CO250: Spring 2019

Proof of the KKT theorem (Theorem 7.9 in the text)

Theorem 1 (Theorem 7.9 in the text). Let (NLP): $\min\{f(x): \ g_i(x) \le 0 \ \text{ for all } i = 1, \dots, m\}$ be a convex NLP, where $f, g_1, \dots, g_m : \mathbb{R}^n \mapsto \mathbb{R}$ are (convex and) differentiable. Let $\bar{x} \in \mathbb{R}^n$ be a feasible solution.

(a) If the *KKT optimality conditions*

$$-\nabla f(\bar{x}) \in \text{cone}\{\nabla g_i(\bar{x}) : i \in J(\bar{x})\} \tag{1}$$

hold, then $\bar{x}$ is an optimal solution to (NLP).

(b) Suppose that (NLP) has a Slater point. Then, if $\bar{x}$ is an optimal solution to (NLP), the KKT conditions (1) hold.

Thus, Theorem 1 states that for a convex NLP, the KKT conditions (1) are always *sufficient* for optimality, and if the NLP has a Slater point, then they are also *necessary* for optimality.

Proof.

Part (a). Suppose that the KKT conditions (1) hold at $\bar{x}$. This means there exist $y_i \ge 0$, $i = 1, \dots, m$, such that

$$-\nabla f(\bar{x}) = \sum_{i \in J(\bar{x})} y_i \nabla g_i(\bar{x}). \tag{2}$$

Consider any feasible solution x to (NLP). Consider the g_is for $i \in J(\bar{x})$. Since the g_is and f are convex functions, we have

$$0 \ge g_i(x) \ge \underbrace{g_i(\bar{x})}_{=0} + (\nabla g_i(\bar{x}))^\top (x - \bar{x}) = (\nabla g_i(\bar{x}))^\top (x - \bar{x}) \qquad \text{for all } \ i \in J(\bar{x}) \tag{3}$$

$$f(x) \ge f(\bar{x}) + (\nabla f(\bar{x}))^\top (x - \bar{x}). \tag{4}$$

Substituting the expression for $\nabla f(\bar{x})$ from (2) into (4) and utilizing (3), we obtain that

$$f(x) \ge f(\bar{x}) - \sum_{i \in J(\bar{x})} y_i (\nabla g_i(\bar{x}))^\top (x - \bar{x}) \ge f(\bar{x}).$$

Thus, $\bar{x}$ is an optimal solution to (NLP).

Part (b). Suppose that $\bar{x}$ is an optimal solution to (NLP). Suppose, for a contradiction, that the KKT conditions (1) are not satisfied. We will show that we can then find a direction $\bar{d} \in \mathbb{R}^n$ and a suitably small $\kappa > 0$ such that $\bar{x} + \kappa\bar{d}$ is a feasible solution and $f(\bar{x} + \kappa\bar{d}) < f(\bar{x})$, thereby contradicting the optimality of $\bar{x}$.
The following claim will be used to show that if $\bar{d}$ satisfies certain conditions, then we can find a suitable $\kappa > 0$. We present the claim without proof; the proof however is not difficult and follows from Taylor's expansion.

Claim 2. Let $h : \mathbb{R}^n \mapsto \mathbb{R}$ be a differentiable function. Let $\bar{x}, \bar{d} \in \mathbb{R}^n$.

(i) For any $\delta > 0$, there exists a small enough $\epsilon > 0$ such that

$$h(\bar{x} + \kappa\bar{d}) \le h(\bar{x}) + \delta \qquad \text{for all } \ 0 \le \kappa \le \epsilon$$

(ii) Suppose that $\bar{d}^\top \nabla h(\bar{x}) < 0$. Then, there exists a small enough $\epsilon > 0$ such that

$$h(\bar{x} + \kappa\bar{d}) < h(\bar{x}) \qquad \text{for all } \ 0 < \kappa \le \epsilon.$$

To obtain a suitable vector $\bar{d}$, we start with our assumption that the KKT conditions (1) do not hold. This implies that the following LP is infeasible.

$$\begin{array}{lrcl} \min & & \mathbb{0}^\top y & \\ \text{s.t.} & \sum_{i\in J(\bar{x})} y_i \nabla g_i(\bar{x}) & = & -\nabla f(\bar{x}) \\ & y & \geq & \mathbb{0}. \end{array} \tag{P}$$

The dual of (P) is

$$\begin{array}{lll} \max & -d^\top \nabla f(\bar{x}) & \\ \text{s.t.} & d^\top \nabla g_i(\bar{x}) \leq 0 & \text{for all } \; i \in J(\bar{x}). \end{array} \tag{D}$$

Now $d = \mathbb{0}$ is feasible for (D), and because (P) is infeasible, (D) must be unbounded (by duality). In particular, this means there exists $\hat{d} \in \mathbb{R}^n$ such that

$$\hat{d}^\top \nabla f(\bar{x}) < 0, \qquad \hat{d}^\top \nabla g_i(\bar{x}) \leq 0 \qquad \text{for all } \; i \in J(\bar{x}). \tag{5}$$

Now $\hat{d}$ is "almost" the direction we need to apply Claim 2. The problem is that we could have $\hat{d}^\top \nabla g_i(\bar{x}) = 0$ for some $i \in J(\bar{x})$, so we will not be able to ensure that $g_i(\bar{x} + \kappa\hat{d}) \leq 0$, even for arbitrarily small $\kappa > 0$. To fix this, we use the existence of a Slater point $\hat{x}$. By convexity of the g_is, we have

$$0 > g_i(\hat{x}) \geq \underbrace{g_i(\bar{x})}_{=0} + (\nabla g_i(\bar{x}))^\top (\hat{x} - \bar{x}) = (\nabla g_i(\bar{x}))^\top (\hat{x} - \bar{x}) \qquad \text{for all } \; i \in J(\bar{x}). \tag{6}$$

By (5) and (6), we infer that if we choose σ small enough and set $\bar{d} := \hat{d} + \sigma(\hat{x} - \bar{x})$, then we have

$$\bar{d}^\top \nabla f(\bar{x}) < 0, \qquad \bar{d}^\top \nabla g_i(\bar{x}) < 0 \qquad \text{for all } \; i \in J(\bar{x}). \tag{7}$$

We now apply Claim 2 to three groups of functions, taking $\bar{x}$ and $\bar{d}$ to be the vector defined above. For every $i \in J(\bar{x})$, since $\bar{d}^\top \nabla g_i(\bar{x}) < 0$, by part (ii) of Claim 2, there exsits $\epsilon_i > 0$ such that $g_i(\bar{x} + \kappa\bar{d}) < g_i(\bar{x}) = 0$ for all $0 < \kappa \leq \epsilon_i$. Define

$$\delta := \min\{|g_i(\bar{x})| : i \notin J(\bar{x})\},$$

which is positive. For every $i \notin J(\bar{x})$, by part (i) of Claim 2, there exists $\epsilon_i > 0$ such that $g_i(\bar{x} + \kappa\bar{d}) \leq g_i(\bar{x}) + \delta \leq 0$ for all $0 \leq \kappa \leq \epsilon_i$. Finally, for $f(x)$, because $\bar{d}^\top \nabla f(\bar{x}) < 0$, by part (ii) of Claim 2, there exists $\epsilon_f > 0$ such that $f(\bar{x} + \kappa\bar{d}) < f(\bar{x})$ for all $0 < \kappa \leq \epsilon_f$.
Now, take κ to be the smallest among ϵ_f and all ϵ_i for $i = 1, \ldots, m$. We then have $g_i(\bar{x} + \kappa\bar{d}) \leq 0$ for all $i = 1, \ldots, m$, and $f(\bar{x} + \epsilon\bar{d}) < f(\bar{x})$, which yields the contradiction. This means that our first assumption that (1) does not hold is incorrect. $\square$