ECE 273 Convex Optimization and Applications Instructor: Jun-Kun Wang
Scribe: Adi Krishnamoorthy, Jerry Yan, Atefeh Mollabagher May 2, 2024
Editor/TA: Marialena Sfyraki

Lecture 10: Duality Theory Part I: Lagrangian, Dual function, Duality and Part II: KKT (Kahn-Karush-Tucker) Conditions

1 Lagrangians and dualities

We attempt to find the minimizer of a function subject to constraints

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

Definition 1. ***(Lagrangian):***

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

where $\lambda \ge 0$.

Property 1 of the Lagrangian:
Let $\Omega := \{x \in \mathbb{R}^d : f_j(x) \le 0, \forall j \in [m]; h_i(x) = 0, \forall i \in [p]\}$. Consider $x \in \Omega$. Then, the Lagrangian lower-bounds the function $f(x)$.

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

Property 2 of the Lagrangian:
Let $\Omega := \{x \in \mathbb{R}^d : f_j(x) \le 0, \forall j \in [m]; h_i(x) = 0, \forall i \in [p]\}$, and

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

Then,

$$\inf_{x\in\mathbb{R}^d} \sup_{\lambda\ge 0;\mu} L(x,\lambda,\mu) = \inf_{x\in\Omega} \sup_{\lambda\ge 0;\mu} L(x,\lambda,\mu) = \inf_{x\in\Omega} f(x)$$

Definition 2. ***(Dual function):***

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

Definition 3. ***(Dual problem):***

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

Theorem 1. ***(Weak duality):*** *Weak duality is defined by*

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

Proof. Note that by Property 2:

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

Then for any fixed $\lambda \geq 0, \mu$:

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

Therefore:

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

□

Remark: In general, for any function $Q(\cdot, \cdot)$, the following is true:

$$\sup_{y \in Y} \inf_{x \in X} Q(x, y) \leq \inf_{x \in X} \sup_{y \in Y} Q(x, y)$$

Lemma 1. *Dual function is concave:*

$$\begin{aligned} g(\lambda, \mu) &= \inf_{x \in \mathbb{R}^d} L(x, \lambda, \mu) \\ &= \inf_{x \in \mathbb{R}^d} f(x) + \sum_{j=1}^{m} \lambda_j f_j(x) + \sum_{i=1}^{p} \mu_i h_i(x) \end{aligned}$$

Remark: The dual function is the infimum of affine functions, i.e. $g(\lambda, \mu)$ can be written as $g(\theta) := \inf_{x \in \mathbb{R}^d} q_x(\theta)$ where $q_x(\theta) = f(x) + \sum_{j=1}^{m} \lambda_j f_j(x) + \sum_{i=1}^{p} \mu_i h_i(x)$.

Proof. (Lemma 1.) We prove this by showing that $g((1-\alpha)\theta_1 + \alpha\theta_2) \geq (1-\alpha)g(\theta_1) + \alpha g(\theta_2)$. We know that

$$\begin{aligned} & g((1-\alpha)\theta_1 + \alpha\theta_2) \\ &= \inf_{x \in \mathbb{R}^d} q_x((1-\alpha)\theta_1 + \alpha\theta_2) \\ &= \inf_{x \in \mathbb{R}^d} ((1-\alpha) q_x(\theta_1) + \alpha q_x(\theta_2)) \\ &\geq (1-\alpha) \inf_{x \in \mathbb{R}^d} q_x(\theta_1) + \alpha \inf_{x \in \mathbb{R}^d} q_x(\theta_2) \\ &= (1-\alpha) g(\theta_1) + \alpha g(\theta_2) \end{aligned}$$

where the third equality is true due to q_x being affine. □

Definition 4. ***(Strong duality):*** *Strong duality is defined as*

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

Definition 5. ***(Slater condition):*** *There exists a point $\bar{x}\in\Omega$ such that all the inequality constraints defining Ω 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. ***(Slater condition and strong duality):*** *If $f, f_1, \ldots, f_m$ are convex functions and $h_i(\cdot)$ are affine, the Slater condition guarantees the strong duality.*

Proof. Proof of this theorem can be found in Chapter 5.4 of [Vishnoi (2021)]. □

2 KKT (Kahn-Karush-Tucker) Conditions

Define the optimization problem

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

Definition 6. ***(KKT conditions):*** *We say the primal variables $x_*\in\mathbb{R}^d$ and the dual variables $\lambda_*\in\mathbb{R}^m$, $\mu_*\in\mathbb{R}^p$ satisfy KKT conditions if all of the following are satisfied:*

- *(Primal feasibility)* $\forall j\in[m]: f_j(x_*)\leq 0$ *and* $\forall i\in[p]: h_i(x_*)=0$.
- *(Dual feasibility)* $\lambda_*\geq 0$.
- *(Stationarity)* $\partial_x L(x_*,\lambda_*,\mu_*)=0$.
- *(Complementary slackness)* $\forall j\in[m]: \lambda_j^* f_j(x_*)=0$.

Remark: Complementary slackness has the implication that one of the terms in $\lambda_j^* f_j(x_*) = 0, \quad \forall j\in[m]$ must be zero if the other is nonzero. In other words, $f_j(x_*)\neq 0 \Rightarrow f_j(x_*)<0 \Rightarrow \lambda_j^*=0$, and $\lambda_j^*\neq 0\Rightarrow f_j(x_*)=0$.

Theorem 3. ***(Strong Duality and KKT conditions):*** *Let $x_*\in\mathbb{R}^d$ be the primal feasible points and let $\lambda_*\in\mathbb{R}^m$ and $\mu_*\in\mathbb{R}^p$ be the dual feasible points.* ***Strong duality****, i.e.,*

$$f(x_*) = g(\lambda_*,\mu_*)$$

implies that x_*,λ_*,μ_* satisfy ***KKT conditions****. Furthermore, if $f(\cdot), f_1(\cdot),\ldots,f_m(\cdot)$ are* ***convex*** *and $h_1(\cdot)$, $h_2(\cdot)$, $\ldots, h_p(\cdot)$ are affine, then the converse is true:* ***KKT conditions imply strong duality****.*

Remark: Recall the dual value is always not greater than the primal value

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

Thus, when we have zero duality gap ($f(x_*) = g(\lambda_*, \mu_*)$), which implies that x_* is the primal optimal, and λ_*, μ_* are dual optimals. Thus, if $f(\cdot), f_1(\cdot), \ldots, f_m(\cdot)$ are **convex**, $h_1(\cdot)$, $h_2(\cdot)$, $\ldots, h_p(\cdot)$ are affine, and x_*, λ_*, μ_* satisfy KKT, it follows that strong duality is satisfied, and x_* is the primal optimal, and λ_*, μ_* are dual optimals.

3 Example applications of the KKT conditions

In the following examples we work with the following projection function

$$\text{Proj}_C(y) := \arg\min_{x \in C} \|y - x\|_2$$

3.1 Example 1

Projection to ℓ_2 norm ball: $C := \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$**:**

$$\text{Proj}_C(y) := \frac{y}{\max\{1, \|y\|_2\}}$$

Step 1: Getting the Lagrangian $L(x, \lambda)$ by introducing the Lagrangian Multiplier $\lambda \in \mathbb{R}^d_{\geq 0}$. Given the constrained problem

$$\begin{aligned} &\min_{x \in C} \|x - y\|_2^2 \\ &\text{s.t. } \|x\|_2^2 \leq 1, \end{aligned}$$

the Lagrangian can be constructed as:

$$L(x, \lambda) = \|x - y\|_2^2 + \lambda(\|x\|_2^2 - 1).$$

What is the "primal feasibility" of the KKT conditions in this case?

$$x : \|x\|_2^2 \leq 1$$

What is the "dual feasibility" of the KKT conditions in this case?

$$\lambda : \lambda \geq 0$$

Recall the KKT conditions:

- (Stationarity) $\partial_x L(x_*, \lambda_*, \mu_*) = 0.$

- (Complementary slackness) $\forall j \in [m] : \lambda_j^* f_j(x_*) = 0$.

What is the "stationarity" of the KKT conditions in this case?

$$\nabla_x L(x, \lambda) = 2(x - y) + 2\lambda x = 0$$
$$\Leftrightarrow y = (1 + \lambda)x \tag{1}$$

What is the "complementary slackness" of the KKT conditions in this case?

$$\lambda_j^* f_j(x_*) = 0$$
$$\Leftrightarrow \lambda(\|x\|_2^2 - 1) = 0 \tag{2}$$

Now, let us derive the projection formulation based on the KKT conditions.

$$\text{Proj}_C(y) := \arg\min_{x \in C} \|y - x\|_2$$

$$\text{Proj}_C(y) := \frac{y}{\max\{1, \|y\|_2\}}$$

We differentiate the following cases:

- **Case 1:**

  $$\|x\|_2^2 < 1$$

  Then, by complementary slackness (equation (2)):

  $$\lambda = 0$$

  and by stationarity (equation (1)):

  $$y = x$$

  As a result, we have:

  $$\|y\|_2^2 < 1$$

- **Case 2:**

  $$\|x\|_2^2 = 1 \tag{3}$$

  Then, by complementary slackness (equation (2)):

  $$\lambda \neq 0$$

  and by stationarity (equation (1)):

  $$y = (1 + \lambda)x \tag{4}$$

Using equations (3) and (4):

$$\frac{\|y\|_2^2}{(1+\lambda)^2} = 1$$

$$\Leftrightarrow \|y\|_2 = 1+\lambda \tag{5}$$

As a result of equations (4) and (5), we have:

$$x = \frac{y}{\|y\|_2}$$

3.2 Example 2

Projection to ℓ_1 norm ball: $C := \{x \in \mathbb{R}^d : \|x\|_1 \leq 1\}$:
Denote $x = \text{Proj}_C(y)$. Then

$$\begin{cases} x = y & \text{, if } \|y\|_1 \leq 1; \\ x[i] = \text{sign}(y[i])(|y[i]| - \lambda)_+, \forall i \in [d] & \text{, otherwise} \end{cases}$$

where λ is a number such that $\sum_{i=1}^{d}(|y[i]| - \lambda)_+ = 1$, where $(z)_+ := \max\{0, z\}$.

Step 1: Getting the Lagrangian $L(x, \lambda)$ by introducing the Lagrangian Multiplier $\lambda \in \mathbb{R}^d_{\geq 0}$,

$$L(x, \lambda) = \frac{1}{2}\|x - y\|_2^2 + \lambda(\|x\|_1 - 1).$$

What is the "stationarity" of the KKT conditions in this case?

$$\partial_x L(x, \lambda) = 0$$

Recall that:

$$\|x\|_1 = \sum_{i=1} |x[i]|$$

and

$$\text{Subgradient of } \|x\|_1 = \begin{bmatrix} g_x[1] \in \partial(|x[1]|) \\ g_x[2] \in \partial(|x[2]|) \\ \vdots \\ g_x[d] \in \partial(|x[d]|) \end{bmatrix} \in \mathbb{R}^d$$

By the stationarity condition:

$$\partial_x L(x, \lambda) = x - y + \lambda\partial(\|x\|_1) = 0$$

For each $i \in [d]$, we have: $x[i] - y[i] + \lambda g_x[i] = 0$, or equivalently, for each $i \in [d]$: $y[i] = x[i] + \lambda g_x[i]$, where $g_x[i] \in \partial(|x[i]|)$.
We differentiate the following cases:

- **Case 1:** $x[i] = 0$

$$g_0 \in \{-1, 1\}$$

$$y[i] = x[i] + \lambda\{-1, 1\} = \{-\lambda, \lambda\}$$

- **Case 2:** $x[i] > 0$

$$g_x = 1$$

$$y[i] = x[i] + \lambda > \lambda \geq 0$$

- **Case 3:** $x[i] < 0$

$$g_x = -1$$

$$y[i] = x[i] - \lambda < -\lambda \leq 0$$

Thus, for each $i \in [d] : y[i] = x[i] + \lambda g_x[i]$, where $g_x[i] \in \partial(|x[i]|)$, we have

$$x[i] = \text{sign}(y[i])(|y[i]| - \lambda)_+, \forall i \in [d]$$

We will verify the above expression for Case 2, that is for $y[i] > \lambda$, we have $\forall i \in [d]$

$$\begin{aligned} x[i] &= \text{sign}(y[i])(|y[i]| - \lambda)_+ \\ &= \text{sign}(y[i])(y[i] - \lambda) \\ &= y[i] - \lambda \end{aligned}$$

The other two cases can be verified similarly.
Now, by the primal feasibility:

$$\|x\|_1 \leq 1$$

By complementary slackness:

$$\lambda(\|x\|_1 - 1) = 0$$

Then,

- $\|x\|_1 < 1 \Rightarrow \lambda = 0 \Rightarrow x[i] = y[i]$.
- $\|x\|_1 = 1 \Rightarrow \sum_{i=1}^{d} |\text{sign}(y[i])(|y[i]| - \lambda)_+| = 1$ (together with primal feasibility) $\Leftrightarrow \sum_{i=1}^{d} (|y[i]| - \lambda)_+ = 1$.

Bibliographic notes

More information can be found in [Drusvyatskiy (2020)] and [Vishnoi (2021)].

References

[Drusvyatskiy (2020)] Dmitriy Drusvyatskiy. Convex Analysis and Nonsmooth Optimization. 2020.

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