Class 3: KKT conditions and constraint qualifications

With the tools given in the last two classes, we can now prove the Karush-Kuhn-Tucker (KKT) conditions for the following NLP problem:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X, \end{array} \tag{NLP}$$

where

$$X := \{x \in \mathbb{R}^n \mid g(x) \leq 0\} \tag{1}$$

is the feasible set. Here, for simplicity, we deal only with inequality constraints. The following analysis is similar when we add equalities. Recall that we assume that the functions $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g\colon \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. Moreover, we write $g := (g_1, \ldots, g_m)$ with $g_i\colon \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$.

Definition 1. *Let $A(x) := \{i \in \{1, \ldots, m\} \mid g_i(x) = 0\}$ be the set of* active indices *at $x \in \mathbb{R}^n$. The* linearized cone *of X at x is defined by*

$$L_X(x) := \{d \in \mathbb{R}^n \mid \langle \nabla g_i(x), d\rangle \leq 0, i \in A(x)\}.$$

The above definition says that the linearized cone $L_X(x)$ is the set of vectors that make obtuse angle with all the gradients $\nabla g_i(x)$ associated to the active constraints. It can be seen that $L_X(x)$ is a convex cone. Furthermore, although $L_X(x)$ is not equivalent to the tangent set $T_X(x)$, the inclusion $T_X(x) \subseteq L_X(x)$ does hold. The so-called *constraint qualifications* (CQ) are conditions under which the linearized cone $L_X(x)$ is similar to the tangent cone $T_X(x)$. In the next theorem, we will prove the KKT conditions for (NLP) by assuming that one of these CQs holds.

Theorem 1. *Let $x^* \in \mathbb{R}^n$ be a local minimizer of problem (NLP) and assume that $L_X(x^*) \subseteq \operatorname{co} T_X(x^*)$ holds. Then, there exist $\lambda \in \mathbb{R}^m$ such that*

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) = 0,$$
$$\lambda_i \geq 0, \quad g_i(x^*) \leq 0, \quad \lambda_i g_i(x^*) = 0, \qquad i = 1, \ldots, m.$$

In other words, (x^*, λ) satisfies the Karush-Kuhn-Tucker (KKT) conditions.

Proof. Since x^* is a local minimizer, we have

$$-\nabla f(x^*) \in N_X(x^*)$$

from [Class 1, Theorem 1]. Moreover, we obtain

$$L_X(x^*)^\circ \supseteq (\operatorname{co} T_X(x^*))^\circ = T_X(x^*)^\circ = N_X(x^*),$$

where the first inclusion holds from this theorem's assumption and [Class 2, Proposition 1(a)], the second equality is satisfied from [Class 2, Proposition 1(b)], and the third one is just the definition of normal cone, given in [Class 1, Definition 5]. Then, we can write

$$-\nabla f(x^*) \in L_X(x^*)^\circ. \tag{2}$$

Now, using Farkas' lemma [Class 2, Lemma 2], we obtain

$$L_X(x^*)^\circ = \left\{ d \in \mathbb{R}^n \;\middle|\; d = \sum_{i \in A(x^*)} \lambda_i \nabla g_i(x^*), \lambda_i \geq 0 \right\}.$$

From (2), it means that there exists $\lambda_i \geq 0$ for $i \in A(x^*)$ such that

$$-\nabla f(x^*) = \sum_{i \in A(x^*)} \lambda_i \nabla g_i(x^*).$$

By defining $\lambda_i = 0$ for $i \in \{1, \dots, m\} \setminus A(x^*)$, we have

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) = 0.$$

Moreover, since x^* is a local minimizer, it is also feasible, that is, $g_i(x^*) \leq 0$ for all i. Furthermore, from the definition of λ, the conditions $\lambda_i \geq 0$ and $\lambda_i g_i(x^*) = 0$, $i = 1, \dots, m$ clearly hold. Therefore, the KKT conditions are satisfied. □

In the next result, we show that the condition $L_X(x) \subseteq \operatorname{co} T_X(x)$ is automatically satisfied when the constraints are all linear.

Proposition 1. *Assume that the feasible set (1) is defined with $g_i(x) := \langle a^i, x \rangle + b_i$, with $a^i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ for all $i = 1, \dots, m$. Then, the condition $L_X(x) \subseteq \operatorname{co} T_X(x)$ holds for all $x \in X$.*

Proof. Let $d \in L_X(x)$. Since $\nabla g_i(x) = a^i$, from the definition of $L_X(x^*)$, we have $\langle a^i, d \rangle \leq 0$ for all $i \in A(x)$. Define $x^k := x + t_k d$, with $t_k > 0$ and $t_k \to 0$. Clearly, we obtain $x^k \to x$. Also, since

$$g_i(x^k) = \langle a^i, x^k \rangle + b_i = \langle a^i, x \rangle + b_i + t_k \langle a^i, d \rangle = g_i(x) + t_k \langle a^i, d \rangle$$

holds and $x \in X$, for sufficiently large k, we have $g_i(x^k) \leq 0$. This means that $x^k \in X$ when k is large enough. Moreover, defining $\alpha_k := 1/t_k$, then $\alpha_k \geq 0$ and the following holds:

$$\lim_{k \to \infty} \alpha_k (x^k - x) = d.$$

Therefore, $d \in L_X(x) \subseteq \operatorname{co} T_X(x)$, and the conclusion follows. □

Definition 2. *Given $x \in X$, the following are CQs for problem (NLP):*

(a) Guignard CQ: $L_X(x) \subseteq \operatorname{co} T_X(x)$.

(b) Abadie CQ: $L_X(x) \subseteq T_X(x)$.

The condition used in the previous proof of the KKT conditions is the Guignard CQ, which is in a sense the weakest CQ that ensures that KKT conditions are necessary optimality conditions. However, in general, for nonlinear constraints, it is difficult to verify the Guignard CQ or even the Abadie CQ. Because of this, other CQs were proposed in the literature, and here we list some of them that are well-known.

Definition 3. *Given $x \in X$, the following are CQs for problem (NLP):*

(a) Linear independence CQ (LICQ): $\nabla g_i(x)$ with $i \in A(x)$ are linearly independent.

(b) Slater CQ: g_i is convex and there exists $\hat{x} \in \mathbb{R}^n$ such that $g_i(\hat{x}) < 0$ for all $i = 1, \dots, m$.

(c) Mangasarian-Fromovitz CQ (MFCQ): There exists $d \in \mathbb{R}^n$ such that

$$\langle \nabla g_i(x), d \rangle < 0, \quad i \in A(x).$$

When the problem has only inequality constraints, the MFCQ is also called Cottle CQ. Let us now show the relation between these CQs.

Theorem 2. *The linear independence CQ implies Mangasarian-Fromovitz CQ.*

Proof. Assume that LICQ holds at $x \in \mathbb{R}^n$. Let G be a matrix with columns $\nabla g_i(x)$ with $i \in A(x)$. Then, from LICQ, the rank of G is $|A(x)|$. Thus, the matrix $G^\top G \in \mathbb{R}^{|A(x)| \times |A(x)|}$ is nonsingular, and so there exists a vector z such that

$$G^T G z = \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix}.$$

By defining $d := Gz$, we obtain

$$\langle \nabla g_i(x), d \rangle = -1 < 0, \quad i \in A(x),$$

and we conclude that MFCQ is satisfied at x. □

Theorem 3. *Slater CQ implies Mangasarian-Fromovitz CQ.*

Proof. Given $x \in \mathbb{R}^n$, since g_i is convex,

$$g_i(\hat{x}) \geq g_i(x) + \langle \nabla g_i(x), \hat{x} - x \rangle$$

holds for all i. Letting $i \in A(x)$, we have $g_i(x) = 0$. Thus, if $d = \hat{x} - x$, we obtain $g_i(\hat{x}) \geq \langle \nabla g_i(x), d \rangle$. Now, from Slater CQ, we conclude that $\langle \nabla g_i(x), d \rangle \leq 0$, as we claimed. □

It is clear that LICQ or MFCQ does not necessarily imply Slater CQ because of its convexity assumption. It is also known that MFCQ implies Abadie CQ, which in turn implies Guignard CQ.

1st part by Ellen H. Fukuda (e-mail: ellen(at)i.kyoto-u.ac.jp, where (at) = @)