

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)
 - 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**)
- Sensitivity (Shadow-Price) (**Perturbation**)
- Generalized Inequalities

Duality

Primal Problem (Solution x is feasible)

$$\min f_0(x) \quad x \in R^n$$

$$\text{s. t. } f_i(x) \leq 0 \quad i = 1, \dots, m \quad \text{domain } D$$

$$h_i(x) = 0 \quad i = 1, \dots, p \quad = \text{dom } f_0 \cap_i \text{dom } f_i \cap_i \text{dom } h_i$$

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

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

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

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

Lagrange dual function (Solution x may not be feasible)

$$g(\lambda, v) = \inf_{x \in D} L(x, \lambda, v)$$

Dual Problem

$$\max_{\lambda, v} g(\lambda, v) \quad \text{s. t. } \lambda \in R_+^m, v \in R^p$$

Duality

Dual Problem (Solution x may not be feasible)

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

1. $g(\lambda, v)$ is concave

2. $g(\lambda, v) \leq p^*$ an optimal value where $\lambda \geq 0$

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 $\lambda \geq 0$

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

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

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

Example: Linear Programming

Prime:

$$\min c^T x$$

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

Lagrangian

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

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

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

Dual:

$$\begin{aligned} \max & -b^T \lambda_I \quad (\min b^T \lambda_I) \\ \text{subject to} & A^T \lambda_I + c \geq 0 \\ & \lambda_I \geq 0 \end{aligned}$$

Example: Linear Programming

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

$$\text{subject to } \begin{bmatrix} 1 & 3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \preceq \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

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

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

$$\lambda_1, \lambda_2 \geq 0$$

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

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

Example: Linear Programming

$$\min c^T x$$

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

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

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

Properties:

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

$$\max -b^T v$$
$$A^T v + c \geq 0$$

or

$$\max b^T v$$
$$A^T v \leq c$$

Example: Quadratic Programming

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

Lagrangian:

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

To minimize L over x , we set $\nabla_x L(x, v) = 2x + A^T v = 0$

$$\Rightarrow x = -\frac{1}{2} A^T v \quad (1)$$

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 $g(v) = -\frac{1}{4} v^T A A^T v - b^T v = b^T (AA^T)^{-1} b \quad (2)$

(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

$$\begin{aligned} \min x^T P x & \quad P \in S_{++}^n \\ \text{s. t. } Ax & \leq b \end{aligned}$$

Lagrange Dual Function:

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

Dual Problem:

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

Example: Quadratic Program (nonconvex prob.)

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

Dual Function:

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

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

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

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

Dual Problem:

$$\begin{array}{l|l} \max -b^T (A + \lambda I)^+ b - \lambda & \max -t - \lambda \\ \text{s.t. } A + \lambda I \geq 0 & \text{s.t. } \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \geq 0 \\ b \in R(A + \lambda I) & \end{array}$$

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

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

Example: Discrete Problem

Two-Way Partitioning Problem

Primal:

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

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:

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

Lagrange dual function:

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

Example: Discrete Problem

Dual:

$$\begin{aligned} \max g(v) &= -1^T v \\ \text{s.t. } W + \text{diag}(v) &\succeq 0 \end{aligned}$$

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

Examples: Conjugate Function

$$\begin{aligned} \min f_0(x) \\ \text{s.t. } Ax \leq b \\ Cx = d \end{aligned}$$

Dual function

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

Conjugate function

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

$$\begin{aligned} \min c^T x \\ \text{s.t. } Ax = b \\ x \geq 0 \end{aligned}$$

$$\begin{aligned} \max -b^T v \\ \text{s.t. } A^T v + c \geq 0 \end{aligned}$$

$$\begin{aligned} \min c^T x \\ \text{s.t. } Ax \leq b \end{aligned}$$

$$\begin{aligned} \max -b^T \lambda \\ \text{s.t. } A^T \lambda + c = 0 \\ \lambda \geq 0 \end{aligned}$$

Examples: Entropy Maximization

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

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

Thus, the dual function is

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

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

Dual Problem

$$\begin{aligned} \max \quad & -b^T \lambda - \log(\sum_{i=1}^n e^{-a_i^T \lambda}) \\ \text{s. t. } \quad & \lambda \geq 0 \end{aligned}$$

Examples: Minimum Volume Covering Ellipsoid

$$\begin{aligned} \min f_0(x) &= \log \det X^{-1}, \quad X \in S_{++}^n \\ \text{s. t. } &a_i^T X a_i \leq 1, \quad i = 1, \dots, m \end{aligned}$$

Lagrangian

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

Lagrange dual function

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

Dual Problem

$$\begin{aligned} \max & \log \det \left(\sum_{i=1}^m \lambda_i a_i a_i^T \right) - \mathbf{1}^T \lambda + n \\ \text{s. t. } & \sum_{i=1}^m \lambda_i a_i a_i^T \succ 0, \quad \lambda \in R_+^m \end{aligned}$$

Interpretation: Saddle-point

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

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

Example : $f(w, z)$

| | | | | | | |
|----------|----|---|---|----|----|---|
| $w = 1$ | 1 | [| 1 | -1 | 3 |] |
| $w = 2$ | 2 | [| 2 | 2 | -1 |] |
| $w = 3$ | 3 | [| 3 | 1 | -2 |] |
| $z = 1,$ | 2, | 3 | | | | |

I. z decides first

$$\left\{ \begin{array}{l} \min_{w \in W} f(w, 1) = 1 \\ \min_{w \in W} f(w, 2) = -1 \\ \min_{w \in W} f(w, 3) = -2 \end{array} \right. \quad \max_{z \in Z} \min_{w \in W} f(w, z) = 1$$

II. w decides first

$$\left\{ \begin{array}{l} \max_{z \in Z} f(1, z) = 3 \\ \max_{z \in Z} f(2, z) = 2 \\ \max_{z \in Z} f(3, z) = 3 \end{array} \right. \quad \min_{w \in W} \max_{z \in Z} f(w, z) = 2$$

Interpretation: Saddle-point

Claim : Result of II > Result of I

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

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

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

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

Interpretation: Saddle-point

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

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

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

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

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

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

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

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

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

Interpretation: Saddle-point

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

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

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

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

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

$$\max_{z \in Z} f(1, z) = 3$$

$$\max_{z \in Z} f(2, z) = 3$$

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

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

Interpretation: Saddle-point

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

Geometric Interpretation

$$\begin{aligned} \min f_o(x) & \quad (t) \\ \text{s.t. } f_1(x) & \leq 0 \quad (u \leq 0) \end{aligned}$$

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

$$g(\lambda) = \lambda u + t$$

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

u

Duality via Separating Hyperplane

Set $G = \{(f_1(x), \dots, f_m(x), h_1(x), \dots, h_p(x), f_o(x)) \mid x \in D\}$,
 $G \in R^m \times R^p \times R, p^* = \inf\{t \mid (u, w, t) \in g, u \leq 0, w = 0\}$

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

Separating hyperplane: Example

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

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

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

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

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

Lagrange dual problem

$$\begin{aligned} & \max g(\lambda, v) \\ & \text{s. t. } \lambda \geq 0 \end{aligned}$$

Properties

This is a convex problem.

The opt. solution is denoted as d^*

$$p^* - d^* = \text{gap} \geq 0$$

If $\text{gap} > 0$, it is a weak duality.

If $\text{gap} = 0$, it is a strong duality.

Slater's condition

relint: relative interior of set D

Given that the primal problem is convex,

$$\text{If } f_i(x) < 0, i = 1, \dots, m, \exists x \in \text{relint } D$$

Then strong duality holds.

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

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

↑ any norm

Shadow Price Interpretation: Food vs. Vitamin

Primal $\min c^T x$
 s. t. $Ax \geq b$
 $x \geq 0$

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

$$\min c_1 x_1 + c_2 x_2 + c_3 x_3$$

$$\min c^T x$$

$$\text{s. t. } -Ax + b \leq 0$$

$$-x \leq 0$$

$$\begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \geq \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, x_i \geq 0, \forall i$$

Dual $\max \lambda^T b$
 s. t. $A^T \lambda \leq c$
 $\lambda \geq 0$

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

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

Lagrangian $L(x, \lambda) = c^T x + \lambda_1^T (-Ax + b) + \lambda_2^T (-x)$
 $= [c^T + \lambda_1^T (-A) - \lambda_2^T] x + \lambda_1^T b$
 $c^T = \lambda_1^T (A) + \lambda_2^T, \text{ or } A^T \lambda_1 \leq c$

Shadow Price Interpretation: Spring Energy & Force

$$\min f_o(x_1, x_2) = \frac{1}{2}k_1x_1^2 + \frac{1}{2}k_2(x_2 - x_1)^2 + \frac{1}{2}k_3(l - x_2)^2$$

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

$$w/2 - x_1 \leq 0$$

$$w + x_1 - x_2 \leq 0$$

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

$$\min \frac{1}{2}(k_1x_1^2 + k_2(x_2 - x_1)^2 + k_3(l - x_2)^2)$$

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

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

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

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

zero gradient condition

$$\begin{bmatrix} k_1x_1 - k_2(x_2 - x_1) \\ k_2(x_2 - x_1) - k_3(l - x_2) \end{bmatrix} + \lambda_1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$

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 : $\lambda \geq 0$

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

4. Gradient of Lagrangian with respect to x variables

$$\nabla_x f_0(x) + \sum_{i=1}^m \lambda_i \nabla_x f_i(x) + \sum_{i=1}^p v_i \nabla_x h_i(x) = 0$$

KKT (Karush-Kuhn-Tucker) Conditions

1. Primal, Lagrangian, and Dual

$$\begin{aligned}
 \min f_o(x) \quad & L(x, \lambda, v) \\
 f_i \leq 0 \quad & \\
 h_i = 0 \quad & \\
 & = f_o(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p v_i h_i(x), \lambda_i \geq 0
 \end{aligned}$$

$$\begin{aligned}
 \max_{\lambda, v} \min_x &: (x, \lambda, v) \\
 &= \max_{\lambda, v} g(\lambda, v)
 \end{aligned}$$

$$\min_x \max_{\lambda, v} \quad \max_{\lambda, v} \min_x \quad (\text{dual})$$

1. Feasibility (x, λ, v)

$$2. L(x, \lambda, v) = f_o(x) \begin{cases} \lambda_i > 0 & \text{if } f_i = 0 \\ \lambda_i = 0 & \text{if } f_i < 0 \end{cases}$$

$$3. g(\lambda, v) = \min_x L(x, \lambda, v) = f_o(x)$$

Necessary condition for local optimality

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

Sensitivity

Perturbed Problem

$$\min f_o(x)$$

$$\text{s. t. } f_i \leq u_i$$

$$h_i(x) = w_i$$

$$\max \tilde{g} = g(\lambda, v) - u^T \lambda - w^T v$$

$$\text{s. t. } \lambda \geq 0$$

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

Unperturbed Problem

$$u_i = w_i = 0$$

$$\max g(\lambda, v)$$

$$\text{s. t. } \lambda \geq 0$$

$$p^*(0,0), \lambda^*, v^*$$

$$p^*(u, w) \geq g(\lambda^*, v^*) - u^T \lambda^* - w^T v^* = p^*(0,0) - u^T \lambda^* - w^T v^*$$

$$\lambda_i^* = - \left. \frac{\partial p^*(u, w)}{\partial u_i} \right|_{(u, w) = (0,0)}, \quad v_i^* = - \left. \frac{\partial p^*(u, w)}{\partial w_i} \right|_{(u, w) = (0,0)}$$

$$\text{Proof: } \frac{\partial p^*(0,0)}{\partial u_i} = \lim_{t \searrow 0} \frac{p^*(te_i, 0) - p^*(0,0)}{t} \geq -\lambda_i^*$$

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

hence, equality

Generalized Inequalities

Primal

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

Lagrangian

$$\begin{aligned} L(x, \lambda, v) = f_o(x) + \sum_{i=1}^m \lambda_i^T f_i(x) + \sum_{i=1}^p v_i h_i(x), \\ \lambda_i \succeq_{K_i^*} 0, i = 1, \dots, m, \lambda_i \in R^{k_i} \end{aligned}$$

Lagrange Dual

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

Generalized Inequality: KKT Conditions

$$\min f_o(x)$$

$$s.t \ f_i \preceq_{K_i} 0, i = 1, \dots, m$$

$$h_i = 0, i = 1, \dots, p$$

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

1. Primal constraints : $f_i(x) \preceq_{K_i} 0, i = 1, \dots, m.$

$$h_i(x) = 0, i = 1, \dots, p.$$

2. Dual constraints : $\lambda \succeq_{K_i^*} 0$

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

$$\min f^T x$$

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

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

Lagrangian

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

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

Lagrange Dual

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

$$\|z_i\| \leq w_i, \quad i = 1, \dots, k$$

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

Generalized Inequalities: Semidefinite Program

Primal

$$\min c^T x$$

$$x_1 F_1 + \cdots + x_n F_n + G \preceq 0, \text{ where } F_1, \dots, F_n, G \in S^k$$

Lagrangian

$$L(x, \lambda, v) = \inf_x (c^T x + \text{tr}((x_1 F_1 + \cdots + x_n F_n + G)Z)),$$

$$Z \in S_+^k$$

Lagrange Dual

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

Dual

$$\max_Z \text{tr}(GZ)$$

$$\text{tr}(F_i Z) + c_i = 0, i = 1, \dots, n$$

$$Z \succeq 0$$

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.