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

Duality

Primal Problem (Feasible Solution)

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

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

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

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

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

Lagrange dual function

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

Duality

Dual Problem (Infeasible Solution)

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

1. $g(\lambda, v)\ is\ concave$
2. $g(\lambda, v) \leq p^*$ 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 $\lambda \succcurlyeq 0$

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

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

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

Example: Linear Programming

Prime: $\min c^T x$

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

Lagrangian

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

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

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

Dual:

$$\max \; -\mathrm{b}^{\mathrm{T}} \lambda_I \; (\min b^T \lambda_I)$$

$$\text{subject to } A^T \lambda_I + c \geq 0$$

$$\lambda_I \geq 0$$

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: $\mathrm{m}ax -[3\ 2]\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$

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

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

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

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**	$\max b^T v$ $A^T v \preccurlyeq c$

Example: Quadratic Programming

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

Lagrangian:

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

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:

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

A concave function of v

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

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

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 $\begin{cases} x^* = A^T (AA^T)^{-1} b \\ p^* = {x^*}^T x^* = b^T (AA^T)^{-1} b \end{cases}$

Example: Quadratic Program

Quadratic Program

$$\min x^T P x \qquad P \in S_{++}^n$$
$$s.t.\ Ax \preccurlyeq b$$

Lagrange Dual Function:

$$\mathrm{g}(\lambda) = \min_x L(x,\lambda) = x^T P x + \lambda^T (Ax - b)$$
$$= -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda,\ \lambda \succcurlyeq 0.$$

Dual Problem:

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

Example: Quadratic Program (nonconvex prob.)

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

<u>Dual Function:</u>

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

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

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

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

<u>Dual Problem:</u>

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

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

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

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

Example: Discrete Problem

Two-Way Partitioning Problem

Primal:

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

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

We can assume that

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

Lagrangian:

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

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:

$$\max g(v) = -I^T v$$
$$s.t.\ W + diag(v) \succcurlyeq 0$$

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

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

Examples: Conjugate Function

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

Dual function

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

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

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

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: the i[th] column of A.

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

<u>Dual Problem</u>

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

Examples: Minimum Volume Covering Ellipsoid

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

Lagrangian

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

Lagrange dual function

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

Dual Problem

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

Interpretation: Saddle-point

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

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

$$g(z) = \min_{w} f(w,z)$$

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

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

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

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

Interpretation: Saddle-point

Claim : Result of II $>$ Result of I

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

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

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

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

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

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

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

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

Interpretation: Saddle-point

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

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

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

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

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

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

Interpretation: Saddle-point

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

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

Saddle Point (ref: Von Neumann, Dantzig, Nash)

1. Definition

Given function $f(w, z)$,

$(\tilde{w}, \tilde{z})$ is a saddle point of $f(w, z)$

if $\max_{z} f(\tilde{w}, z) = f(\tilde{w}, \tilde{z})$

$min_w f(w, \tilde{z}) = f(\tilde{w}, \tilde{z})$

2. Theorem I

$$\max_{z} \min_{w} f(w, z) = \min_{w} \max_{z} f(w, z)$$

iff a saddle point of $f(w, z)$ exists.

Saddle Point: Theorem I Proof

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

iff a saddle point of $f(w,z)$ exists.

Proof: => Let

$$\tilde{w} = \arg_w \min_w \max_z f(w,z)$$

$$\tilde{z} = \arg_z \max_z \min_w f(w,z)$$

We have

$$f(\tilde{w},\tilde{z}) \le \max_z f(\tilde{w},z) = \min_w f(w,\tilde{z}) \le f(\tilde{w},\tilde{z})$$

By definition $(\tilde{w},\tilde{z})$ is a saddle point.

Saddle Point: Theorem I Proof

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

iff a saddle point of $f(w,z)$ exists.

Proof: <= Assume that $(\widetilde{w}, \tilde{z})$ is a saddle point.

We have

$$\max_z \min_w f(w,z) \geq \min_w f(w,\tilde{z}) = f(\widetilde{w},\tilde{z})$$

$$\min_w \max_z f(w,z) \leq \max_z f(\widetilde{w},z) = f(\widetilde{w},\tilde{z})$$

Thus, $\max_z \min_w f(w,z) = \min_w \max_z f(w,z)$

Saddle Point: 2-D Table Formulation

The row and column selection is formulated as a bilinear optimization problem

- $f(w, z) = \sum_i \sum_j a_{ij} w_i z_j$

Constraint I: row and column selection

- $w_i, z_j \in \{0,1\}, \sum_i w_i = 1, \sum_j z_j = 1.$

Constraint II: Relaxed discrete constraints

$\sum_i w_i = 1, \sum_j z_j = 1,\ w_i \geq 0, z_j \geq 0, \forall i, j$

Saddle Point: 2-D Table Formulation

Find a saddle point of $f(w,z)$ under constraint I.

Theorem II: A saddle point of 2-D table formulation can be solved if $f(w,z)$ is convex w.r.t. w, and concave w.r.t. z.

Saddle Point:

1. The optimization problem with relaxed constraints can be solved with algorithms (Dantzig)

- $\max_z \min_w f(w,z) = \min_w \max_z f(w,z)$

2. Since $f(w,z)$ is convex w.r.t. w, and concave w.r.t. z, the solution can be reduced to constraint I (row and column selection), $(\widetilde{w}, \tilde{z})$.

3. From 2, $(\widetilde{w}, \tilde{z})$ is a saddle point by definition.

Geometric Interpretation

$min\, f_o(x) \qquad (t)$
$s.t.\, f_1(x) \leq 0 \quad (u \leq 0)$

$g(\lambda) = \min_{(u,t)\in G} t + \lambda u \qquad G = \{(f_1(x), f_o(x)) | x \in D\}$

$g(\lambda) = \lambda u + t$
supporting hyperplane to G
that intersects t axis at $t = g(\lambda)$

u

Duality via Separating Hyperplane

Set $G = \{\left(f_1(x), \dots, f_m(x), h_1(x), \dots, h_p(x), f_o(x)\right) | \mathrm{x} \in D\}$,

$G \in R^m \times R^p \times R, p^* = \inf\{t | (u, w, t) \in g, u \leq 0, w = 0\}$

Lagrangian $L = (\lambda, v, 1)^T (u, w, t) = \sum_{i=1}^{m} \lambda_i u_i + \sum_{i=1}^{p} v_i w_i + t$

Dual Problem $g(\lambda, v) = \inf\{(\lambda, v, 1)^T (u, w, t) | (u, w, t) \in G\}$

Separating hyperplane: Example

$$\{(u, t) | f_o(x) \leq t, f_1(x) \leq u, \exists x \in D\}$$

$$\left(\tilde{\lambda}, \tilde{v}, \tilde{\mu}\right)^T (u, w, t) \geq \alpha, \qquad \forall (u, w, t) \in A$$

$$\left(\tilde{\lambda}, \tilde{v}, \tilde{\mu}\right)^T (u, w, t) \leq \alpha, \qquad \forall (u, w, t) \in B$$

Since $\tilde{\mu} \neq 0$, we can have $(\lambda, v, 1) = (\frac{\tilde{\lambda}}{\tilde{\mu}}, \frac{\tilde{v}}{\tilde{\mu}}, 1)$

$$A = \{(u, w, t) | \exists x \in D, f_i(x) \leq u_i, i = 1, \dots, m, \\ h_i(x) = w_i, i = 1, \dots, p, f_o(x) \leq t\}$$

Lagrange dual problem

$$\max g(\lambda, v)$$
$$s.t. \quad \lambda \geq 0$$

Properties

This is a convex problem.

The opt. solution is denoted as d^*

$p^* - d^* = gap \geq 0$

If $gap > 0$, it is a weak duality.

If $gap = 0$, it is a strong duality.

Slater's condition

relint: relative interior of set D

Given that the primal problem is convex,

If $f_i(x) < 0, i = 1, \ldots, m, \exists x \in relint\ D$

Then strong duality holds.

$$relint\ C = \{x \in C | B(x, r) \cap aff\ C \subseteq C\ for\ some\ r > 0\}$$

$$B(x, r) = \{y | \|y - x\| \leq r\}$$

↖ any norm

Shadow Price Interpretation: Food vs. Vitamin

Flour protein powder
Veg. vitamins A,B,D,E,K
Fruits minerals

Primal

$$\begin{aligned} &\min c^T x \\ &s.t.\ Ax \succcurlyeq b \\ &\qquad x \succcurlyeq 0 \end{aligned} \qquad \begin{aligned} &\min c^T x \\ &s.t.\ -Ax + b \preccurlyeq 0 \\ &\qquad -x \preccurlyeq 0 \end{aligned}$$

$$\min c_1 x_1 + c_2 x_2 + c_3 x_3 \qquad \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \succcurlyeq \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, x_{\mathrm{i}} \geq 0, \forall i$$

Dual

$$\begin{aligned} &\max \lambda^T b \\ &s.t.\ A^T \lambda \preccurlyeq c \\ &\qquad \lambda \succcurlyeq 0 \end{aligned}$$

$$\max \lambda_1 b_1 + \lambda_2 b_2 + \lambda_3 b_3$$

$$\text{s.t.} \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \preccurlyeq \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

Lagrangian

$$\begin{aligned} L(x, \lambda) &= c^T x + \lambda_I^T(-Ax + b) + \lambda_{II}^T(-x) \\ &= [c^T + \lambda_I^T(-A) - \lambda_2^T]x + \lambda_{II}^T b \\ c^T &= \lambda_I^T(A) + \lambda_{II}^T,\ or\ A^T \lambda_I \leq c \end{aligned}$$

Shadow Price Interpretation: Spring Energy & Force

$$min\, f_o(x_1, x_2) = \frac{1}{2}k_1 x_1^2 + \frac{1}{2}k_2(x_2 - x_1)^2 + \frac{1}{2}k_3(l - x_2)^2$$

f_o : potential energy $k_i > 0$: stiffness constant of spring *i*

$w/2 - x_1 \le 0$

$w + x_1 - x_2 \le 0$

$w/2 - l + x_2 \le 0$

$$min\, \frac{1}{2}(k_1 x_1^2 + k_2(x_2 - x_1)^2 + k_3(l - x_2)^2)$$

$\lambda_1 \quad w/2 - x_1 \le 0$

$\lambda_2 \quad w + x_1 - x_2 \le 0$

$\lambda_3 \quad w/2 - l + x_2 \le 0$

$\lambda_1(w/2 - x_1) = 0, \lambda_2(w - x_2 + x_1) = 0, \lambda_3(w/2 - l + x_2) = 0$

zero gradient condition

$$\begin{bmatrix} k_1 x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{bmatrix} + \lambda_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$

λ_i : contact forces between the walls & blocks

KKT (Karush-Kuhn-Tucker) Conditions

2. $f_i(x), i = 1, \dots, m,\ \ h_i(x),\ \ i = 1, \dots, p\ are\ differentiable$

1. Primal constraints : $f_i(x) \leq 0, i = 1, \dots, m.$
   $h_i(x) = 0, i = 1, \dots, p.$

2. Dual constraints : $\lambda \succcurlyeq 0$

3. Complementary slackness : $\lambda_i f_i(x) = 0, i = 1, \dots, m.$

4. Gradient of Lagrangian with respect to *x* variables
   $\nabla_x f_o(x) + \sum_{i=1}^{m} \lambda_i \nabla_x f_i(x) + \sum_{i=1}^{p} v_i \nabla_x h_i(x) = 0$

KKT (Karush-Kuhn-Tucker) Conditions

1. Primal, Lagrangian, and Dual

$\max_{\lambda,v} \min_{x}: (x, \lambda, v)$

$= \max_{\lambda,v} g(\lambda, v)$

$min\ f_o(x)$

$f_i \leq 0$

$h_i = 0$

$L(x, \lambda, v)$

$= f_o(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x), \lambda_i \geq 0$

$\min_{x} \max_{\lambda,v}$ $\quad$ $\max_{\lambda,v} \min_{x}$ ($dual$)

1. Feasibility (x, λ, v)
2. $L(x, \lambda, v) = f_o(x)$ $\begin{cases} \lambda_i > 0\ if\ f_i = 0 \\ \lambda_i = 0\ if\ f_i < 0 \end{cases}$
3. $g(\lambda, v) = \min_{x} L(x, \lambda, v) = f_o(x)$

Necessary condition for local optimality

Sufficient when the problem is convex & satisfy regularity conditions (Slater condition)

Sensitivity

Perturbed Problem

$$min\, f_o(x)$$
$$s.t.\ \ f_i \le u_i$$
$$h_i(x) = w_i$$

$$max\, \tilde{g} = g(\lambda, v) - u^T\lambda - w^T v$$
$$s.t.\quad \lambda \ge 0$$
$$p^*(u,w) = \max_{\lambda,v} g(\lambda, v) - u^T\lambda - w^T v$$

Unperturbed Problem

$$u_i = w_i = 0$$

$$\max g(\lambda, v)$$
$$s.t.\quad \lambda \ge 0$$
$$p^*(0,0), \lambda^*, v^*$$

$$p^*(u,w) \ge g(\lambda^*, v^*) - u^T\lambda^* - w^T v^* = p^*(0,0) - \mathrm{u}^{\mathrm{T}}\lambda^* - w^T v^*$$

$$\lambda_i^* = -\frac{\partial p^*(u,w)}{\partial u_i}\Big|\,(u,w) = (0,0),\ \ v_i^* = -\frac{\partial p^*(u,w)}{\partial w_i}\Big|\,(u,w) = (0,0)$$

Proof : $\dfrac{\partial p^*(0,0)}{\partial u_i} = \lim\limits_{t \searrow 0} \dfrac{p^*(te_i,0) - p^*(0,0)}{t} \ge -\lambda_i^*$

$$\frac{\partial p^*(0,0)}{\partial u_i} = \lim_{t \nearrow 0} \frac{p^*(te_i,0) - p^*(0,0)}{t} \le -\lambda_i^*$$

hence, equality

Generalized Inequalities

Primal

$$min\, f_o(x)$$
$$f_i \preccurlyeq_{K_i} 0, i = 1, \dots, m$$
$$h_i = 0, i = 1, \dots, p$$

Lagrangian

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

$$\lambda_i \succcurlyeq_{K_i^*} 0, i = 1, \dots, m,\ \lambda_i \in R^{k_i}$$

Lagrange Dual

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

Generalized Inequality: KKT Conditions

$$min\, f_o(x)$$
$$s.t\ f_i \preccurlyeq_{K_i} 0, i = 1, \dots, m$$
$$h_i = 0, i = 1, \dots, p$$

$f_i(x), i = 1, \dots, m,\ h_i(x),\ i = 1, \dots, p$ are differentiable

1. Primal constraints : $f_i(x) \preccurlyeq_{K_i} 0, i = 1, \dots, m$.
   $h_i(x) = 0, i = 1, \dots, p$.
2. Dual constraints : $\lambda \succcurlyeq_{K_i^*} 0$

3. Complementary slackness : $\lambda_i^T f_i(x) = 0, i = 1, \dots, m$.

4. Gradient of Lagrangian with respect to *x* variables
   $$\nabla_x f_o(x) + \sum_{i=1}^{m} \lambda_i^T \nabla_x f_i(x) + \sum_{i=1}^{p} v_i \nabla_x h_i(x) = 0$$

Generalized Inequalities: SOCP

Primal

$$min\ f^T x$$

$$||A_i x + b_i||_2 \leq c_i^T x + d, i = 1, \ldots, m$$

i.e. $(A_i x + b_i, c_i^T + d_i) \in K_i,\ i = 1, \ldots, m$

Lagrangian

$$L(x, \lambda, v) = f^T x - \sum (z_i^T (A_i x + b_i) + w_i (c_i^T + d_i))$$

$$(z_i, w_i) \in K_i^*,\ i = 1, \ldots, m$$

Lagrange Dual

$$\max - \sum_i (b_i^T z_i + d_i w_i)$$

$$||z_i|| \leq w_i, \qquad i = 1, \ldots, k$$

$$\sum_i A_i^T z_i + c_i w_i = f$$

Generalized Inequalities: Semidefinite Program

Primal

$$min\, c^T x$$

$$x_1F_1 + \cdots + x_nF_n + G \preccurlyeq 0, \text{where } F_1, \ldots, F_n, G \in S^k$$

Lagrangian

$$L(x, \lambda, v) = \inf_x (c^T x + tr\big((x_1F_1 + \cdots + x_nF_n + G\big)Z)),$$

$$Z \in S_+^k$$

Lagrange Dual

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

Dual

$$\max_Z tr(GZ)$$

$$tr(F_iZ) + c_i = 0, i = 1, \ldots, n$$

$$Z \succcurlyeq 0$$

Chapter 5 Summary

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

Chapter 5 Summary

- Duality provides a lower bound of the problem even the primal may not be convex.
- KKT conditions convert the minimization problem into equations.
- Lagrange multipliers provide the sensitivity of the constraints.
- Generalized inequality broadens the scope of convex optimization.