Class 1: Preliminaries

The first part of the class consists in the proof of the so-called Karush-Kuhn-Tucker (KKT) conditions for a general nonlinear programming problem, with equality and inequality constraints. More specifically, we consider the following *nonlinear programming* (NLP) problem:

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

where

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

is the feasible set of problem (NLP). We assume that the functions $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{R}^m$ and $h\colon \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable. We also write $g := (g_1, \dots, g_m)$ with $g_i\colon \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$, and $h := (h_1, \dots, h_p)$ with $h_j\colon \mathbb{R}^n \to \mathbb{R}$, $j = 1, \dots, p$.

To prove KKT, we need the following ingredients: (a) a necessary optimality condition for (NLP) with normal cones characterization, (b) a property about cones and their polar, and (c) the Farkas' lemma. Before that, let us start with some simple notations and recall some basic definitions.

Notations

- ✓ The Euclidean inner product and norm are denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively.
- ✓ For any matrix $A \in \mathbb{R}^{m \times n}$, its transpose is denoted by $A^\top \in \mathbb{R}^{n \times m}$.
- ✓ A (column) vector $x \in \mathbb{R}^n$ with entries $x_i \in \mathbb{R}$, $i = 1, \dots, n$, is written as $(x_1, \dots, x_n)^\top$, or simply $(x_1, \dots, x_n)$. Note that we use subscripts for scalars.
- ✓ The sequence of vectors $x^0, x^1, x^2, \dots$ is denoted by $\{x^k\}_k$, or simply $\{x^k\}$. Note that we use superscripts here, to differentiate with the entries of a vector x.
- ✓ The gradient of a function $r\colon \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ is denoted by

  $$\nabla r(x) := \left(\frac{\partial r(x)}{\partial x_1}, \dots, \frac{\partial r(x)}{\partial x_n} \right)^\top,$$

  where $\partial r(x)/\partial x_i$ are the partial derivatives of r for $i = 1, \dots, n$.

Definition 1. *We say that a set S is* convex *if*

$$\alpha x + (1 - \alpha) y \in S \quad \textit{for all } \alpha \in [0, 1] \textit{ and } x, y \in S.$$

Definition 2. *We say that a set K is a* cone *if for all $x \in K$ and $\alpha \geq 0$, we have $\alpha x \in K$. Moreover, it is called* convex cone *when this cone is also convex.*

Exercise 1. *Prove that K is a convex cone if and only if*

$$\alpha x + \beta y \in K \quad \textit{for all } \alpha, \beta \geq 0 \textit{ and } x, y \in K.$$

By using Exercise 1, we can show that the set

$$\mathbb{S}^n_+ := \{A \in \mathbb{R}^{n \times n} \mid A \text{ is positive semidefinite}\}$$

is a convex cone. In fact, let $A, B \in \mathbb{S}^n_+$ and $\alpha, \beta \geq 0$. Then, for all v, we have

$$v^\top(\alpha A + \beta B)v = \alpha v^\top A v + \beta v^\top B v \geq 0$$

which means that $\alpha A + \beta B \in \mathbb{S}^n_+$.

Definition 3. *Let $x \in S$ be a vector with $S \subseteq \mathbb{R}^n$. The* tangent cone *of S at x is defined by*

$$T_S(x) := \left\{ y \in \mathbb{R}^n \;\middle|\; y = \lim_{k\to\infty} \alpha_k(x^k - x),\ \lim_{k\to\infty} x^k = x,\ x^k \in S \setminus \{x\},\ \alpha_k \geq 0,\ k = 0, 1, \ldots \right\}.$$

Moreover, each $y \in T_S(x)$ is called tangent vector *of S at x.*

The above tangent cone can be understood in the following sense. Consider a sequence $\{x^k\}$ in S that converges to x. In this situation, we can also think in a sequence of nonnegative scalars $\{\alpha_k\}$ and define $\{\alpha_k(x^k - x)\}$. If this last sequence converges to a point y, then y is the tangent vector of S at x. We can also imagine $T_S(x)$ as a "linear approximation" of the set S at the point x.

Proposition 1. *The tangent cone $T_S(x)$ is a closed set.*

Proof. Let $\{y^\ell\} \subseteq T_S(x)$ be a sequence such that $y^\ell \to y$. We need to show that $y \in T_S(x)$. We can assume without loss of generality that

$$\|y^\ell - y\| < \frac{1}{\ell}. \tag{2}$$

From Definition 3, since $y^\ell \in T_S(x)$, there exist $\{x^{\ell,k}\}_k$ and $\{\alpha_{\ell,k}\}_k$ such that

$$y^\ell = \lim_{k\to\infty} \alpha_{\ell,k}(x^{\ell,k} - x), \quad \lim_{k\to\infty} x^{\ell,k} = x, \quad x^{\ell,k} \in S \setminus \{x\}, \quad \alpha_{\ell,k} \geq 0$$

for all k. Then, there exists an index k_ℓ satisfying

$$\|\alpha_{\ell,k_\ell}(x^{\ell,k_\ell} - x) - y^\ell\| < \frac{1}{\ell} \quad \text{and} \quad \|x^{\ell,k_\ell} - x\| < \frac{1}{\ell}. \tag{3}$$

Now letting $\alpha_\ell := \alpha_{\ell,k_\ell}$ and $x^\ell := x^{\ell,k_\ell}$, we have $\alpha_\ell \geq 0$, $x^\ell \in S \setminus \{x\}$. Also, since

$$\|x^\ell - x\| = \|x^{\ell,k_\ell} - x\| < \frac{1}{\ell},$$

we obtain $x^\ell \to x$. From inequalities (2), (3), and the triangle inequality, we have

$$\begin{aligned}
\|\alpha_\ell(x^\ell - x) - y\| &= \|\alpha_{\ell,k_\ell}(x^{\ell,k_\ell} - x) - y\| \\
&\leq \|\alpha_{\ell,k_\ell}(x^{\ell,k_\ell} - x) - y^\ell\| + \|y^\ell - y\| \\
&\leq \frac{1}{\ell} + \frac{1}{\ell}.
\end{aligned}$$

Therefore, $y = \lim_{\ell\to\infty} \alpha_\ell(x^\ell - x)$ and we finally conclude that $y \in T_S(x)$. □

Definition 4. *Given a cone K, the set*

$$K^\circ := \left\{x \mid \langle x, y\rangle \le 0 \text{ for all } y \in K\right\}$$

is called the polar cone *of K.*

Recall that for given vectors x and y, we have $\langle x, y\rangle = \|x\|\,\|y\|\cos(\theta)$, where θ is the angle between x and y. Thus, K° is simply the set of elements that make obtuse angle with every element of K.

Exercise 2. *For any cone K, prove that the polar cone K° is convex.*

Definition 5. *Let $x \in S$ be a vector with $S \subseteq \mathbb{R}^n$. The* normal cone *of S at x is defined by*

$$N_S(x) := T_S(x)^\circ.$$

From Exercise 2, we observe that normal cones are always convex. For tangent cones, however, this is not necessarily true. For instance, if the set S is defined by

$$S = \left\{(x_1, x_2) \;\middle|\; \left((x_1+1)^2 - x_2\right)\left((x_1-1)^2 - x_2\right) = 0\right\},$$

then at $x = (0, 1)$, the cone $T_S(x)$ is not convex.

With the above definitions and remarks, we can now show the first ingredient that is necessary to prove KKT. Recalling the optimization problem (NLP), we prove a condition that is necessary for optimality, based on the normal cone of the feasible set (1).

Theorem 1. *Let $x^* \in \mathbb{R}^n$ be a local minimizer of problem (NLP). Then, we have*

$$-\nabla f(x^*) \in N_X(x^*). \tag{4}$$

Proof. Let $y \in T_X(x^*)$ be an arbitrary vector. From Definitions 4 and 5, we need to prove that $\langle -\nabla f(x^*), y\rangle \le 0$. From Definition 3, there exist sequences $\{x^k\}$ and $\{\alpha_k\}$ such that $x^k \in X \setminus \{x^*\}$ and $\alpha_k \ge 0$ for all k, with $x^k \to x^*$ and $\alpha_k(x^k - x^*) \to y$. Now, since f is differentiable (in particular at x^*), we have

$$f(x^k) = f(x^*) + \langle \nabla f(x^*), x^k - x^*\rangle + o(\|x^k - x^*\|),$$

where $o\colon [0,\infty) \to \mathbb{R}$ is a function such that $\lim_{t\to 0} o(t)/t = 0$. Moreover, because x^* is a local minimizer, $f(x^*) \le f(x^k)$ for large enough k. From the above equality, it means that

$$\langle \nabla f(x^*), x^k - x^*\rangle + o(\|x^k - x^*\|) \ge 0.$$

Using the fact that $\alpha_k \ge 0$ and $x^k \ne x$ for all k, we can also write

$$\langle \nabla f(x^*), \alpha_k(x^k - x^*)\rangle + \alpha_k\|x^k - x^*\|\frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|} \ge 0$$

with k large enough. Now, taking $k \to \infty$ in the above expression yields

$$\langle \nabla f(x^*), y\rangle + \|y\| \cdot 0 \ge 0,$$

and the claim holds. □

Now, assume that $X = \mathbb{R}^n$, which means that (NLP) is unconstrained. In this case, for any $x \in \mathbb{R}^n$, we have $T_X(x) = \mathbb{R}^n$ and $N_X(x) = \{0\}$. In particular, the normal cone of X at a local minimum x^* is given by $N_X(x^*) = \{0\}$. Thus, in this case, the condition (4) can be written as $\nabla f(x^*) = 0$, which is the classical first-order optimality condition for uncontrained problems. It is then natural to use the following definition.

Definition 6. *A point $x \in X$ is called stationary for* (NLP) *if $-\nabla f(x) \in N_X(x)$.*

Operations Research, Advanced (Graduate School of Informatics, Kyoto University)
1st part by Ellen H. Fukuda (e-mail: `ellen(at)i.kyoto-u.ac.jp`, where `(at)` = `@`)