \left(-2(AA^T)^{-1} b\right) - b^T\left(-2(AA^T)^{-1}b\right)$
$= -b^T (AA^T)^{-1} b + 2 b^T (AA^T)^{-1} b$

Example: Quadratic Program

Quadratic Program

$$\begin{aligned} &min\ x^T P x \qquad P \in S^n_{++} \\ &s.t.\ Ax \preccurlyeq b \longrightarrow Ax - b \leq 0 \end{aligned}$$

Lagrange Dual Function:

$$g(\lambda) = \min_x L(x,\lambda) = x^T P x + \lambda^T (Ax - b) \qquad \nabla_x L = 2Px + A^T\lambda$$

$$= -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda,\ \lambda \succcurlyeq 0. \qquad x = -(2P)^{-1}A^T\lambda$$

Dual Problem:

$$\begin{aligned} &\max -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ &s.t.\ \lambda \succcurlyeq 0 \end{aligned}$$

Example: Quadratic Program (nonconvex prob.)

$$\begin{aligned} &min\ x^T A x + 2b^T x \\ &s.t.\ x^T x \leq 1 \qquad A \in S^n, A \not\succcurlyeq 0 \end{aligned}$$

Dual Function:

$$g(\lambda) = \min_x L(x,\lambda) = x^T (A + \lambda I) x + 2b^T x - \lambda$$

Unbounded below if $A + \lambda I \not\succcurlyeq 0$ or if $A + \lambda I \succcurlyeq 0$ & $b \notin R(A + \lambda I)$

Minimized by $x = -(A + \lambda I)^+ b$

Otherwise $g(\lambda) = -b^T (A + \lambda I)^+ b - \lambda$

Dual Problem:

$$\begin{aligned} &\boldsymbol{max}\ -b^T (A + \lambda I)^+ b - \lambda \\ &\boldsymbol{s.t.}\ \ A + \lambda I \succcurlyeq 0 \\ &\qquad b \in R(A + \lambda I) \end{aligned} \qquad \Bigg| \qquad \begin{aligned} &\boldsymbol{max}\ -t - \lambda \\ &\boldsymbol{s.t.}\ \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succcurlyeq 0 \end{aligned}$$

$$\begin{bmatrix} I & 0 \\ -((A + \lambda I)^+ b)^T & I \end{bmatrix} \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \begin{bmatrix} I & -(A + \lambda I)^+ b \\ 0 & I \end{bmatrix} \succcurlyeq 0$$

$$\begin{bmatrix} A + \lambda I & 0 \\ 0 & -b^T (A + \lambda I)^+ b + t \end{bmatrix} \succcurlyeq 0$$

Example: Discrete Problem

Two-Way Partitioning Problem

Primal:

$$\begin{aligned} &\min x^T W x && x \in R^n, w_{ij} \in R \\ &s.t.\ \ x_i^2 = 1 && i = 1, \dots, m \\ &i.e.\ \ x_i \in \{-1, 1\},\ x^T W x = \textstyle\sum_{ij} x_i x_j w_{ij} \end{aligned}$$

Not a convex opt problem (Partitioning is an NP complete problem)

We can assume that

$$W \in S^n \quad (x^T W x = \frac{1}{2} x^T W x + \frac{1}{2} x^T W^T x = \frac{1}{2} x^T (W + W^T) x)$$

Lagrangian:

$$L(x, v) = x^T W x + \sum_{i=1}^{n} v_i (x_2^2 - 1) = x^T \big(W + diag(v)\big) x - I^T v$$

Lagrange dual function:

$$g(v) = \inf_x x^T \big(W + diag(v)\big) x - I^T v = \begin{cases} -I^T v, & W + diag(v) \succcurlyeq 0 \\ -\infty, & otherwise \end{cases}$$

Example: Discrete Problem

Dual:

$$\begin{aligned} &\max g(v) = -I^T v \\ &s.t.\ W + diag(v) \succcurlyeq 0 \end{aligned}$$

Solution $\quad v = -\lambda_{min}(W) 1$

$p^* \geq d^* = -1^T v = n \lambda_{min}(W)$

Examples: Minimum Volume Covering Ellipsoid

$$\min f_0(x) = \log\det X^{-1}, \; X \in S_{++}^n$$
$$s.t.\; a_i^T X a_i \le 1, i = 1, \dots, m$$

<u>Lagrangian</u>

$$L(x, \lambda) = logdetX^{-1} + \sum_{i=1}^m \lambda_i a_i^T X a_i - 1^T\lambda, \; \lambda \in R_+^m$$

<u>Lagrange dual function</u>

$$g(\lambda) = \min_x L(x, \lambda), \lambda \in R_+^m$$

<u>Dual Problem</u>

$$\max \; logdet\left(\sum_{i=1} \lambda_i a_i a_i^T\right) - 1^T\lambda + n$$
$$s.t. \; \sum_{i=1} \lambda_i a_i a_i^T \succ 0, \lambda \in R_+^m$$

Interpretation: Saddle-point

x, λ, ν

$$\max_{z\in Z} \min_{w\in W} f(w,z) \le \min_{w\in W} \max_{z\in Z} f(w,z)$$

$\max_{z\in Z} g(z)$

$g(z) = \min_w f(w,z)$

$$\text{Example}: f(w,z) \quad w = \begin{array}{c c} 1 & \\ 2 & \\ 3 & \\ & \end{array}\!\!\begin{array}{c} \begin{bmatrix} 1 & -1 & 3 \\ 2 & 2 & -1 \\ 3 & 1 & -2 \end{bmatrix} \\ z = 1, \;\; 2, \;\; 3 \end{array}$$

max min
z w

$$\text{I. } z \text{ decides first} \begin{cases} \min_{w\in W} f(w,1) = 1 \\ \min_{w\in W} f(w,2) = -1 \\ \min_{w\in W} f(w,3) = -2 \end{cases} \max_{z\in Z} \min_{w\in W} f(w,z) = 1$$

min max
w z

$$\text{II. w decides first} \begin{cases} \max_{z\in Z} f(1,z) = 3 \\ \max_{z\in Z} f(2,z) = 2 \\ \max_{z\in Z} f(3,z) = 3 \end{cases} \min_{w\in W} \max_{z\in Z} f(w,z) = 2$$

Examples: Conjugate Function

$$\min f_0(x)$$
$$s.t.\ Ax \preccurlyeq b$$
$$Cx = d$$

Dual function

$$g(\lambda, v) = \inf_{x \in dom f_0} (f_0(x) + \lambda^T(Ax - b) + v^T(Cx - d))$$
$$= \inf_{x \in dom f_0} (f_0(x) + (A^T\lambda + C^Tv)^T x - b^T\lambda - d^Tv)$$
$$= -f_0^*(-A^T\lambda - C^Tv) - b^T\lambda - d^Tv$$

Conjugate function

Where $f_0^*(y) = \max_{x \in dom f} y^T x - f_0(x)$

$\min c^T x$ $s.t.\ \ Ax = b$ $x \succcurlyeq 0$ $\max\ -b^T v$ $s.t.\ \ A^T v + c \succcurlyeq 0$	$\min c^T x$ $s.t.\ Ax \preccurlyeq b$ $\max\ -b^T \lambda$ $s.t.\ \ A^T\lambda + c = 0$ $\lambda \succcurlyeq 0$

Examples: Entropy Maximization

$$\min f_0(x) = \sum_{i=1}^n x_i log x_i, \quad x \in R_{++}^n$$
$$s.t.\ Ax \preccurlyeq b$$
$$1^T x = 1$$

Since $f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}, y_i \in R$

Thus, the dual function is

$$g(\lambda, v) = -b^T\lambda - v - \sum_{i=1}^n e^{-a_i^T\lambda - v - 1}$$
$$= -b^T\lambda - v - e^{-v-1}\sum_{i=1}^n e^{-a_i^T\lambda}, \ \ a_i\text{: the i}^{\text{th}}\text{column of A.}$$

To maximize $g(\lambda, v)$, we set $v = log \sum_{i=1}^n e^{-a_i^T\lambda} - 1$

Dual Problem

$$\max\ \ -b^T\lambda - \log(\sum_{i=1}^n e^{-a_i^T\lambda})$$
$$s.t.\ \ \ \lambda \succcurlyeq 0$$

Interpretation: Saddle-point

$\min_{x} \max_{\lambda,\nu}$ $\max_{\lambda,\nu} \min_{x} L(x,\lambda,\nu)$ $f(w,z)$ $\min_{w} \max_{z}$

Claim : Result of II $\geq$ Result of I

Given an arbitrary pair $(\tilde{w}, \tilde{z}) \in D$

$$\min_{w \in W} f(w, \tilde{z}) \leq f(\tilde{w}, \tilde{z}) \leq \max_{z \in Z} f(\tilde{w}, z) \quad \forall \tilde{w}, \tilde{z} \in D$$

$$\min_{w \in W} f(w, \tilde{z}) \leq \max_{z \in Z} f(\tilde{w}, z)$$

Thus $\max_{z \in Z} \min_{w \in W} f(w, z) \leq \min_{w \in W} \max_{z \in Z} f(w, z)$

Interpretation: Saddle-point

Example : $f(w,z)$ $w = \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix}$

$z = 1, \quad 2, \quad 3$

$\min_{w \in W} f(w, 1) = 1$

$\min_{w \in W} f(w, 2) = 1$

$\min_{w \in W} f(w, 3) = 1$

$\max_{z \in Z} \min_{w \in W} f(w, z) = 1$

$\max_{z \in Z} f(1, z) = 1$

$\max_{z \in Z} f(2, z) = 2$

$\max_{z \in Z} f(3, z) = 3$

$\min_{w \in W} \max_{z \in Z} f(w, z) = 1$

Interpretation: Saddle-point

$$\text{Example}: f(w,z) \quad w = \begin{array}{c} 1 \\ 2 \\ 3 \end{array}\begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix}$$
$$z = 1, \quad 2, \quad 3$$

$$\min_{w\in W} f(w,1) = 1 \qquad \max_{z\in Z} f(1,z) = 3$$
$$\min_{w\in W} f(w,2) = 1 \qquad \max_{z\in Z} f(2,z) = 3$$
$$\min_{w\in W} f(w,3) = 1 \qquad \max_{z\in Z} f(3,z) = 3$$

$$\max_{z\in Z}\min_{w\in W} f(w,z) = 1 \qquad \min_{w\in W}\max_{z\in Z} f(w,z) = 3$$

Interpretation: Saddle-point

$$\text{Example}: f(w,z) \quad w = \begin{array}{c} 1 \\ 2 \\ 3 \end{array}\begin{bmatrix} 4 & 6 & 3 \\ 2 & 3 & 1 \\ 3 & 5 & 2 \end{bmatrix}$$
$$z = 1, \quad 2, \quad 3$$