Lecture-04: Review of constrained optimization

1 Constrained Optimization

Problem 1 (Primal problem). Consider a cost function $f : \mathbb{R}^d \to \mathbb{R}$ and a constraint function $g : \mathbb{R}^d \to \mathbb{R}^m$. The **primal problem** is $p^* \triangleq \inf\{f(x) : x \in \mathcal{X}\}$, where the constraint set is

$$\mathcal{X} \triangleq \bigcap_{i=1}^{m} \left\{x \in \mathbb{R}^d : g_i(x) \leqslant 0\right\}. \quad (1)$$

Definition 1.1 (Lagrangian). For the Problem 1, we define an associated Lagrangian function $\mathcal{L} : \mathbb{R}^d \times \mathbb{R}^m_+ \to \mathbb{R}$ for Lagrange or dual variables $\alpha \in \mathbb{R}^m_+$ and primal variables $x \in \mathbb{R}^d$, as

$$\mathcal{L}(x,\alpha) \triangleq f(x) + \langle \alpha, g(x) \rangle. \quad (2)$$

Definition 1.2 (Dual function). The dual function $F : \mathbb{R}^m_+ \to \mathbb{R}$ associated with the Problem 1 is defined for dual variables $\alpha \in \mathbb{R}^m_+$ as

$$F(\alpha) \triangleq \inf\left\{\mathcal{L}(x,\alpha) : x \in \mathbb{R}^d\right\}. \quad (3)$$

Definition 1.3. A function $h : \mathcal{X} \to \mathbb{R}$ is said to be affine if it can be defined as $x \mapsto h(x) \triangleq \langle w, x \rangle + b$ for all $x \in \mathcal{X} \subseteq \mathbb{R}^d$ and some $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$.

Theorem 1.4. *The following are true for the dual function $F : \mathbb{R}^m_+ \to \mathbb{R}$ defined in* (3) *for the Problem 1.*
(a) F is concave in $\alpha \in \mathbb{R}^m_+$.
(b) $F(\alpha) \leqslant \mathcal{L}(x,\alpha)$ for all $\alpha \in \mathbb{R}^m_+$ and $x \in \mathbb{R}^d$.
(c) $F(\alpha) \leqslant p^*$ for all $\alpha \in \mathbb{R}^m_+$.

Proof. Recall that $\mathcal{L}(\alpha) = f(x) + \langle \alpha, g(x) \rangle$ is an affine function of $\alpha \in \mathbb{R}^m_+$, and $F(\alpha) = \inf_x \mathcal{L}(x,\alpha)$.
(a) Let $\beta \in [0,1]$, $\alpha_1, \alpha_2 \in \mathbb{R}^m_+$, and $x \in \mathcal{X}$. Since Lagrangian $\mathcal{L}$ is affine in α, we obtain that

$$F(\beta\alpha_1 + \bar{\beta}\alpha_2) = \inf_x \left[\beta\mathcal{L}(x,\alpha_1) + \bar{\beta}\mathcal{L}(x,\alpha_2)\right] \geqslant \beta \inf_x \mathcal{L}(x,\alpha_1) + \bar{\beta} \inf_x \mathcal{L}(x,\alpha_2) = \beta F(\alpha_1) + \bar{\beta} F(\alpha_2).$$

(b) From the definition of F, it follows that $F(\alpha) \leqslant \mathcal{L}(x,\alpha)$ for all $x \in \mathbb{R}^d$.
(c) Recall that $g_i(x) \leqslant 0$ for all $x \in \mathcal{X}$ and $i \in [m]$, and hence $\langle \alpha, g(x) \rangle \leqslant 0$ for all $x \in \mathcal{X}$ and $\alpha \in \mathbb{R}^m_+$. Therefore, $\mathcal{L}(x,\alpha) \leqslant f(x)$ for all $x \in \mathcal{X}$, and hence $F(\alpha) \leqslant \inf_{x \in \mathcal{X}} \mathcal{L}(x,\alpha) \leqslant p^*$ and the result follows. □

Problem 2 (Dual problem). The dual problem associated with primal problem defined in Problem 1 is

$$d^* \triangleq \max\{F(\alpha) : \alpha \in \mathbb{R}^m_+\}.$$

Remark 1. From the properties of dual function $F : \mathbb{R}^m_+ \to \mathbb{R}$ in Theorem 1.4, we obtain that F is concave in $\alpha \in \mathbb{R}^m_+$. Since $\mathbb{R}^m_+$ is a convex set, it follows that the dual problem is convex. We further observe that the optimal value of dual problem $d^* \leqslant p^*$. The difference of optimal values $(p^* - d^*)$ is called the **duality gap**. For a primal problem, the **strong duality** holds if the duality gap is zero, or $d^* = p^*$.

1.1 Saddle point and necessary conditions

Definition 1.5 (Saddle point). For a Lagrangian $\mathcal{L} : \mathbb{R}^d \times \mathbb{R}^m_+ \to \mathbb{R}$, a saddle point (x^0, α^0) sastifies

$$\sup_{\alpha \in \mathbb{R}^m_+} \mathcal{L}(x^0,\alpha) \leqslant \mathcal{L}(x^0,\alpha^0) \leqslant \inf_{x \in \mathbb{R}^d} \mathcal{L}(x,\alpha^0). \quad (4)$$

Theorem 1.6 (Sufficient condition). *For the primal problem defined in Problem 1, if (x^0, α^0) is a saddle point of the associated Lagrangian $\mathcal{L}$, then the following statements are true.*
(a) $F(\alpha^0) = \mathcal{L}(x^0, \alpha^0)$.
(b) $x^0 \in \mathcal{X}$.
(c) $\langle \alpha^0, g(x^0) \rangle = 0$.
(d) $p^* = f(x^0) = F(\alpha^0)$.

Proof. Let (x^0, α^0) be the saddle point of the Lagrangian $\mathcal{L}$ associated with the Problem 1.

(a) Recall that $F(\alpha) = \inf_{x \in \mathbb{R}^d} \mathcal{L}(x, \alpha)$ for all $\alpha \in \mathbb{R}^m_+$ from the definition of dual function. Hence from the second saddle point condition in (4), we get that $\mathcal{L}(x^0, \alpha^0) \leqslant F(\alpha^0)$. Applying Theorem 1.4(b) at $\alpha = \alpha^0$, we have $F(\alpha^0) \leqslant \mathcal{L}(x^0, \alpha^0)$. It follows that $F(\alpha^0) = \mathcal{L}(x^0, \alpha^0)$.

(b) We assume that $x^0 \notin \mathcal{X}$, then there exists an $i \in [m]$ such that $g_i(x^0) > 0$. We take $\alpha \in \mathbb{R}^m_+$ such that $\alpha_i > \alpha^0_i$ and $\alpha_j = \alpha^0_j$ for all $j \neq i$. Then, we observe that

$$\mathcal{L}(x^0, \alpha) - \mathcal{L}(x^0, \alpha^0) = \left\langle \alpha - \alpha^0, g(x^0) \right\rangle = (\alpha_i - \alpha^0_i) g_i(x^0) + \sum_{j \neq i} (\alpha_j - \alpha^0_j) g_j(x^0) \geqslant 0.$$

This contradicts the first saddle point condition in (4), and hence $x^0 \in \mathcal{X}$.

(c) Since $x^0 \in \mathcal{X}$, we have $\langle \alpha, g(x^0) \rangle \leqslant 0$ for all $\alpha \in \mathbb{R}^m_+$. It follows that $\mathcal{L}(x^0, \alpha) \leqslant f(x^0)$ for all $\alpha \in \mathbb{R}^m_+$ with the upper bound achieved for α such that $\langle \alpha, g(x^0) \rangle = 0$, i.e. $\sup \{\mathcal{L}(x^0, \alpha) : \alpha \in \mathbb{R}^m_+\} = f(x^0)$. This condition together with the first saddle point condition in (4) implies that

$$\mathcal{L}(x^0, \alpha^0) \leqslant \sup_{\alpha \in \mathbb{R}^m_+} \mathcal{L}(x^0, \alpha) \leqslant \mathcal{L}(x^0, \alpha^0).$$

The implies that α^0 is the supremum achieving dual variable for $x^0 \in \mathcal{X}$ and $\langle \alpha^0, g(x^0) \rangle = 0$.

(d) Recall that $p^* = \inf_{x \in \mathcal{X}} f(x)$. Since $x^0 \in \mathcal{X}$, we have $p^* \leqslant f(x^0)$. From the previous part, we have $\mathcal{L}(x^0, \alpha^0) = f(x^0)$. From the the second saddle point condition in (4) and the definition of dual function F, we have $\mathcal{L}(x^0, \alpha^0) \leqslant F(\alpha^0)$. Further, we have $F(\alpha^0) \leqslant p^*$ from Theorem 1.4(c). Combining these results, we get

$$p^* \leqslant f(x^0) = \mathcal{L}(x^0, \alpha^0) \leqslant F(\alpha^0) \leqslant p^*.$$

This implies that there is no duality gap and the saddle point (x^0, α^0) solves the primal and the dual problems. □

Corollary 1.7. *Consider the primal problem in Problem 1 with the cost function f and constraint functions $(g_i : i \in [m])$ being differentiable. Let (x^0, α^0) be a saddle point of the associated Lagrangian, then $\nabla_x \mathcal{L}(x^*, \alpha^*) = \nabla_x f(x^0) + \langle \alpha^0, \nabla_x g(x^0) \rangle = 0$.*

1.2 Convexity, constraint qualification, and sufficient conditions

Definition 1.8. Consider a set A. The interior of a set A is defined as $A^o \triangleq \cup \{U : U \subseteq A, U \text{ open}\}$. The closure of a set A is defined as $\bar{A} \triangleq \cap \{C : A \subseteq C, C \text{ closed}\}$.

Definition 1.9 (Strong constraint qualification). The strong constraint qualification or **Slater's condition** is defined as the existence of a point $x \in \mathcal{X}^o$ such that $g_i(x) < 0$ for all $i \in [m]$.

Theorem 1.10 (Strong necessary condition). *Let the cost function f and constraints g_i for $i \in [m]$ be convex functions, such that the Slater's condition holds, and x^* be the solution of the Problem 1 . Then, there exists $\alpha^* \in \mathbb{R}^m_+$ such that (x^*, α^*) is a saddle point of the associated Lagrangian $\mathcal{L}$.*

Definition 1.11 (Weak constraint qualification). The weak constraint qualification or **weak Slater's condition** is defined as the existence of a point $x \in \mathcal{X}^o$ such that for each $i \in [m]$ either $g_i(x) < 0$ or $g_i(x) = 0$ and g_i affine.

Theorem 1.12 (Weak necessary condition). *Let the cost function f and constraints g_i for $i \in [m]$ be convex differentiable functions, such that the weak Slater's condition holds, and x^* be the solution of the Problem 1 . Then, there exists $\alpha^* \in \mathbb{R}^m_+$ such that (x^*, α^*) is a saddle point of the associated Lagrangian $\mathcal{L}$.*

Remark 2. The strong duality holds when the primal problem is convex with qualifying constraints.

Theorem 1.13 (Karush-Kuhn-Tucker (KKT)). *Let the cost function f and constraint functions g_i for all $i \in [m]$ be convex and differentiable functions, such that the constraints are qualified. Then $x^* \in \mathbb{R}^d$ is a solution of the constrained problem iff there exists $\alpha^* \in \mathbb{R}_+^m$ such that*

$$\nabla_x f(x^*) + \sum_{i=1}^{m} \alpha_i^* \nabla_x g_i(x^*) = 0, \qquad g_i(x^*) \leqslant 0, i \in [m], \qquad \sum_{i=1}^{m} \alpha_i^* g_i(x^*) = 0. \tag{5}$$

Proof. We will show this in two steps. We will first assume that $x^* \in \mathcal{X}$ is solution of the constrained problem and show that the there exists $\alpha^* \in \mathbb{R}_+^m$ such that the three KKT conditions are met. Conversely, we will assume that the pair $(x^*, \alpha^*) \in \mathbb{R}^d \times \mathbb{R}_+^m$ satisfies three KKT conditions, and show that $f(x) \geqslant f(x^*)$ for all $x \in \mathcal{X}$.

$\Longrightarrow$ From the necessary condition theorem, it follows that if $x^* \in \mathcal{X}$ is a solution to the primal problem, then there exists dual variables $\alpha^* \in \mathbb{R}_+^m$ such that (x^*, α^*) is a saddle point of the Lagrangian. Then it follows from Corollary 1.7, the first KKT condition holds. From Theorem 1.6, we have $x^* \in \mathcal{X}$, and hence constraint $g_i(x^*) \leqslant 0$ is satisfied for all $i \in [m]$ and the second KKT condition holds. From Theorem 1.6, we have $\langle \alpha^*, g(x^*) \rangle = 0$ and the third KKT condition holds.

$\Longleftarrow$ We first observe from the convexity of f, that for any $x \in \mathbb{R}^{\mathbb{N}}$, we have

$$f(x) - f(x^*) \geqslant \langle \nabla_x f(x^*), x - x^* \rangle. \tag{6}$$

From the first KKT condition, we get $\langle \nabla_x f(x^*), x - x^* \rangle = -\sum_{i=1}^{m} \alpha_i^* \langle \nabla_x g_i(x^*), x - x^* \rangle$. From the convexity of g_i for all $i \in [m]$ and the third KKT condition, we have

$$-\sum_{i=1}^{m} \alpha_i^* \langle \nabla_x g_i(x^*), x - x^* \rangle \geqslant -\sum_{i=1}^{m} \alpha_i^* (g_i(x) - g_i(x^*)) = -\sum_{i=1}^{m} \alpha_i^* g_i(x). \tag{7}$$

Recall that any point $x \in \mathcal{X}$ satisfies the constraint $g_i(x) \leqslant 0$ for each $i \in [m]$. Thus, combining the inequalities in (6) and (7), we get $f(x) - f(x^*) \geqslant 0$ for all $x \in \mathcal{X}$.

□