IE 539: Convex Optimization **KAIST, Fall 2022**
Lecture #18: KKT conditions, Lagrangian duality November 8, 2022
Lecturer: Dabeen Lee

1 Outline

In this lecture, we study

- KKT conditions,
- Lagrangian duality.

2 Karush-Kuhn-Tucker conditions

Remember that x^* is an optimal solution to

$$\min_{x\in C} \quad f(x)$$

where C is a convex set and f is differentiable if and only if

$$\nabla f(x^*)^\top (x - x^*) \geq 0 \quad \forall x \in C.$$

However, the structure of C may be arbitrary, which makes the condition difficult to verify. In this section, we present another way of verifying optimality. Namely, Karu-Kuhn-Tucker conditions, often referred to as KKT conditions.

2.1 Linear constraints

We consider problems of the following structure.

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax \leq b \\ & Cx = d \end{aligned} \tag{18.1}$$

where

- $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$,
- $C \in \mathbb{R}^{\ell\times d}$ and $d \in \mathbb{R}^\ell$.

Theorem 18.1 (KKT conditions for linearly constrained problems)**.** *The linearly constrained problem as in (18.1) satisfies the following.*

1. *(Necessity) If x^* is a feasible solution to (18.1) and $f(x^*)$ is a local minimum, then there exist $\lambda^* \in \mathbb{R}^m_+$ and $\mu^* \in \mathbb{R}^\ell$ such that*

$$\nabla f(x^*)^\top + \lambda^{*\top} A + \mu^{*\top} C = 0 \quad \& \quad \lambda^{*\top}(Ax - b) = 0. \tag{$\star$}$$

2. *(Sufficiency) If f is convex, x^* is a feasible solution to (18.1), and there exist $\lambda^* \in \mathbb{R}^m_+$ and $\mu^* \in \mathbb{R}^\ell$ satisfying ($\star$), then x^* is an optimal solution to (18.1).*

2.2 General convex constraints

We consider problems of the following structure.

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0 \quad \text{for } i = 1, \dots, m \\ & h_j(x) = 0 \quad \text{for } j = 1, \dots, \ell \end{aligned} \tag{18.2}$$

where

- f is convex,
- $g_1, \dots, g_m$ are convex,
- $h_1, \dots, h_\ell$ are affine.

Definition 18.2 (Slater's condition)**.** Suppose that $g_1, \dots, g_k$ are affine and $g_{k+1}, \dots, g_m$ are convex functions that are not affine. Then we say that the problem (18.2) satisfies Slater's condition if there exists a solution $\bar{x}$ such that

$$g_i(\bar{x}) \le 0 \text{ for } i = 1, \dots, k, \quad g_i(\bar{x}) < 0 \text{ for } i = k+1, \dots, m, \quad h_j(\bar{x}) = 0 \text{ for } j = 1, \dots, \ell.$$

Theorem 18.3 (KKT conditions for convex constrained problems)**.** *The convex programming problem as in* (18.2) *satisfies the following.*

1. *(Necessity) Assume that Slater's condition is satisfied. If x^* is a feasible optimal solution to* (18.2)*, then there exist $\lambda^* \in \mathbb{R}^m_+$ and $\mu^* \in \mathbb{R}^\ell$ such that*

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^\ell \mu_j^* \nabla h_j(x^*) = 0 \quad \& \quad \lambda_i^* g_i(x^*) = 0 \text{ for all } i = 1, \dots, m. \tag{$\star\star$}$$

2. *(Sufficiency) If x^* is a feasible solution to* (18.2) *and there exist $\lambda^* \in \mathbb{R}^m_+$ and $\mu^* \in \mathbb{R}^\ell$ satisfying* ($\star\star$)*, then x^* is an optimal solution to* (18.2).

3 Lagrangian duality

We again consider the following optimization problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0 \quad \text{for } i = 1, \dots, m \\ & h_j(x) = 0 \quad \text{for } j = 1, \dots, \ell. \end{aligned} \tag{18.3}$$

We consider the most general setting for which we do not impose the condition that the objective and constraint functions are convex.

3.1 Lagrangian dual problem

The *Lagrangian function* of (18.3) is given by

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

When the objective function f is convex, constraint functions $g_1, \ldots, g_m$ are convex, constraint functions $h_1, \ldots, h_\ell$ are affine, and the multiplier $\lambda \geq 0$, the Lagrangian function is convex in x for any fixed λ and μ. Moreover, the Lagrangian function is affine in λ and μ for any fixed x.

The *Lagrangian dual function* of (18.3) is

$$q(\lambda,\mu) = \inf_x \mathcal{L}(x,\lambda,\mu) = \inf_x \left\{ f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x) \right\}.$$

Notice that the Lagrangian dual function is concave in (λ,μ), regardles of f, $g_1, \ldots, g_m$, and $h_1, \ldots, h_\ell$. This is because $\mathcal{L}(x,\lambda,\mu)$ is affine in λ and μ for any fixed x, and $q(\lambda,\mu)$ is a point-wise minimum of affine functions.

Proposition 18.4. *Let x be a feasible solution to (18.3), and $\lambda \geq 0$. Then*

$$f(x) \geq q(\lambda,\mu).$$

Proof. Since x is feasible, $g_i(x) \leq 0$ for $i = 1, \ldots, m$ and $h_j(x) = 0$ for $j = 1, \ldots, \ell$. Then for any $\lambda \geq 0$, we have

$$\sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x) \leq 0.$$

This implies that

$$f(x) \geq \mathcal{L}(x,\lambda,\mu).$$

Note that

$$q(\lambda,\mu) = \inf_x \mathcal{L}(x,\lambda,\mu) \leq \mathcal{L}(x,\lambda,\mu).$$

Therefore, $f(x) \geq q(\lambda,\mu)$. □

By Proposition 18.4, if (18.3) is unbounded below, the Lagrangian dual function $q(\lambda,\mu) = -\infty$ for any $\lambda \geq 0$.

With the Lagrangian dual function, we can provide a lower bound on the problem (18.3). The *Lagrangian dual problem* is defined as

$$\begin{aligned} \text{maximize} \quad & q(\lambda,\mu) \\ \text{subject to} \quad & \lambda \geq 0. \end{aligned} \tag{18.4}$$

We often call (18.3) as *primal* and (18.4) as the *associated (Lagrangian) dual*. The following result states that the optimal value of the primal is lower bounded by the optimal value of the dual.

Theorem 18.5 (Weak duality)**.** *Consider the problem (18.3) and the associated Lagrangian dual problem (18.4). Then the following statement holds.*

$$\min_{x \in C} f(x) \geq \max_{\lambda \geq 0} q(\lambda,\mu)$$

where $C = \{x : g_i(x) \leq 0 \text{ for } i = 1, \ldots, m,\ h_j(x) = 0 \text{ for } j = 1, \ldots, \ell\}$.

Proof. By proposition 18.4, we know that $f(x) \geq q(\lambda, \mu)$ for any $x \in C$ and $\lambda \geq 0$. Then taking the minimum of $f(x)$ over $x \in C$, it follows that $\min_{x\in C} f(x) \geq q(\lambda, \mu)$. Then taking the maximum of $q(\lambda, \mu)$ over $\lambda \geq 0$, we obtain the desired inequality. □

Theorem 18.5 holds regardless of whether the objective and constraint functions are convex or not. In fact, if we further assume that the objective f is convex and the constraint functions satisfy Slater's condition, then the inequality given in Theorem 18.5 holds with equality.

Theorem 18.6 (Strong duality)**.** *Consider the primal problem* (18.3) *and the associated Lagrangian dual problem* (18.4)*. Assume that the objective function f and the constraint functions $g_1, \ldots, g_m, h_1, \ldots, h_\ell$ are convex. If the primal problem* (18.3) *has a finite optimal value and Slater's condition, given in Definition 18.2, is satisfied, then there exist $\lambda^* \geq 0$ and μ^* such that*

$$\min_{x\in C} f(x) = q(\lambda^*, \mu^*) = \max_{\lambda\geq 0} q(\lambda, \mu)$$

where $C = \{x : g_i(x) \leq 0 \text{ for } i = 1, \ldots, m,\ h_j(x) = 0 \text{ for } j = 1, \ldots, \ell\}$.

3.2 Examples

Consider the following linear program in standard form.

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0. \end{aligned} \tag{18.5}$$

Then the Lagrangian dual function is given by

$$\begin{aligned} q(\lambda, \mu) &= \inf_x \mathcal{L}(x, \lambda, \mu) \\ &= \inf_x \left\{ c^\top x - \lambda^\top x + \mu^\top (Ax - b) \right\} \\ &= -b^\top \mu + \inf_x \left\{ (c - \lambda + A^\top \mu)^\top x \right\}. \end{aligned}$$

Note that $\inf_x \left\{(c - \lambda + A^\top \mu)^\top x\right\} = -\infty$ unless $c - \lambda + A^\top \mu = 0$. Hence, to maximize the Lagrangian dual function $q(\lambda, \mu)$, it is sufficient to consider (λ, μ) satisfying $c - \lambda + A^\top \mu = 0$. Therefore, the associated Lagrangian dual problem is equivalent to

$$\begin{aligned} \text{maximize} \quad & -b^\top \mu \\ \text{subject to} \quad & c - \lambda + A^\top \mu = 0, \\ & \lambda \geq 0. \end{aligned} \tag{18.6}$$

In fact, we can eliminate the variable from the constraints $c + A^\top \mu \geq \lambda$ and $\lambda \geq 0$, and they can be equivalently written as $c + A^\top \mu \geq 0$. Moreover, the variables μ are unrestricted, so we can replace μ by $-\mu$. Then (18.6) is equivalent to

$$\begin{aligned} \text{maximize} \quad & b^\top \mu \\ \text{subject to} \quad & A^\top \mu \leq c, \end{aligned} \tag{18.7}$$

which is the dual linear program for (18.5).

Next we consider the following quadratic program.

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}x^\top Qx + p^\top x \\ \text{subject to} \quad & Ax = b \end{aligned} \tag{18.8}$$

where Q is positive definite and thus is invertible. The corresponding Lagrangian function is given by

$$\begin{aligned} \mathcal{L}(x,\mu) &= \frac{1}{2}x^\top Qx + p^\top x + \mu^\top(Ax-b) \\ &= -b^\top\mu + \left(\frac{1}{2}x^\top Qx + (p + A^\top\mu)^\top x\right). \end{aligned}$$

Then

$$\nabla_x \mathcal{L}(x,\mu) = Qx + (p + A^\top\mu),$$

and therefore, $\nabla_x \mathcal{L}(x,\mu) = 0$ if and only if $x = -Q^{-1}(p + A^\top\mu)$. This in turn implies that the Lagrangian dual function is given by

$$\begin{aligned} q(\mu) &= \inf_x \mathcal{L}(x,\mu) \\ &= \mathcal{L}\left(-Q^{-1}(p + A^\top\mu), \mu\right) \\ &= -b^\top\mu - \frac{1}{2}(p + A^\top\mu)^\top Q^{-1}(p + A^\top\mu) \\ &= -\frac{1}{2}\mu^\top AQ^{-1}A^\top\mu - (b + AQ^{-1}p)^\top\mu - \frac{1}{2}p^\top p. \end{aligned}$$

Hence, the Lagrangian dual problem is

$$\max_\mu \left\{-\frac{1}{2}\mu^\top AQ^{-1}A^\top\mu - (b + AQ^{-1}p)^\top\mu\right\}.$$

3.3 Lagrangian dual for conic programming

Consider the following conic programming problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \preceq_{K_i} 0 \quad \text{for } i = 1,\ldots,m \\ & h_j(x) = 0 \quad \text{for } j = 1,\ldots,\ell \end{aligned} \tag{18.9}$$

where $K_1,\ldots,K_m$ are proper cones. Remember that $g_i(x) \preceq_{K_i} 0$ means $-g_i(x) \in K_i$. Moreover, recall that the dual cone of a cone K is given by

$$K^* = \{y : \ y^\top x \geq 0 \ \forall x \in K\}.$$

As we picked a nonnegative multiplier $\lambda \geq 0$ to define the Lagrangian function, we pick a multiplier λ from the dual cone K^*. The Lagrangian function of (18.9) is given by

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

where $\lambda_i \in K_i^*$ is now a vector from the dual cone of K_i for each i. Then the Lagrangian dual function is similarly defined as $q(\lambda, \mu) = \inf_x \mathcal{L}(x, \lambda, \mu)$. The Lagrangian dual problem is given by

$$\begin{aligned} \text{maximize} \quad & q(\lambda, \mu) \\ \text{subject to} \quad & \lambda_i \geq_{K_i^*} 0 \quad \text{for } i = 1, \ldots, m. \end{aligned} \tag{18.10}$$

As an example, we consider the following semidefinite program.

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & \sum_{i=1}^{d} x_i A_i \geq_{S_+^m} B \end{aligned} \tag{18.11}$$

where S_+^m denotes the PSD cone containing all $m \times m$ PSD matrices. We learned that the PSD cone is self-dual, so the dual of S_+^m is itself. Let $Y \in S_+^m$, and consider the associated Lagrangian dual function.

$$q(Y) = \inf_x \mathcal{L}(x, Y) = \inf_x \left\{ c^\top x - \sum_{i=1}^{d} x_i \mathrm{tr}(Y^\top A_i) + \mathrm{tr}(Y^\top B) \right\}.$$

Note that the Lagrangian dual function $q(Y)$ has a finite value if and only if $c_i = \mathrm{tr}(Y^\top A_i)$ for every $i \in [d]$. Then the Lagrangian dual problem is given by

$$\begin{aligned} \text{maximize} \quad & \mathrm{tr}(Y^\top B) \\ \text{subject to} \quad & \mathrm{tr}(Y^\top A_i) = c_i \quad \text{for } i = 1, \ldots, d. \\ & Y \in S_+^m \end{aligned} \tag{18.12}$$