CSE203B Convex Optimization: Chapter 5 Duality

Project: Problem Statement
Convexity
Duality

One key concept of the textbook

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

Extension of dual cones and conjugate functions

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) Concepts + 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>

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

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

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

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

convert constraints to costs

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

Lagrange dual function or shadow price

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

Not relevant to x.

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

Example: Linear Programming

Prime: $min\ c^T x$

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

Lagrangian

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

-c falls into the dual cone of feasible region

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

$A^T\lambda_1 + c = \lambda_{II} \geq 0$

Dual:

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

$max\ -b^T\lambda_I$ → $min\ b^T\lambda_I$

$A^T\lambda_1 \geq -c$ → $A^T\lambda_1 \geq -c$

$\lambda_1 \geq 0$ → $\lambda_1 \geq 0$

W6A

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

$K = \{x \mid Ax \leq 0\}$

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

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

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: $g(\lambda, v) = \inf_x L(x, \lambda, v)$

$A^T v - \lambda = -c$

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 $\{(\lambda, v) \mid A^T v - \lambda + c = 0\}$, hence concave.
2. If $\lambda \succcurlyeq 0 \Rightarrow A^T v + c \succcurlyeq 0$
   $p^* \geq -b^T v \ \ if\ A^T v + c \succcurlyeq 0$

$-v$

$$\boxed{\begin{aligned} &\max\ -b^T v \\ &\quad A^T v + c \succcurlyeq 0 \end{aligned}} \quad \textbf{or} \quad \boxed{\begin{aligned} &\max\ b^T v \\ &\quad A^T v \preccurlyeq c \end{aligned}}$$

Example: Quadratic Programming

$$\min x^T x$$
$$\text{subject to } Ax = b$$

Lagrangian:
$$L(x, v) = x^T x + v^T(Ax - b) \rightarrow \left(-\tfrac{1}{2}A^T v\right)^T\left(-\tfrac{1}{2}A^T v\right) + v^T\left(A\left(-\tfrac{1}{2}A^T v\right) - b\right)$$

To minimize L over x, we set $\nabla_x L(x, v) = 2x + A^T v = 0$
$$\Rightarrow x = -\tfrac{1}{2}A^T v \ \ (1) \rightarrow -\tfrac{1}{2}A^T\left[-2(AA^T)^{-1}b\right] = A^T(AA^T)^{-1}b$$

Lagrange Dual Function:
$$g(v) = L\left(x = -\frac{1}{2}A^T v, v\right) = -\frac{1}{4}v^T AA^T v - b^T v, \quad \nabla g(v) = -\tfrac{1}{2}AA^T v - b = 0$$

(A concave function of v)

$$-\tfrac{1}{2}AA^T v = b$$

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

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

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

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

A feasible solution

A lower bound

Example: Quadratic Program

Quadratic Program

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

Lagrange Dual Function:

$$\begin{aligned} g(\lambda) &= \min_x x^T P x + \lambda^T (Ax - b) \\ &= -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda, \ \lambda \geqslant 0 \end{aligned}$$

$\nabla L(x,\lambda) = 2Px + A^T\lambda = 0$

$x = -\frac{1}{2} P^{-1} A^T \lambda$

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.)

$$\begin{aligned} &min\ x^T A x + 2b^T x \\ &s.t.\ x^T x \leq 1 \qquad A \in S^n, A \nsucccurlyeq 0 \end{aligned}$$

<u>Dual Function:</u>

$L(x,\lambda) = x^T A x + 2b^T x + \lambda(x^T x - 1)$

$= x^T A x + 2b^T x + \lambda(x^T I x - 1)$

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

Unbounded below if $A + \lambda I \nsucccurlyeq 0$ or if $A + \lambda I \succcurlyeq 0$ & $b \notin R(A + \lambda I)$

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

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

<u>Dual Problem:</u>

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

$$\begin{aligned} &\boldsymbol{max}\ -t - \lambda \\ &\boldsymbol{s.t.}\ \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succcurlyeq 0 \end{aligned}$$

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

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

Example: Discrete Problem

Two-Way Partitioning Problem

Primal:

$$\begin{array}{ll} min\ x^TWx & x \in R^n, w_{ij} \in R \\ s.t.\ \ x_i^2 = 1 & i = 1, \dots, m \end{array}$$

$$i.e.\ \ x_i \in \{-1, 1\},\ x^TWx = \sum_{ij} x_i x_j w_{ij}$$

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

We can assume that

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

Lagrangian:

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

Lagrange dual function:

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

Example: Discrete Problem

Dual:

$$\begin{array}{l} max\ g(v) = -I^Tv \\ s.t.\ W + diag(v) \succcurlyeq 0 \end{array}$$

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

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