TMA947 / MMG621 — Nonlinear optimization

Lecture 6 — The Karush-Kuhn-Tucker conditions

Emil Gustavsson, Zuzana Nedělková

November 13, 2017

Consider a problem of the form

$$\min f(\boldsymbol{x}), \tag{1a}$$

$$\text{subject to } g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \dots, m \tag{1b}$$

where now $f : \mathbb{R}^n \to \mathbb{R}$, and $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$ are all C^1, i.e., we take S to be of the form $S := \{\boldsymbol{x} \in \mathbb{R}^n \mid g_i(\boldsymbol{x}) \leq 0, i = 1, \dots, m\}$.

1 KKT conditions

We begin by developing the KKT conditions when we assume some regularity of the problem. We assume that the problem considered is well behaved, and postpone the issue of whether any given problem is well behaved until later.

Definition 1 (Abadie's constraint qualification)**.** *We say that the problem* (1) *satifies Abadie's constraint qualification if $T_S(\boldsymbol{x}) = G(\boldsymbol{x})$ for all $\boldsymbol{x} \in S$.*

Remark: Abadie's constraint qualification should be viewed as an abstract condition expressing "(1) is well-behaved"

Theorem 1. *Assume that the problem* (1) *satisfies Abadie's CQ, then at any locally optimal point $\boldsymbol{x}^*$ the system*

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(\boldsymbol{x}^*) = 0, \tag{2}$$

$$\mu_i g_i(\boldsymbol{x}^*) = 0, \quad i = 1, \dots, m, \tag{3}$$

$$\mu_i \geq 0, \quad i = 1, \dots, m. \tag{4}$$

has a solution $\boldsymbol{\mu}$.

Proof. See Theorem 5.29 in the book. □

Note that the KKT conditions are precisely the Fritz-John conditions, with the added requirement that $\mu_0 = 1$. We call the vector $\boldsymbol{\mu}$ solving the KKT system for some fixed $\boldsymbol{x} \in S$ a *Lagrange multiplier*.

2 Constraint qualifications

Our final task in showing that the KKT conditions are usable is to find practical conditions under which we can guarantee that Abadie's CQ holds. We start with one of the simplest and most useful ones.

Definition 2 (LICQ)**.** *We say that the linear independence constraint qualification (LICQ) holds at $\boldsymbol{x} \in S$ if the gradients $\nabla g_i(\boldsymbol{x}), i \in \mathcal{I}$ are linearly independent.*

Proposition 1. *The LICQ implies Abadie's CQ.*

Proof. See Proposition 5.40 and 5.46 in the book □

The other constraint qualifications we will consider in this course comes from comparing the cones defined previously. As a first example we consider the Mangasarian-Fromowitz CQ.

Definition 3 (MFCQ)**.** *The Mangasarian-Fromowitz constraint qualification (MFCQ) holds at $\boldsymbol{x} \in S$ if $\mathring{G}(\boldsymbol{x})$ is nonempty.*

Proposition 2. *The MFCQ implies the Abadie's CQ.*

Proof. See Proposition 5.40 in the book. □

The MFCQ can be used to get other constraint qualifications as well.

Definition 4 (Slater CQ)**.** *Slater CQ holds for* (1) *if g_i, $i = 1, \dots, m$ are convex functions, and there exists an interior point, i.e., a point $\boldsymbol{x}_0$ such that $g_i(\boldsymbol{x}_0) < 0$ for all $i \in \{1, \dots, m\}$.*

Proposition 3. *The Slater CQ implies Abadie's CQ.*

Proof. See Proposition 5.43 in the book. □

Definition 5 (Affine constraints CQ)**.** *The affine constraints holds for* (1) *if g_i, $i = 1, \dots, m$ are affine functions.*

Proposition 4. *The affine constraints CQ imply Abadie's CQ.*

Proof. If the constraints are affine then for any $\boldsymbol{p}$ and any $\boldsymbol{x} \in S$ and any $i \in \mathcal{I}(\boldsymbol{x})$ we have $g_i(\boldsymbol{x} + t\boldsymbol{p}) = g_i(\boldsymbol{x}) + t\nabla g_i(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{p} = t\nabla g_i(\boldsymbol{p})^{\mathrm{T}}\boldsymbol{p}$. So if $\boldsymbol{p} \in G(\boldsymbol{x})$ we have $g_i(\boldsymbol{x} + t\boldsymbol{p}) \leq 0$, and thus $\boldsymbol{x} + t\boldsymbol{p}$ is feasible for all small enough $t \geq 0$, i.e., $\boldsymbol{p} \in R_S(\boldsymbol{x})$. Hence $G(\boldsymbol{x}) = R_S(\boldsymbol{x})$ from which it follows that $T_S(\boldsymbol{x}) = G(\boldsymbol{x})$. □

3 Equality constraints

So far we have only talked about problems with inequality constraints, we briefly outline how to apply the above theory for the problem to

$$
\begin{aligned}
& \min f(\boldsymbol{x}), && (5a)\\
& \text{subject to } g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m, && (5b)\\
& h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, l. && (5c)
\end{aligned}
$$

where all the functions above are C^1. The main idea is to replace the equality constraints $h_j(\boldsymbol{x})=0$, with two inequality constraints, i.e., by $h_j(\boldsymbol{x}) \leq 0$ and $h_j(\boldsymbol{x}) \geq 0$, and apply the KKT theory to the problem

$$
\begin{aligned}
& \min f(\boldsymbol{x}), && (6a)\\
& \text{subject to } g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m, && (6b)\\
& h_j(\boldsymbol{x}) \leq 0, \quad j = 1, \ldots, l, && (6c)\\
& -h_j(\boldsymbol{x}) \leq 0, \quad j = 1, \ldots, l. && (6d)
\end{aligned}
$$

The main observation is that the equality constraints are *always* active in any feasible solution, and they will enter the KKT system with non-negative multipliers of opposite sign, which we can rewrite as a multiplier with any sign restrictions. The KKT system becomes

$$
\begin{aligned}
& f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(\boldsymbol{x}^*) = 0, && (7a)\\
& \mu_i g_i(\boldsymbol{x}^*) = 0, \quad i = 1, \ldots, m, && (7b)\\
& \mu_i \geq 0, \quad i = 1, \ldots, m. && (7c)
\end{aligned}
$$

The main difficulty is what happens to the constraint qualifications when we add equality constraints.

The main CQ was the MFCQ previously, i.e., $\overset{\circ}{G}(\boldsymbol{x}) \neq \emptyset$. But in the presence of equality constraints we need to introduce a cone

$$
H(\boldsymbol{x}) := \{\boldsymbol{p} \in \mathbb{R}^n \mid \nabla h_j(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{p} = 0,\ j = 1, \ldots, l\}
$$

and try to develop the theory of the previous sections when replacing the cones $\overset{\circ}{G}(\boldsymbol{x})$ and $G(\boldsymbol{x})$ by $\overset{\circ}{G}(\boldsymbol{x}) \cap H(\boldsymbol{x})$ and $G(\boldsymbol{x}) \cap H(\boldsymbol{x})$ respectively. We cannot immediately generalize the statement $\overset{\circ}{G}(\boldsymbol{x}) \subseteq T_S(\boldsymbol{x})$ to the statement $\overset{\circ}{G}(\boldsymbol{x}) \cap H(\boldsymbol{x}) \subseteq T_S(\boldsymbol{x})$. However, it turns out that we need to require the set of gradients $\{\nabla h_j(\boldsymbol{x})\}_{j=1}^{l}$ to be linearly independent. The constraint qualifications above then have to be modified to include the statement "and the set of gradients $\{\nabla h_j(\boldsymbol{x})\}_{j=1}^{l}$ is linearly independent"[1] . Refer to the book for detailed statements of all the CQs.

[1] Except the affine constraints CQ and Abadie's CQ.

4 Sufficiency of the KKT conditions under convexity

We have developed *neccessary* optimality conditions for the problem (1) so far. However, we want to know whether the KKT conditions are *sufficient* for optimality, that is, if the KKT system is solvable at $\boldsymbol{x}^*$ can we conclude that $\boldsymbol{x}^*$ is optimal in (1)? In the unconstrained case, we saw that the property that allows such statements is convexity, and it turns out that what we need in the constrained case is also convexity, but in terms of both the objective function and the constraints.

Theorem 2. *If, in* (1), *the objective function f and all constraint functions g_i, $i = 1, \dots, m$ are convex, then the KKT conditions are a sufficient optimality condition.*

Proof. See Theorem 5.49 in the book. □

Note that if we apply the above theorem to problems with equality constraints, then we must require that the equality constraints are affine.