TMA4180 Optimization: Tangent Cone and Constraint Qualifications

Elisabeth Köbis

NTNU, TMA4180, Optimering, Spring 2021

February 19th, 2021

Problem and Feasible Sequences

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{subject to } c_i(x) = 0, \quad i \in \mathcal{E}. \qquad \text{(P)}$$
$$c_i(x) \geq 0, \quad i \in \mathcal{I}.$$

Feasible set: $\Omega = \{x | c_i(x) = 0,\ i \in \mathcal{E},\ c_i(x) \geq 0,\ i \in \mathcal{I}\}$

Definition: Feasible Sequences

Given a feasible point x, we call $\{z_k\}$ a **feasible sequence approaching** x if $z_k \in \Omega$ for all k sufficiently large and $z_k \longrightarrow x$.

Recall that p was called a **feasible direction** at x if there exists $t > 0$ such that $x + tp \in \Omega$.

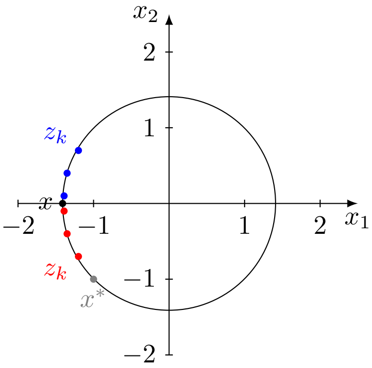
If z_k is a feasible sequence approaching x, then z_k is also a feasible direction: Define $p_k := z_k - x$. Then $z_k = x + p_k \in \Omega$, and so, z_k is a feasible direction at x.

Problem and Feasible Sequences

Example: A Single Equality Constraint

$$\min x_1 + x_2 \quad \text{s. t.} \quad x_1^2 + x_2^2 - 2 = 0.$$

$x^* = (-1, -1)^T$ is the global minimizer. Consider $x = (-\sqrt{2}, 0)^T$. Note that Ω is **not** convex.



Problem and Feasible Sequences

Example: A Single Equality Constraint

Then

$$z_k = \begin{pmatrix} -\sqrt{2 - 1/k^2} \\ -1/k \end{pmatrix}$$

is a feasible sequence, as for big enough k (here: for all k), it fulfills the constraints:

$$(-\sqrt{2 - 1/k^2})^2 + (-1/k)^2 - 2 = 2 - 1/k^2 + 1/k^2 - 2 = 0$$

Note that the objective function $f(x) = x_1 + x_2$ increases as we move along the sequence:

$$\begin{aligned} f(z_{k+1}) &= (-\sqrt{2 - 1/(k+1)^2})^2 + (-1/(k+1))^2 - 2 \\ &> (-\sqrt{2 - 1/k^2})^2 + (-1/k)^2 - 2 = f(z_k). \end{aligned}$$

Since we with increasing k, $f(z_k)$ increases and $f(z_k) < f(x)$ for all $k = 1, 2, 3, \ldots$ we see that x cannot be a (local) minimizer.

Problem and Feasible Sequences

Example: A Single Equality Constraint

Let us look at the sequence

$$z_k = \begin{pmatrix} -\sqrt{2 - 1/k^2} \\ 1/k \end{pmatrix}$$

which is also a feasible sequence, as for big enough k (here: for all k), it fulfills the constraints:

$$(-\sqrt{2 - 1/k^2})^2 + (1/k)^2 - 2 = 2 - 1/k^2 + 1/k^2 - 2 = 0$$

Note that the objective function $f(x) = x_1 + x_2$ decreases as we move along the sequence:

$$\begin{aligned} f(z_{k+1}) &= (-\sqrt{2 - 1/(k+1)^2})^2 + (-1/(k+1))^2 - 2 \\ &< (-\sqrt{2 - 1/k^2})^2 + (-1/k)^2 - 2 = f(z_k). \end{aligned}$$

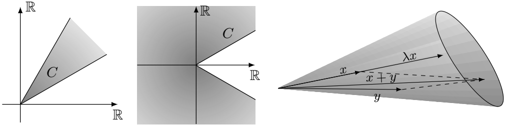
Note that here, $f(z_k) > f(x)$, which (falsely) might suggest that x is a local minimizer. But: A local minimizer is a point x at which **all** feasible sequences have the property that $f(z_k) \geq f(x)$ for all k sufficiently large.

Cones

Definition: Cone

A **cone** C is a set such that

$$c \in C \Longrightarrow \forall \lambda > 0: \ \lambda c \in C.$$



Tangents

Definition: Tangent / Tangent Vector / Tangent Cone

The vector d is said to be a **tangent** (or tangent vector) to Ω at a point x if there are a feasible sequence $\{z_k\}$ approaching x and a sequence of positive scalars $\{t_k\}$ with $t_k \longrightarrow 0$ such that

$$\lim_{k \to \infty} \frac{z_k - x}{t_k} = d.$$

The set of all tangents to Ω at x is called the **tangent cone** and is denoted by $T_\Omega(x)$.

Tangents

$T_\Omega(x)$ is indeed a cone: Let $d \in T_\Omega(x)$. Then λd ($\lambda > 0$) also belongs to $T_\Omega(x)$, because t_k can be replaced by t_k/λ:

$$\lim_{k\to\infty} \frac{z_k - x}{t_k/\lambda} = \lambda \lim_{k\to\infty} \frac{z_k - x}{t_k} = \lambda d.$$

So, to understand what vectors $T_\Omega(x)$ contains, it is enough to understand what unit vectors it contains. Then the rest of them are the multiples of those unit vectors. The reason to focus on the unit vectors in $T_\Omega(x)$ is that they are easier to find: we can take $t_k = ||z_k - x||$ in the definition. In other words, a unit vector u belongs to $T_\Omega(x)$ if and only if there is a sequence $z_k \to x$ such that

$$\frac{z_k - x}{||z_k - x||} \to u.$$

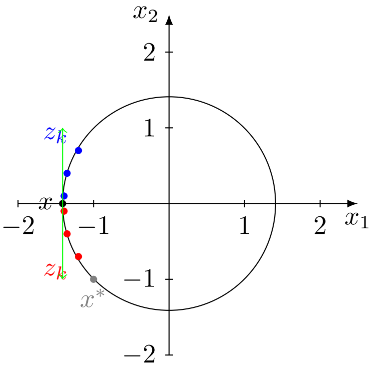
We can visualize this formula: You stand at x looking at approaching z_k, and mark the directions from which they hit you. The set of all these directions (normalized vectors $z_k - x$) gives all the unit vectors contained in the tangent cone.

Tangent Cone

Example: A Single Equality Constraint

$$\min x_1 + x_2 \quad \text{s. t.} \quad x_1^2 + x_2^2 - 2 = 0.$$

$x^* = (-1,-1)^T$ is the global minimizer. Consider $x = (-\sqrt{2}, 0)^T$.



$T_\Omega(x) = \{(0, d_2)^T | d_2 \in \mathbb{R}\}$.

Tangent Cone: Example I

Example: A Single Equality Constraint

$$\min x_1 + x_2 \quad \text{s. t.} \quad x_1^2 + x_2^2 - 2 = 0.$$

Consider $x = (-\sqrt{2}, 0)^T$. The tangent cone is given by all vectors of the form $(0, d_2)^T$ with $d_2 \in \mathbb{R}$: $T_\Omega(x) = \{(0, d_2)^T | d_2 \in \mathbb{R}\}$. To show this (in part), we consider the sequence $(z_k) := \begin{pmatrix} -\sqrt{2 - 1/k^2} \\ 1/k \end{pmatrix}$. Let $g_1(x) := -\sqrt{2 - x^2}$, $g_2(x) := x$. It holds $g_1'(x) = \frac{x}{\sqrt{2-x^2}}$, $g_2'(x) \equiv 1$. For the first component of $\frac{z_k - x}{t_k}$ with $t_k := \frac{1}{k}$, we obtain

$$\begin{aligned} \lim_{k\to\infty} \left(\frac{z_k - x}{t_k} \right)_1 &= \lim_{\frac{1}{k} =: h \to 0} \frac{g_1(h) - g_1(0)}{h} \\ &= g_1'(0) = 0\,. \end{aligned}$$

Tangent Cone: Example II

and for the second component, we get

$$\lim_{k\to\infty}\left(\frac{z_k - x_0}{t_k}\right)_2 = \lim_{k\to\infty}\frac{1/k - 0}{1/k} = 1\,,$$

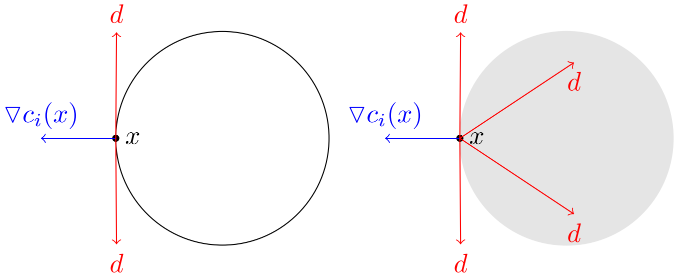
and therefore $d := (0,1)^T \in T_\Omega(x)$.

Linearized Feasible Directions

Definition: Set of Linearized Feasible Directions

Given a feasible point x and the active constraint set $\mathcal{A}(x)$, the set of linearized feasible directions $\mathcal{F}(x)$ is

$$\mathcal{F}(x) = \left\{ d | d^T \nabla c_i(x) = 0, \text{ for all } i \in \mathcal{E}, d^T \nabla c_i(x) \geq 0 \text{ for all } i \in \mathcal{A}(x) \cap \mathcal{I} \right\}$$



Left: Equality constraint $c_i(x) = 0$. Right: Inequality constraint $c_i(x) \geq 0$.

Linearized Feasible Directions

It is important to note that the linearized feasible direction set depends on the definition of the constraint functions c_i, $i \in \mathcal{E} \cup \mathcal{I}$. For example, consider the following constraint functions:

Example: A Single Equality Constraint

$$\min x_1 + x_2 \quad \text{s. t.} \quad c_1(x) := x_1^2 + x_2^2 - 2 = 0.$$

Consider $x = (-\sqrt{2}, 0)^T$. Then

$d^T \nabla c_1(x) = d^T (2x_1, 2x_2)|_{x=(-\sqrt{2},0)^T} = d^T(-2\sqrt{2}, 0) = -2\sqrt{2} d_1 \overset{!}{=} 0.$

Thus, $\mathcal{F}(x) = \{(0, d_2)^T | d_2 \in \mathbb{R}\}$.

If we define $c_1(x) := (x_1^2 + x_2^2 - 2)^2 = 0$, then Ω is unchanged, but:

$d^T \nabla c_1(x) = d^T (4x_1(x_1^2 + x_2^2 - 2), 4x_2(x_1^2 + x_2^2 - 2))|_{x=(-\sqrt{2},0)^T} = d^T(0, 0) = 0$. Thus, $\mathcal{F}(x) = \mathbb{R}^2$.

Linear Independence Constraint Qualification

Definition: Linear Independence Constraint Qualification (LICQ)

Given a feasible point x and the active constraint set $\mathcal{A}(x)$, we say that the **linear independence constraint qualification (LICQ)** holds if the set of active constraint gradients $\{\nabla c_i(x), i \in \mathcal{A}(x)\}$ is linearly independent.

Linear Independence Constraint Qualification

Example: A Single Equality Constraint

$$\min x_1 + x_2 \quad \text{s. t.} \quad x_1^2 + x_2^2 - 2 = 0.$$

Consider $x = (-\sqrt{2}, 0)^T$. Then
$\nabla c_1(x) = (2x_1, 2x_2)^T|_{x=(-\sqrt{2},0)^T} = (-2\sqrt{2}, 0)^T$. The set $\{(-2\sqrt{2}, 0)^T\}$ is linearly independent, and so, $T_\Omega(x) = \mathcal{F}(x)$ (see next slide).

Relation between Tangent Cone and Set of Linearized Feasible Directions

Lemma

Let x be a feasible point. The following two statements are true.

1. $T_\Omega(x) \subseteq \mathcal{F}(x)$.
2. If the LICQ condition is satisfied at x, then $T_\Omega(x) = \mathcal{F}(x)$.

Proof: See Nocedal & Wright, Lemma 12.2 (page 323).

First-Order Necessary Optimality Conditions

A local minimizer is a point x at which **all** feasible sequences have the property that $f(z_k) \geq f(x)$ for all k sufficiently large. We have the following result:

Theorem

If x^* is a local minimizer, then we have

$$\nabla f(x^*)^T d \geq 0 \ \forall d \in T_\Omega(x^*).$$

Proof: See Nocedal & Wright, Lemma 12.3 (page 325).

Farkas Lemma

Lemma: Farkas

Let $K := \{By + Cw | y \geq 0\}$. Given any vector $g \in \mathbb{R}^n$, we have either that $g \in K$ or that there exists $d \in \mathbb{R}^n$ satisfying

$$g^T d < 0, \; B^T d \geq 0, \; C^T d = 0.$$

but not both.

Proof: See Nocedal & Wright, Lemma 12.4 (page 327).

First-Order Necessary Optimality Conditions

Lagrangian function:

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x).$$

Theorem: First-Order Necessary Optimality Conditions

Suppose that x^* is a local minimizer of (P), that the functions f and c_i in are continuously differentiable, and that the LICQ holds at x^*. Then there is a Lagrange multiplier vector λ^* with components λ_i^*, $i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at (x^*, λ^*):

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0 \\
c_i(x^*) &= 0, \quad \text{for all } i \in \mathcal{E}, \\
c_i(x^*) &\geq 0, \quad \text{for all } i \in \mathcal{I}, \\
\lambda_i &\geq 0, \quad \text{for all } i \in \mathcal{I}, \\
\lambda_i^* c_i(x^*) &= 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}.
\end{aligned}$$

Proof

Let x^* be a local minimizer of (P). Then $\nabla f(x^*)^T d \geq 0 \ \forall d \in T_\Omega(x^*)$. As the LICQ condition is satisfied at x^*, then $T_\Omega(x^*) = \mathcal{F}(x^*)$. By applying Farkas Lemma to the cone N defined by

$$\sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*),\ \lambda_i \geq 0 \text{ for } i \in \mathcal{A}(x^*) \cap \mathcal{I},$$

and setting $g = \nabla f(x^*)$, we have that either

$$\nabla f(x^*) = \lambda_i \nabla c_i(x^*) = A(x^*)^T \lambda^*,\ \lambda_i \geq 0 \text{ for } i \in \mathcal{A}(x^*) \cap \mathcal{I}, \tag{1}$$

or else there is a direction d such that $d^T \nabla f(x^*) < 0$ and $d \in \mathcal{F}(x^*)$, which according to above is not true. Thus, there is a vector λ such that (1) holds. We now define the vector λ^* by

$$\lambda_i^* = \begin{cases} \lambda_i & (i \in \mathcal{A}(x)) \\ 0 & (i \in \mathcal{I} \setminus \mathcal{A}(x)) \end{cases}$$

It can be seen quite easily that with this vector λ^*, the conditions given in the theorem are satisfied.