Convex Optimization

Lecture 9: Lagrangian Dual and KKT Condition

Lecturer: *Dr*. Wan-Lei Zhao

Autumn Semester 2025

Contact: wlzhao@xmu.edu.cn

Outline

Lagrange Dual (1)

- Given a typical optimization problem

$$\begin{aligned} \underset{x}{\text{Min.}} \quad & f(x) \quad x \in R^n \\ s.t. \; g_i(x) &\leq 0, \quad i = 1 \cdots m \qquad (1) \\ h_j(x) &= 0, \quad j = 1 \cdots p \end{aligned}$$

- Its optimal value is p^*
- The Lagrange of the problem in $L : R^n \times R^m \times R^p$ is

$$L(x) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

Figure: John Von Neuman (1903 -1957)

Lagrange Dual (2)

- Given a typical optimization problem $\mathcal{P}$

$$\begin{aligned} \underset{x}{\text{Min.}} \quad & f(x) \quad x \in R^n \\ s.t.\ g_i(x) &\leq 0, \quad i = 1 \cdots m \\ h_j(x) &= 0, \quad j = 1 \cdots p \end{aligned} \tag{2}$$

- Its optimal value is P^*
- The Lagrange of the problem in $L: R^n \times R^m \times R^p$ is

$$L(x) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

- Whether the Lagrange is equivalent to the original $\mathcal{P}$

$$\inf_x L(x) = \begin{cases} +\infty & x \notin D \\ \inf_x f(x) & x \in D \end{cases}$$

Lagrange Dual (3)

- Its optimal value is P^*
- The Lagrange of the problem in $L : R^n \times R^m \times R^p$ is

$$L(\lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

- The Lagrange dual is defined as $G : R^m \times R^p$

$$\begin{aligned} G(\lambda, \mu) &= \inf_{x \in D} (x, \lambda, \mu) \\ &= \inf_{x \in D} f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \end{aligned}$$

- Observe that
  1. the infimum is unconstrained
  2. G is concave regardless of original problem
  3. G can be $-\infty$

Poof: First of all, $dom\ G = R_+^m \times R^p$ is convex, $\forall(\lambda_1, \mu_1) \in dom\ G$, $\forall(\lambda_2, \mu_2) \in dom\ G$, $\theta \in [0,\ 1]$

$$
\begin{aligned}
G(\theta\lambda_1 + (1-\theta)\lambda_2, \theta\mu_1 + (1-\theta)\mu_2) = \inf_{x \in D}\{f(x) + \sum_{i=1}^{m}[\theta\lambda_1 + (1-\theta)\lambda_2]_i \\
g_i(x) + \sum_{j=1}^{p}[\theta\mu_1 + (1-\theta)\mu_2]_j h_j(x)\} \\
= \inf_{x \in D}\{f(x) + \sum_{i=1}^{m}[\theta\lambda_1]_i g_i(x) + \sum_{j=1}^{p}[\theta\mu_1]_j h_j(x) \\
\sum_{i=1}^{m}[(1-\theta)\lambda_2]_i g_i(x) + \sum_{j=1}^{p}[(1-\theta)\mu_2]_j h_j(x)\} \\
\geq \theta G(\lambda_1, \mu_1) + (1-\theta) G(\lambda_2, \mu_2)
\end{aligned}
$$

—prooved—

Lagrange Dual (4): the lower bound property

- Its optimal value is P^*
- The Lagrange of the problem in $L : R^n \times R^m \times R^p$ is

$$L(\lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

- **The lower bound property:** if $\lambda \geq 0$, then $G(x, \lambda, \mu) \leq p^*$
- **Proof:**
- Given $\widehat{x}$ is feasible and $\lambda \geq 0$ then
- Based on the fact that $\sum_{i=1}^{m} \lambda_i g_i(\widehat{x}) \leq 0$ and $\sum_{j=1}^{p} \mu_j h_j(\widehat{x}) = 0$

$$f(\widehat{x}) \geq L(\widehat{x}, \lambda, \mu) \geq \inf_{x \in D}(x, \lambda, \mu) = G(\lambda, \mu)$$

Lagrange Dual (5): the dual problem

- Its optimal value is P^*
- The Lagrange of the problem in $L : R^n \times R_+^m \times R^p$ is

$$\begin{aligned} &\underset{\lambda,\mu}{\text{Max.}} \quad G(\lambda, \mu) \\ &s.t. \quad \lambda \geq 0. \end{aligned}$$

- This problem finds the best lower bound on for the original problem $\mathcal{P}$
- It is a convex optimization since λ_i and μ_j are linear in the objective function
- Given the optimal value for G is d^*
- λ, μ are dual feasible if $\lambda \geq 0$ and $(\lambda, \mu) \in domG$

Lagrange Dual (6): example-1

$$\text{Min.} x^T x$$
$$s.t. \quad Ax = b$$

- The Lagrange is $L(x, \lambda) = x^T x + \lambda^T (Ax - b)$
- To minimize L over x, set gradient equal to zero:

$$\bigtriangledown_x L(x, \mu) = 2x + A^T \mu = 0$$
$$\Rightarrow \quad x = -\frac{1}{2} A^T \mu$$

- Plug $x = -\frac{1}{2} A^T \mu$ into L

$$\bigtriangledown_x L(-\frac{1}{2} A^T \mu, \mu) = -\frac{1}{4} \mu^T A A^T \mu - b^T \mu$$

Lagrange Dual (7): example-2

$$\underset{x}{\text{Min.}} \quad c^T x$$
$$s.t. \quad Ax = b \quad x \succeq 0$$

- The Lagrange is:

$$\begin{aligned} L(x, \lambda, \mu) &= c^T x + \mu(Ax - b) - \lambda^T x \\ &= -b^T \mu + (c + A^T \mu - \lambda)^T x \end{aligned}$$

$$\Downarrow$$

$$G(\lambda, \mu) = \inf_x L(x, \lambda, \mu) = \begin{cases} -b^T \mu & A^T \mu - \lambda + c = 0 \\ -\infty & otherwise \end{cases}$$

- G is linear on the affine domain $\{(\lambda, \mu) | A^T \mu - \lambda + c = 0\}$, hence it is concave
- We have $p^* \geq -b^T \mu$ if $A^T + v \succeq 0$

LP with Box Constraints (1)

$$\begin{aligned} \underset{x}{\text{Min.}} \quad & c^T x \\ s.t. \quad & Ax = b \\ & \ell \preceq x \preceq u \end{aligned}$$

$$\Downarrow$$

$$\begin{aligned} L(x, \lambda_1, \lambda_2, v) &= c^T x + v^T (Ax - b) + \lambda_1^T (l - x) + \lambda_2^T (x - u) \\ &= (c + A^T v - \lambda_1 + \lambda_2)^T x - v^T b + \lambda_1^T l - \lambda_2^T u \end{aligned}$$

$$\Downarrow$$

$$g(\mu) = \begin{cases} -v^T b + \lambda_1^T l - \lambda_2^T u & c + A^T v + \lambda_1 - \lambda_2 = 0 \\ -\infty & c + A^T v + \lambda_1 - \lambda_2 \neq 0 \end{cases}$$

LP with Box Constraints (2)

$$
\begin{aligned}
L(x, \lambda_1, \lambda_2, v) &= c^T x + v^T(Ax - b) + \lambda_1^T(l - x) + \lambda_2^T(x - u) \\
&= (c + A^T v - \lambda_1 + \lambda_2)^T x - v^T b + \lambda_1^T l - \lambda_2^T u
\end{aligned}
$$

$$\Downarrow$$

$$
g(\mu) = \begin{cases} -v^T b + \lambda_1^T l - \lambda_2^T u & c + A^T v + \lambda_1 - \lambda_2 = 0 \\ -\infty & c + A^T v + \lambda_1 - \lambda_2 \neq 0 \end{cases}
$$

$$\Downarrow$$

$$
\begin{aligned}
max \quad & -b^T v - \lambda_1^T u + \lambda_2^T l \\
s.t. \quad & A^T v + \lambda_1 - \lambda_2 + c = 0 \\
& \lambda_1 \geq 0 \quad \lambda_2 \geq 0
\end{aligned}
$$

Least Square Problem

- Minimize $\| Ax - b \|_2$

$$
\begin{aligned}
\text{Min.}_x \quad & \frac{1}{2} \| y \|_2 \\
s.t. \quad & Ax - b = y
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
L(x, y, v) &= \frac{1}{2} \| y \|_2^2 + v^T(Ax - b - y) \\
&= \frac{1}{2} \| y \|_2^2 + v^T Ax - v^T b - v^T y
\end{aligned}
$$

$$\Downarrow$$

$$
g(v) = \begin{cases} -v^T b - \frac{1}{2} \| v \|_2^2 & v^T A = 0 \\ -\infty & v^T A \neq 0 \end{cases}
$$

Lagrange Dual (6): the weak duality

- Given p^* is the optimal value for the original problem $\mathcal{P}$

$$L(\lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

- Given d^* is the optimal value for the dual problem $\mathcal{D}$

$$\begin{aligned} G(\lambda, \mu) &= \inf_{x \in D} (x, \lambda, \mu) \\ &= \inf_{x \in D} f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \end{aligned}$$

- We have $p^* \geq d^*$
- Moreover, $p^* - d^*$ is called the dual gap

Lagrange Dual (7): the strong duality

- Given p^* is the optimal value for the original problem $\mathcal{P}$

$$L(\lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

- Given d^* is the optimal value for the dual problem $\mathcal{D}$

$$\begin{aligned} G(\lambda, \mu) &= \inf_{x \in D} (x, \lambda, \mu) \\ &= \inf_{x \in D} f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \end{aligned}$$

- We have $p^* > d^*$
- We call the dual is **strong duality**
- We call the dual is **weak duality**

Lagrange Dual (8): the Slater condition

- Given a typical optimization problem $\mathcal{P}$

$$\begin{aligned} \underset{x}{\text{Min.}} \quad & f(x) \quad x \in R^n \\ s.t.\ g_i(x) &\leq 0, \quad i = 1 \cdots m \\ h_j(x) &= 0, \quad j = 1 \cdots p \end{aligned}$$

- $\forall i$, $\exists x \in Relint(D)$ such that $g_i(x) < 0$ and all $h_j(x) = 0$, we have $p^* = d^*$
- $Relint(D)$: relative interior set
- This is the sufficient condition for $p^* = d^*$
- It is called **Slater** condition

Outline

The Opening Discussion

$$\begin{aligned} \underset{x}{\text{Min.}} \quad & f(x) \quad x \in R^n \\ s.t.\ g_i(x) &\leq 0, \quad i = 1\cdots m \\ h_j(x) &= 0, \quad j = 1\cdots p \end{aligned}$$

- So far we discussed about one important way to solve an optimization problem
- Given any optimization problem
  1. There could be no solution
  2. There could be optimal solution and solvable in finite steps
  3. There could be optimal solution but cannot be easily found
- We are going to give the condition for the 2nd case

Karush-Kuhn-Tucker Condition (1)

$$\underset{x}{\text{Min.}} \quad f(x) \quad x \in R^n$$
$$s.t.\ g_i(x) \leq 0, \quad i = 1 \cdots m$$
$$h_j(x) = 0, \quad j = 1 \cdots p$$

- Given $f(x)$, and all $g_i(x)$s and $h_j(x)$s are differentiable
- The Lagrange function is

$$L(\lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

- x^*, λ^*, and μ^* is the optimal solution for above problem
- They should satisfy five conditions

Karush-Kuhn-Tucker Condition (1-3): Primal feasibility

$$L(\lambda,\mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

- x^*, λ^*, and μ^* is the optimal solution for above problem
- Three conditions on the original problem

$$\begin{cases} g_i(x^*) \leq 0 & i = 1 \cdots m \\ h_j(x^*) = 0 & j = 1 \cdots p \\ \lambda \succeq 0 \end{cases} \tag{3}$$

- Sometimes $\lambda \succeq 0$ is called dual feasibility

Karush-Kuhn-Tucker Condition (4): Complementary Slackness

$$L(\lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

- x^*, λ^*, and μ^* is the optimal solution for above problem
- Complementary Slackness condition

$$\lambda^* g_i(x^*) = 0 \tag{4}$$

Karush-Kuhn-Tucker Condition (5): Stationarity

$$L(\lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

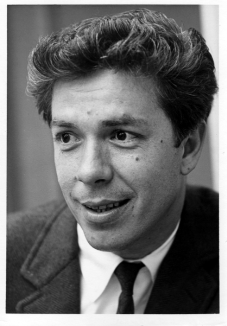
- x^*, λ^*, and μ^* is the optimal solution for above problem
- The solution must be stable, that is

$$\frac{\partial L(x, \lambda^*, \mu^*)}{\partial x}|_{x=x^*} = \bigtriangledown f(x^*) + \sum_{i=1}^{m} \lambda_i^* \bigtriangledown_x g_i(x^*) + \sum_{j=1}^{p} \mu_i^* \bigtriangledown_x h_i(x^*) = 0$$

Karush-Kuhn-Tucker Condition: the Mathematicians behind

(a) W. Karush

(b) H. Kuhn



(c) A. Tucker