DSC 211 Introduction to Optimization Winter 2024 — Instructor: Jun-Kun Wang
Scribe: Zhishang Luo — February 6, 2024
Editors/TAs: Merlin Chang, Marialena Sfyraki

Lecture 9: Lagrangian, Dual Problem, and Duality

1 Lagrangian

Definition 1. *(**Lower/Upper bound**) An element u is an upper bound of a set S if $u \geq s$, for all $s \in S$. Similarly, an element l is a lower bound of a set S if $l \leq s$, for all $s \in S$.*

Definition 2. *(**Infimum**) Let S be a non-empty set of real numbers. The infimum of S, denoted as $m = \inf S$, where $m \in \mathbb{R}$, is defined as the **greatest lower bound** of S, such that:*

1. *$m \leq x$ for all $x \in S$.*
2. *If b is any lower bound of S, then $b \leq m$.*

The infimum of a function is denoted as

$$\inf_{x \in C} f(x),$$

where $f(\cdot) : C \to \mathbb{R}$ and $C \subseteq \mathbb{R}^d$.

Remark: Using the above definition, the optimality gap is defined as

$$\delta_k := f(x_k) - \inf_x f(x).$$

Remark: The motivation to use the concept $\inf_{x \in C} f(x)$ instead of $\min_{x \in C} f(x)$ is that some functions $f(x)$ do not have a minimum. Minimum of a function $\min_{x \in C} f(x)$ needs to be attained at a point in the set C, while the infimum of a function $\inf_{x \in C} f(x)$ does not necessarily need to be attained at a point in the set C.

Example: Some common loss functions $f(x)$ where the minimum does not exist but the infimum exists:

1. Exponential loss function: $f(x) = \exp(-x)$

   The minimum of the function $\min_{x \in C} f(x)$ does not exist.
   The infimum of the function is $\inf_{x \in C} f(x) = 0$.

2. Logistic loss function $f(x) = log(1 + \exp(-x))$

   The minimum of the function $\min_{x \in C} f(x)$ does not exist.
   The infimum of the function is $\inf_{x \in C} f(x) = 0$.

Remark: Note that in practice, we don't know the exact value of $\inf_{x \in C} f(x)$, therefore we need to have an estimate of a lower bound of the $\inf_{x \in C} f(x)$, denoted as y_*, where $y_* \leq \inf_{x \in C} f(x)$.

Question: Why do we want to estimate the lower bound of $\inf_{x \in C} f(x)$?

Answer: Optimality Gap $\delta_k := f(x_k) - \inf_x f(x) \leq f(x_k) - y_*$, where y_* is a lower bound of $\inf_{x \in C} f(x)$ and $f(x_k) - y_*$ is am upper bound of the optimality gap δ_k. We want to estimate an upper bound of the optimality gap δ_k, which is equivalent to estimating a lower bound of the infimum $\inf_{x \in C} f(x)$.

We now consider the following general constrained optimization with functional constraints:

$$\begin{aligned} &\inf_{x \in \mathbb{R}^d} f(x) \\ &\quad \text{s.t. } f_j(x) \leq 0, \quad j = 1, \dots, m. \\ &\quad \text{s.t. } \textbf{affine } \; h_i(x) = 0, \quad i = 1, \dots, p. \end{aligned}$$

Note that the problem can be rewritten as $\min_{x \in C} f(x)$, where the set C is defined by the functional constrains:

$$C := \{x \in \mathbb{R}^d : f_j(x) \leq 0, \forall j \in [m]; h_i(x) = 0, \forall i \in [p]\}$$

Definition 3. ***(Affine Function)*** *Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a function. We say that h is an affine function, if for all $\mathbf{x} \in \mathbb{R}^n$, the function can be written as:*

$$h_i(\mathbf{x}) = \mathbf{a_i^T x} + b$$

where $\mathbf{a_i} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, $i = 1, \dots, n$.

Remark: We can reformulate the problem of finding the minimizer of a function subject to constraints using the **Lagrangian**, $L(x, \lambda, \mu)$.

Definition 4. ***(Lagrangian)*** *The Lagrangian is defined as:*

$$L(x, \lambda, \mu) := f(x) + \sum_{j=1}^{m} \lambda_j f_j(x) + \sum_{i=1}^{p} \mu_i h_i(x).$$

where $\lambda_j \geq 0$, $f_j(x) \leq 0$, $\quad j = 1, \dots, m$ and ***affine*** *$h_i(x) = 0$, $\quad i = 1, \dots, p$.*

Property 1 of the Lagrangian: The Lagrangian lower-bounds the function $f(\cdot)$, that is if $x \in C$, then

$$L(x, \lambda, \mu) \leq f(x)$$

Proof: if $x \in C$ ($h_i(x) = 0, \quad i = 1, \ldots, p.$)

$$L(x, \lambda, \mu) = f(x) + \sum_{j=1}^{m} \lambda_j f(x) + 0$$

We have

$$\lambda_j \geq 0, f_j(x) \leq 0, \quad j = 1, \ldots, m$$

Therefore,

$$\lambda_j f(x) \leq 0 \quad j = 1, \ldots, m$$

Therefore,

$$L(x, \lambda, \mu) \leq f(x)$$

Remark: When is $L(x, \lambda, \mu) = f(x)$ when $x \in C$?

- Case 1: $\lambda = 0$, then $L(x, 0, \mu) = f(x)$
- Case 2: $\lambda \neq 0$, then $L(x, \lambda, \mu) \leq f(x)$

Property 2 of the Lagrangian: Let set C contain all the $x \in \mathbb{R}^d$ under the constraints

$$C := \{x \in \mathbb{R}^d : f_j(x) \leq 0, \forall j \in [m]; h_i(x) = 0, \forall i \in [p]\}$$

We have

$$\sup_{\lambda_j \geq 0; \mu} L(x, \lambda, \mu) = \begin{cases} f(x) & \text{, if } x \in C \\ \infty & \text{, otherwise} \end{cases}.$$

This property implies that

Implication 1.

$$\inf_{x \in \mathbb{R}^d} \sup_{\lambda_j \geq 0, \mu} L(x, \lambda, \mu) = \inf_{x \in C} \sup_{\lambda_j \geq 0, \mu} L(x, \lambda, \mu) = \inf_{x \in C} f(x)$$

Implication 2. For any dual variables λ, μ

$$\inf_{x \in \mathbb{R}^d} L(x, \lambda, \mu) \leq \inf_{x \in C} f(x).$$

2 Dual Problem

Dual function is obtained by minimizing $L(x, \lambda, \mu)$ over the primal variables x.

Definition 5. *(**Dual Function**) Let $L(x, \lambda, \mu)$ be the Lagrangian function of $f(x)$, then dual function $g(\lambda, \mu)$ is defined as*

$$g(\lambda, \mu) \coloneqq \inf_{x \in \mathbb{R}^d} L(x, \lambda, \mu)$$

Definition 6. *(**Dual Problem**) From the **Dual Function** $g(\lambda, \mu)$ we defined above, the dual problem is defined as*

$$\sup_{\lambda \geq 0; \mu} g(\lambda, \mu)$$

Theorem 1. *Weak duality of dual function*

$$\sup_{\lambda \geq 0; \mu} g(\lambda, \mu) \leq \inf_{x \in C} f(x)$$

Remark: This theorem tells us that the dual value is not greater than the primal value.

Proof. To prove theorem 1, from **Implication 2**, we have

$$\inf_{x \in \mathbb{R}^d} L(x, \lambda, \mu) \leq \inf_{x \in C} f(x),$$

which holds for any λ and μ; hence, we have

$$\sup_{\lambda \geq 0; \mu} g(\lambda, \mu) = \inf_{x \in \mathbb{R}^d} L(x, \lambda, \mu) \leq \inf_{x \in C} f(x).$$

□

Definition 7. *(**Strong Duality**) Strong duality means that*

$$\sup_{\lambda \geq 0; \mu} g(\lambda, \mu) = \inf_{x \in C} f(x).$$

Remark: On the left-hand side is maximizing the **dual problem** and on the right-hand side is minimizing the **primal problem**.

3 Example of a Dual Problem

Example: Obtaining the dual problem of the following primal:

$$\min_{x \in \mathbb{R}^d} \langle c, x \rangle$$
$$\text{s.t. } Ax \geq b$$

Step 1: Get the Lagrangian

$$\text{Set } b - Ax \leq 0$$
$$L(x, \lambda) = \langle c, x \rangle + \langle \lambda, b - Ax \rangle = c^T x + \lambda^T b - \lambda^T Ax$$

Step 2: Get the dual function

$$g(\lambda) = \inf_x L(x, \lambda) = \inf_x \left[\langle c - A^T \lambda, x \rangle + b^T \lambda \right] = \begin{cases} 0 + b^T \lambda & \text{, when } c - \lambda^T A = 0, \text{then } c = A^T \lambda \\ -\infty & \text{, otherwise} \end{cases}$$

Step 3: Dual Problem: maximizing $g(\lambda)$

$$\sup_\lambda b^T \lambda \ \ s.t. \ \lambda \geq 0, c = A^T \lambda$$

Remark: Dual function is a concave function (no matter the original function is convex or not). In other words, dual function is infimum of the affine functions with respect to dual variables, i.e„ $g(\lambda, \mu)$ can be written as $g(\theta) := \inf_x L(x, \lambda, \mu) = \inf_x [f(x) + \sum_{j=1}^m \lambda_j f_j(x) + \sum_{i=1}^p \mu_i h_i(x)] = \inf_x g_x(\theta)$, where θ is (λ, μ). Note that the fact that the dual function is concave in no way suggest that solving a non-convex problem is easy (as its dual problem is about maximizing a concave function). The catch/caveat is that getting the dual function is actually an optimization problem, which might be computationally challenging.

Proof: Show $g((1-\alpha)\theta_1 + \alpha\theta_2) \geq (1-\alpha)g(\theta_1) + \alpha g(\theta_2)$, where $g(\theta)$ is dual function, $\alpha \in [0, 1]$.

$$\begin{aligned} g((1-\alpha)\theta_1 + \alpha\theta_2) &= \inf_x g_x((1-\alpha)\theta_1 + \alpha\theta_2) \\ &= \inf_x [(1-\alpha)g_x(\theta_1) + \alpha g_x(\theta_2)], \text{ since } g_x \text{ is affine} \\ &\geq (1-\alpha) \inf_x g_x(\theta_1) + \alpha \inf_x g_x(\theta_2) \\ &= (1-\alpha)g_x(\theta_1) + \alpha g_x(\theta_2) \end{aligned}$$

Definition 8. *(**Slater condition**) Given Strong duality means that*

$$\sup_{\lambda \geq 0; \mu} g(\lambda, \mu) = \inf_{x \in C} f(x).$$

The condition of strong duality is that
There exists a point $\bar{x} \in C$ such that all the inequality constraints defining C are strict at $\bar{x}$, i.e., $f_j(\bar{x}) < 0, \forall j \in [m]$, and $h_i(\bar{x}) = 0, \forall i \in [p]$.

Theorem 2. *If $f, f_1, \ldots, f_m$ are convex functions and $h_i(\cdot)$ are affine, the Slater condition guarantees the strong duality.*

Bibliographic notes

For the proof of strong duality under Slater's condition, see Chapter 5.4 of Algorithms for Convex Optimization. Nisheeth K. Vishnoi. [1].

References

[1] Nisheeth K. Vishnoi. Algorithms for Convex Optimization Cambridge University Press, 2021