TMA947 / MMG621 — Nonlinear optimization

Lecture 5 — Optimality conditions

Emil Gustavsson, Zuzana Nedělková

November 11, 2016

Consider a constrained optimization problem of the form

$$\min f(\boldsymbol{x}), \tag{1a}$$
$$\text{subject to } \boldsymbol{x} \in S, \tag{1b}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $S \subset \mathbb{R}^n$. We have already derived an optimality condition for the case where S is convex and $f \in C^1$, i.e.,

$$\boldsymbol{x}^* \text{is a local minimum} \Longrightarrow \boldsymbol{x}^* \text{is a stationary point}$$

The stationary point was defined in several different ways, one of the definitions was that if $\boldsymbol{x}^* \in S$ is a stationary point of f over S then

$$-\nabla f(\boldsymbol{x}^*) \in N_S(\boldsymbol{x}^*),$$

where $N_S(\boldsymbol{x}^*)$ is the normal cone of S at $\boldsymbol{x}^*$, i.e.,

$$N_S(\boldsymbol{x}^*) := \{\boldsymbol{p} \in \mathbb{R}^n \mid \boldsymbol{p}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}^*) \leq 0, \forall \boldsymbol{y} \in S\}.$$

The optimality condition $-\nabla f(\boldsymbol{x}^*) \in N_S(\boldsymbol{x}^*)$ says that it should not be possible to move from $\boldsymbol{x}^*$ in a direction allowed by S, such that f decreases.

This approach allows also to develop optimality conditions for more general non-linearly constrained problems. We first need to formalize the notion of a "direction allowed by S", and then require that these allowed directions do not contain any descent directions for f. Formulating a good notion of "allowed direction" is possibly the most challenging part of this course!

1 Geometric optimality conditions

First we introduce the most natural definition of allowed directions.

Definition 1 (cone of feasible directions). *Let $S \subset \mathbb{R}^n$ be a nonempty closed set. The cone of feasible directions for S at $\boldsymbol{x} \in S$ is defined as*

$$R_S(\boldsymbol{x}) := \{\boldsymbol{p} \in \mathbb{R}^n \mid \exists \delta > 0, \boldsymbol{x} + \alpha \boldsymbol{p} \in S, \forall 0 \leq \alpha \leq \delta\}. \tag{2}$$

Thus, $R_s(\boldsymbol{x})$ is nothing else but the cone containing all feasible directions at $\boldsymbol{x}$. A vector $\boldsymbol{p} \in R_s(\boldsymbol{x})$ if the feasible set S contains a non-trivial part of the half-line $\boldsymbol{x} + \alpha \boldsymbol{p}$, $\alpha \geq 0$. Unfortunately this cone is too small to develop optimality conditions for non-linearly constrained programs[1].

Example 1. *Let $S := \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_2 = x_1^2\}$. Then $R_S(\boldsymbol{x}) = \emptyset$ for all $\boldsymbol{x} \in S$, because the feasible set is a curved line in $\mathbb{R}^2$.*

We consider a significantly more complicated, but bigger and more well-behaving sets to develop optimality conditions.

Definition 2 (tangent cone)**.** *Let $S \subset \mathbb{R}^n$ be a nonempty closed set. The tangent cone for S at $\boldsymbol{x} \in S$ is defined as*

$$\begin{aligned} T_S(\boldsymbol{x}) := \big\{\boldsymbol{p} \mid \exists \{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset S, \{\lambda_k\}_{k=1}^{\infty} \subset (0, \infty),\ &\textit{such that} \\ \lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}&, \\ \lim_{k\to\infty} \lambda_k(\boldsymbol{x}_k - \boldsymbol{x}) = \boldsymbol{p}&\big\}. \end{aligned} \tag{3}$$

The above definition tells us that to check whether a vector $\boldsymbol{p} \in T_S(\boldsymbol{x})$ we should check whether there is a *feasible* sequence of points $\boldsymbol{x}_k \in S$ that approaches $\boldsymbol{x}$, such that $\boldsymbol{p}$ is the tangential to the sequence $\boldsymbol{x}_k$ at $\boldsymbol{x}$; such tangential vector is described as the limit of $\{\lambda_k(\boldsymbol{x}_k - \boldsymbol{x})\}$ for arbitrary positive sequence $\{\lambda_k\}$. Seen this way, $T_S(\boldsymbol{x})$ consists precisely of all the possible directions in which $\boldsymbol{x}$ can be asymptotically approached through S.

Example 2. *Let again $S := \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_2 = x_1^2\}$. Then, $T_S(\boldsymbol{0}) = \{\boldsymbol{p} \in \mathbb{R}^2 \mid p_2 = 0\}$.*

Example 3. *Let $S := \{\boldsymbol{x} \in \mathbb{R}^2 \mid -x_1 \leq 0; (x_1 - 1)^2 + x_2^2 \leq 1\}$. Then, $R_S(\boldsymbol{0}) = \{\boldsymbol{p} \in \mathbb{R}^2 \mid p_1 > 0\}$ and $T_S(\boldsymbol{0}) = \{\boldsymbol{p} \in \mathbb{R}^2 \mid p_1 \geq 0\}$.*

Example 4. *Suppose that we have a smooth curve in S starting at $\boldsymbol{x} \in S$, that is, we have a C^1 map $\gamma : [0, T] \to S$ for some $T > 0$. Then $\gamma'(0) \in T_S(\boldsymbol{x})$ since the definition of (one-sided) derivative is*

$$\gamma'(0) = \lim_{t \to 0} \frac{\gamma(t) - \gamma(0)}{t}. \tag{4}$$

So if we fix any sequence $t_k \to 0$, and let $\boldsymbol{x}_k := \gamma(t_k)$, $\lambda_k = 1/t_k$, we have defined the sequences required in the definition of $T_S(\boldsymbol{x})$.

It remains to formulate a notion of descent directions to the objective function f, fortunately we can use the same characerization as in the unconstrained case.

Definition 3 (cone of descent directions)**.** $\overset{\circ}{F}(\boldsymbol{x}) := \{\boldsymbol{p} \in \mathbb{R}^n \mid \nabla f(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{p} < 0\}$.

The above examples should then make the following theorem intuitively obvious.

[1] It will, however, work perfectly for *linear* programs!

Theorem 1 (geometric optimality conditions)**.** *Consider the problem* (1)*, where* $f \in C^1$*. Then*

$$\boldsymbol{x}^* \text{ is a local minimum of } f \text{ over } S \Longrightarrow \overset{\circ}{F}(\boldsymbol{x}^*) \cap T_S(\boldsymbol{x}^*) = \emptyset. \tag{5}$$

Proof. See theorem 5.10 in the book. □

Example 5. *If we return to our example with smooth curves, we showed that for any smooth curve* γ *though* S *starting at* $\boldsymbol{x}^*$*, we had* $\gamma'(0) \in T_S(\boldsymbol{x}^*)$*. The geometric optimality condition reduces to the statement that* $\frac{d}{dt}|_{t=0} f(\gamma(t)) \geq 0$ *when applied to this tangent vector.*

2 From geometric to useful optimality conditions

Now we have developed an elegant optimality condition, however there is no practical way to compute $T_S(\boldsymbol{x})$ directly from its definition. One way to overcome this difficulty (leading to the Fritz John conditions) is to replace the cone $T_S(\boldsymbol{x})$ by smaller cones.

Lemma 1. *If the cone* $C(\boldsymbol{x}) \subseteq T_S(\boldsymbol{x})$ *for all* $\boldsymbol{x} \in S$*, then* $\overset{\circ}{F}(\boldsymbol{x}^*) \cap C(\boldsymbol{x}^*) = \emptyset$ *is a neccessary optimality condition.*

Proof. Using the geometric optimality condition we have for any locally optimal $\boldsymbol{x}^* \in S$,

$$\overset{\circ}{F}(\boldsymbol{x}^*) \cap C(\boldsymbol{x}^*) \subseteq \overset{\circ}{F}(\boldsymbol{x}^*) \cap T_S(\boldsymbol{x}^*) = \emptyset.$$

□

By introducing smaller cones we get *weaker* optimality conditions than the geometric optimality conditions!

Example 6. *Let* $C(\boldsymbol{x}) = R_S(\boldsymbol{x})$ *and consider again the example* $S := \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_2 = x_1^2\}$*. Since* $R_S(\boldsymbol{x}) = \emptyset$*, the optimality condition* $\overset{\circ}{F}(\boldsymbol{x}) \cap R_S(\boldsymbol{x}) = \emptyset$ *holds for any feasible* $\boldsymbol{x} \in S$*, which is a totally useless optimality condition.*

The second way to overcome the difficulty with computing $T_S(\boldsymbol{x})$ is to introduce regularity conditions, or *constraint qualifications*, which will allow us to actually compute the tangent cone $T_S(\boldsymbol{x})$ by other means. This approach leads to the Karush-Kuhn-Tucker (KKT) conditions. The drawback of this approach is that, although the KKT conditions are equally strong as the geometric conditions, they are *less general*, i.e., they do not apply for irregular problems.

From now on we consider a problem of the form

$$\min f(\boldsymbol{x}), \tag{6a}$$
$$\text{subject to } g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \dots, m \tag{6b}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, and $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$ are all C^1, i.e., the feasible set $S := \{\boldsymbol{x} \in \mathbb{R}^n \mid g_i(\boldsymbol{x}) \leq 0, i = 1, \ldots, m\}$. This allows us to define additional cones related to $T_S(\boldsymbol{x})$. Let $\mathcal{I}(\boldsymbol{x})$ denote the *active set of constraints* at $\boldsymbol{x}$, that is,

$$\mathcal{I}(\boldsymbol{x}) := \{i \in \{1, \ldots, m\} \mid g_i(\boldsymbol{x}) = 0\}. \tag{7}$$

Definition 4 (gradient cones)**.** *We define the inner gradient cone* $\overset{\circ}{G}(\boldsymbol{x})$ *as*

$$\overset{\circ}{G}(\boldsymbol{x}) := \{\boldsymbol{p} \in \mathbb{R}^n \mid \nabla g_i(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{p} < 0, \forall i \in \mathcal{I}(\boldsymbol{x})\}, \tag{8}$$

and the gradient cone $G(\boldsymbol{x})$ *as*

$$G(\boldsymbol{x}) := \{\boldsymbol{p} \in \mathbb{R}^n \mid \nabla g_i(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{p} \leq 0, \forall i \in \mathcal{I}(\boldsymbol{x})\}. \tag{9}$$

Note that the inner gradient cone $\overset{\circ}{G}(\boldsymbol{x})$ consists of all vectors $\boldsymbol{p}$ that can be guaranteed to be descent directions of all defining functions for the active constraints, while the gradient cone $G(\boldsymbol{x})$ consists of all directions that can be guaranteed not to be ascent directions for the active constraints.

Theorem 2 (Relations between cones)**.** *For the problem* (6) *it holds that*

$$\operatorname{cl} \overset{\circ}{G}(\boldsymbol{x}) \subseteq \operatorname{cl} R_S(\boldsymbol{x}) \subseteq T_S(\boldsymbol{x}) \subseteq G(\boldsymbol{x}) \tag{10}$$

Proof. See Proposition 5.4 and Lemma 5.12 in the book. □

3 The Fritz John conditions

We obtain the Fritz John conditions when we replace the tangent cone $T_S(\boldsymbol{x})$ in the geometric optimality condition by $\overset{\circ}{G}(\boldsymbol{x})$.

$$\boldsymbol{x}^* \text{is locally optimal in (6)} \Longrightarrow \overset{\circ}{G}(\boldsymbol{x}) \cap \overset{\circ}{F}(\boldsymbol{x}) = \emptyset. \tag{11}$$

Therefore, the Fritz John conditions are *weaker* than the geometric optimality conditions.

This condition looks fairly abstract, however it is possible to reformulate it to a more practical condition. The above equation states for a fixed $\boldsymbol{x}$ that a linear system of inequalities does not have solution. Fortunately, we have Farkas' Lemma for turning an inconsistent set of linear inequalities into a consistent set of inequalities.

Theorem 3 (The Fritz John conditions). *If $\boldsymbol{x}^* \in S$ is a local minimum in* (6)*, then there exist multipliers $\mu_0 \in \mathbb{R}$, $\boldsymbol{\mu} \in \mathbb{R}^m$, such that*

$$\mu_0 \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(\boldsymbol{x}^*) = \mathbf{0}, \tag{12}$$

$$\mu_i g_i(\boldsymbol{x}^*) = 0, \quad i = 1, \ldots, m, \tag{13}$$

$$\mu_0, \mu_i \geq 0, \quad i = 1, \ldots, m, \tag{14}$$

$$(\mu_0, \boldsymbol{\mu}^{\mathrm{T}})^{\mathrm{T}} \neq \mathbf{0}. \tag{15}$$

Proof. See Theorem 5.17 in the book. □

The main drawback of the Fritz-John conditions is that they are too weak. The Fritz-John system contains a multiplier in front of the objective function term. If there is a solution to the Fritz-John system where the multiplier $\mu_0 = 0$, the objective function does not play any role in the system. This insight gives us at least one reason to think about regularity conditions (constraint qualifications); these conditions guarantee that any solution of the Fritz-John system must satisfy $\mu_0 \neq 0$.