Optimization Theory and Algorithm **Lecture 5 - 09/26/2021**

Lecture 5

Lecturer:Xiangyu Chang *Scribe: Xiangyu Chang*

Edited by: Junbo Hao

1 KKT Optimality Condition

1.1 Conditions for Strong Duality

Recall the primal problem:

$$
\begin{aligned}
p^* = \min_{\mathbf{x}} \ & f_0(\mathbf{x}), \\
s.t. \ & f_i(\mathbf{x}) \leqslant 0, i = 1, \dots, m, \\
& h_j(\mathbf{x}) = 0, j = 1, \dots, l.
\end{aligned}
$$

Lagrangian:

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_i \lambda_i f_i(\mathbf{x}) + \sum_j \nu_j h_j(\mathbf{x}).$$

Lagrange dual function:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in D} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}).$$

Dual problem:

$$
\begin{aligned}
q^* = \max_{\boldsymbol{\lambda}, \boldsymbol{\nu}} \ & g(\boldsymbol{\lambda}, \boldsymbol{\nu}), \\
s.t. \ & \boldsymbol{\lambda} \succeq 0.
\end{aligned}
$$

Weak Duality: $q^* \leqslant p^*$ that is

$$g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \leqslant f_0(\mathbf{x}^*).$$

Thus, q^* is a non-trivial lower bound of p^*. For example, we can use this property to construct a stopping condition as

$$f(\mathbf{x}^t) - f(\mathbf{x}^*) = f(\mathbf{x}^t) - p^* \leqslant f(\mathbf{x}^t) - q^* = f(\mathbf{x}^t) - g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*).$$

Strong Duality: $q^* = p^*$.

Q: What condition can justify the strong duality? Is convexity enough?

Example 1.1. The following example gives us an interesting instance to show the strong duality cannot be justified by the convex property.

$$\begin{aligned} &\min e^{-x} \\ &s.t.\ x^2/y \leqslant 0, \\ &y > 0. \end{aligned}$$

- $p^* = 1$.
- This is a convex problem.
- Lagrange dual function:

$$g(\boldsymbol{\lambda}) = \inf_{(x,y)} \left\{ e^{-x} + \lambda_1 x^2/y - \lambda_2 y \right\} = 0 (\lambda_1 = \lambda_2 = 0).$$

- Dual problem is $q^* = \max_{\boldsymbol{\lambda}} 0 = 0,\ s.t.\ \boldsymbol{\lambda} \succeq 0$.

Convexity alone is not enough to guarantee strong duality.

Theorem 1.2. *Consider the following convex problem*

$$\begin{aligned} &\min_{\mathbf{x}} f_0(\mathbf{x}), \\ &s.t.\ f_i(\mathbf{x}) \leqslant 0, i = 1, \ldots, m, \\ &A\mathbf{x} = \mathbf{b}. \end{aligned}$$

*If there exists $\mathbf{x} \in int(D)$, such that $A\mathbf{x} = \mathbf{b}, f_i(\mathbf{x}) < 0, i = 1, \ldots, m$ (**strictly feasible**), then the strong duality holds.*

Proof. See Page 234 of [**?**]. ■

There are many results that establish conditions on the problem, beyond convexity, under which strong duality holds. These conditions are called **constraint qualifications**. The above condition is called **Slater conditions**.

1.2 Benefit of Strong Duality

Theorem 1.3. *Suppose that $\mathbf{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ are the primal and dual solution of optimization problem of (**??**), and strong duality holds. Then we have the following two facts:*

- $\sum_i \lambda_i^* f_i(\mathbf{x}^*) = 0$. *That is $\lambda_i^* > 0, \Longrightarrow f_i(\mathbf{x}^*) = 0$ or $f_i(\mathbf{x}^*) < 0, \Longrightarrow \lambda_i^* = 0$. This is also called "complementary slackness."*

- $\mathbf{x}^*$ *is the minimizer of* $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$, *that is*

$$\nabla f_0(\mathbf{x}^*) + \sum_i \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_j \nu_j^* \nabla h_j(\mathbf{x}^*) = 0.$$

Proof. Due to the strong duality, then

$$\begin{aligned} p^* = f_0(\mathbf{x}^*) = q^* = \inf_{\mathbf{x}\in D} & \left\{ f_0(\mathbf{x}) + \sum_i \lambda_i^* f_i(\mathbf{x}) + \sum_j \nu_j^* h_j(\mathbf{x}) \right\} \\ & \leqslant f_0(\mathbf{x}^*) + \sum_i \lambda_i^* f_i(\mathbf{x}^*) + \sum_j \nu_j^* h_j(\mathbf{x}^*) \\ & \leqslant f_0(\mathbf{x}^*). \end{aligned}$$

This implies

$$\sum_i \lambda_i^* f_i(\mathbf{x}^*) = 0$$

and $\mathbf{x}^*$ is the minimizer of $L(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$. In addition,

$$\nabla L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = 0 \Longrightarrow \nabla f_0(\mathbf{x}^*) + \sum_i \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_j \nu_j^* \nabla h_j(\mathbf{x}^*) = 0.$$

∎

Under strong duality, given a dual solution $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ any primal solution $\mathbf{x}^*$ solves

$$\min_{\mathbf{x}} f_0(\mathbf{x}) + \sum_i \lambda_i^* f_i(\mathbf{x}) + \sum_j \nu_j^* h_j(\mathbf{x}).$$

This means that we only need to solve an unconstrained problem we have familiar with them.

Example 1.4. Minimization a separable function subject to an equality constraint.

$$\begin{aligned} & \min f_0(\mathbf{x}) = \sum_{i=1}^n f_i(\mathbf{x}_i), \\ & s.t.\ \mathbf{a}^\top \mathbf{x} = b. \end{aligned}$$

$$\begin{aligned} g(\nu) &= \inf_{\mathbf{x}} \left\{ \sum_{i=1}^n f_i(x_i) + \nu(\mathbf{a}^\top \mathbf{x} - b) \right\} \\ &= \inf_{\mathbf{x}} \left\{ \sum_{i=1}^n (f_i(x_i) + \nu a_i x_i) - \nu b \right\} \\ &= -\sum_{i=1}^n \sup_{x_i} \{(-\nu a_i) x_i - f_i(x_i)\} - \nu b \\ &= -\sum_{i=1}^n f_i^*(-\nu a_i) - \nu b. \end{aligned}$$

Then the dual problem is a one-dimensional optimization problem with respect to ν:

$$\max\ -\sum_{i=1}^{n} f_i^*(-\nu a_i) - \nu b.$$

Suppose that we can obtain ν^*, then solve

$$\min_{x_i}\{(f_i(x_i) - a_i\nu^* a_i)\}, i = 1, \ldots, n.$$

Then solve equation $f'(x_i) = -\nu^* a_i$ to obtain x_i^*.

1.3 Karush-Kuhn-Tucker Conditions

- First appeared in publication by Kuhn and Tucker 1951.
- Later people found out that Karush had the condition in his unpublished master's thesis of 1939.
- Finally, it is called the Karush-Kuhn-Tucker conditions.

Theorem 1.5. *(KKT Optimality Conditions) Let* $\mathbf{x}^*$ *and* $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ *be the primal and dual optimal points of optimization problem of* (**??**) *with zero dual gap, then the following KKT conditions hold:*

$$\nabla f_0(\mathbf{x}^*) + \sum_i \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_j \nu_j^* \nabla h_j(\mathbf{x}^*) = 0 \text{ (stationary point)}, \tag{1}$$

$$f_i(\mathbf{x}^*) \leqslant 0, \text{ (primal feasible)} \tag{2}$$

$$h_j(\mathbf{x}^*) = 0, \text{ (primal feasible)} \tag{3}$$

$$\lambda_i^* f_i(\mathbf{x}^*) = 0, \text{ (complementary slackness)} \tag{4}$$

$$\lambda_i \geqslant 0, \text{ (dual feasible)} \tag{5}$$

where $i = 1, \ldots, m$ *and* $j = 1, \ldots, l$.

Proof. Combing the primal and dual feasible conditions and results of Theorem 1.3, we can justify the KKT optimality conditions. ■

Next, let us show some insightful examples

Example 1.6. For the unconstrained optimization, KKT optmality conditions say: $\nabla f(\mathbf{x}^*) = 0$.

Example 1.7. Let us consider the following general convex optimization with linear equality constrains.

$$\min_{\mathbf{x}} f(\mathbf{x}), \tag{6}$$

$$s.t.\ A\mathbf{x} = \mathbf{b}. \tag{7}$$

Based on the KKT optimality conditions, we have

$$\begin{cases} A\mathbf{x}^* = \mathbf{b}, \\ \nabla f(\mathbf{x}^*) + A^\top \lambda^* = 0. \end{cases}$$

Recall that we have obtain these conditions by the general optimality conditions

$$\langle \nabla f(\mathbf{x}^*), \mathbf{y} - \mathbf{x} \rangle \geqslant 0$$

in the previous example.