Lagrange Duality

Xin Deng

ShanghaiTech University

dengxin1@shanghaitech.edu.cn

June 13, 2020

Outline

Outline:

1. Lagrangian
2. Lagrange Dual Function
3. Dual Problem
4. Weak and Strong Duality
5. KKT conditions

Lagrangian

- Consider an optimization problem in standard form (not necessarily convex)

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0 \quad i = 1, \cdots, m \\ & h_i(\boldsymbol{x}) = 0 \quad i = 1, \cdots, p \end{array}$$

  with variable $\boldsymbol{x} \in \mathbb{R}^n$, domain $\mathcal{D}$, and optimal value $p^\star$
- The *Lagrangian* is a function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, with $\operatorname{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$, defined as

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

  where λ_i is the Lagrange multiplier associated with $f_i(\boldsymbol{x}) \leq 0$ and ν_i is the Lagrange multiplier associated with $h_i(\boldsymbol{x}) = 0$.

Lagrange Dual Function

- The *Lagrange dual function* is defined as the infimum of the Lagrangian over $\boldsymbol{x}$: $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$,

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\ &= \inf_{\boldsymbol{x} \in \mathcal{D}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x}) \right) \end{aligned}$$

- Observe that:
  - the infimum is unconstrained (as opposed to the original constrained minimization problem)
  - g is concave regardless of original problem (infimum of affine functions)
  - g can be $-\infty$ for some $\boldsymbol{\lambda}, \boldsymbol{\nu}$

Lagrange Dual Function

- **Lower bound property:** if $\boldsymbol{\lambda} \succeq \mathbf{0}$, then $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star$.

Proof.

Suppose $\tilde{\boldsymbol{x}}$ is feasible and $\boldsymbol{\lambda} \succeq \mathbf{0}$. Then,

$$f_0(\tilde{\boldsymbol{x}}) \geq L(\tilde{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \geq \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$$

Now choose minimizer of $f_0(\tilde{\boldsymbol{x}})$ over all feasible $\tilde{\boldsymbol{x}}$ to get $p^\star \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. □

- We could try to find the best lower bound by maximizing $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. This is in fact the dual problem.

Dual Problem

- The *Lagrange dual problem* is defined as

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}, \boldsymbol{\nu}}{\text{maximize}} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0} \end{array}$$

- This problem finds the best lower bound on $p^\star$ obtained from the dual function
- It is a convex optimization (maximization of a concave function and linear constraints)
- The optimal value is denoted $d^\star$
- $\boldsymbol{\lambda}, \boldsymbol{\nu}$ are dual feasible if $\boldsymbol{\lambda} \succeq \mathbf{0}$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \operatorname{dom} g$ (the latter implicit constraints can be made explicit in problem formulation)

Dual Problem

Example:

- With a given vector y,

$$\begin{array}{ll} \text{minimize}_x & \sum_{k=1}^{n} x_k \log(x_k/y_k) \\ \text{subject to} & Ax = b, \\ & \mathbf{1}^T x = 1, \end{array}$$

  The domain of the objective function is $\mathbb{R}^n_{++}$. The parameters $y \in \mathbb{R}^n_{++}$, $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$ are given.

- Then the Lagrange dual problem of this problem is

$$\text{maximize}_z \quad b^T z - \log \sum_{k=1}^{n} y_k e^{a_k^T z}$$

  (a_k is the kth column of A).

Dual Problem

- The Lagrangian of primal problem is

$$L(x, \lambda, \mu) = \sum_{k=1}^{n} x_k \log(x_k/y_k) + b^T\lambda - \lambda^T Ax + \mu - \mu\mathbf{1}^T x.$$

- Minimizing over x_k gives the conditions

$$1 + \log(x_k/y_k) - a_k^T\lambda - \mu = 0, \quad k = 1, \dots, n,$$

with solution

$$x_k = y_k e^{a_k^T\lambda+\mu-1}.$$

- Plugging this in L gives the Lagrange dual function

$$g(\lambda, \mu) = b^T\lambda + \mu - \sum_{k=1}^{n} y_k e^{a_k^T\lambda+\mu-1}$$

Dual Problem

The corresponding dual problem is

$$\text{maximize}_\lambda \quad b^T\lambda + \mu - \sum_{k=1}^{n} y_k e^{a_k^T\lambda+\mu-1}.$$

This can be simplified a bit if we optimize over μ by setting the derivative equal to zero:

$$\mu = 1 - \log \sum_{k=1}^{n} y_k e^{a_k^T\lambda}.$$

After this simplification the dual problem reduces to the previous problem presented.

Weak and Strong Duality

- From the lower bound property, we know that $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star$ for feasible $(\boldsymbol{\lambda}, \boldsymbol{\nu})$. In particular, for a $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ that solves the dual problem.
- Hence, **weak duality** always holds (even for nonconvex problems):

$$d^\star \leq p^\star$$

- The difference $p^\star - d^\star$ is called **duality gap**.
- Solving the dual problem may be used to find nontrivial lower bounds for difficult problems.
- Even more interesting is when equality is achieved in weak duality. This is called **strong duality**:

$$d^\star = p^\star$$

Weak and Strong Duality

- Strong duality means that the duality gap is zero.
- Strong duality:
  - is very desirable (we can solve a difficult problem by solving the dual)
  - does not hold in general
  - usually holds for convex problems
  - conditions that guarantee strong duality in convex problems are called **constraint qualifications**.

Weak and Strong Duality

- Slater's constraint qualification is a very simple condition that is satisfied in most cases and ensures strong duality for convex problems.

- Strong duality hold for a convex problem

$$
\begin{array}{lll}
\underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) & \\
\text{subject to} & f_i(\boldsymbol{x}) \leq 0 & i = 1, \cdots, m \\
 & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} &
\end{array}
$$

  if it is strictly feasible, i.e.,

$$\exists\, \boldsymbol{x} \in \operatorname{int} \mathcal{D} : \quad f_i(\boldsymbol{x}) < 0 \quad i = 1, \cdots, m, \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

- There exist many other types of constraint qualifications.

Weak and Strong Duality

If strong duality holds, then we can solve a problem indirectly via solving its corresponding dual problem.

Example:

Primal:

$$\begin{aligned} \text{minimize}_x \quad & c^T x \\ \text{subject to} \quad & Ax \succeq b, \\ & x \succeq 0. \end{aligned}$$

Dual:

$$\begin{aligned} \text{maximize}_\lambda \quad & \lambda^T b \\ \text{subject to} \quad & \lambda^T A \preceq c^T, \\ & \lambda \succeq 0. \end{aligned}$$

KKT conditions

KKT conditions (for differentiable f_i, h_i):

1. primal feasibility:

$$f_i(\boldsymbol{x}) \leq 0,\ i = 1, \cdots, m,\ h_i(\boldsymbol{x}) = 0,\ i = 1, \cdots, p$$

2. dual feasibility: $\boldsymbol{\lambda} \succeq \mathbf{0}$

3. complementary slackness: $\lambda_i^\star f_i(\boldsymbol{x}^\star) = 0$ for $i = 1, \cdots, m$

4. zero gradient of Lagrangian with respect to $\boldsymbol{x}$:

$$\nabla f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i \nabla h_i(\boldsymbol{x}) = \mathbf{0}$$

KKT conditions

- We already known that if strong duality holds and $\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}$ are optimal, then they must satisfy the KKT conditions.
- What about the opposite statement?
- If $\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}$ satisfy the KKT conditions for a convex problem, then they are optimal.

Proof.

From complementary slackness, $f_0(\boldsymbol{x}) = L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ and, from 4th KKT condition and convexity, $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$. Hence, $f_0(\boldsymbol{x}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. □

Theorem

If a problem is convex and Slater's condition is satisfied, then x is optimal if and only if there exists $\boldsymbol{\lambda}, \boldsymbol{\nu}$ that satisfy the KKT conditions.

Thanks.