TMA4180 Optimization: Optimization with convex constraints II

Elisabeth Köbis

NTNU, TMA4180, Optimering, Spring 2021

February 12th, 2021

Optimization with convex constraints

We consider *constrained* optimization problems of the form

$$\min_{x \in \Omega} f(x), \tag{P}$$

where $\Omega \subset \mathbb{R}^n$ is some convex set. Unless specified otherwise, we will always assume that:

- The function $f\colon \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^1$.
- The set $\Omega \subset \mathbb{R}^n$ is nonempty, convex, and closed.

Feasible Directions

In the case of unconstrained optimization, that is, $\Omega = \mathbb{R}^n$, we have seen that a necessary condition for a point $x^* \in \mathbb{R}^n$ to be a minimizer of f is that $\nabla f(x^*) = 0$. One way to interpret this equation is that all directional derivatives of f at x^* are equal to zero. In other words, if we perturb the point x^* a bit in any direction p, the function values do not decrease significantly. Or, we can say that a necessary condition for x^* to be a minimizer of f is that there exists no descent direction p of f at x^*. In the case of constrained optimization, the situation is notably different, because we do not need to consider every possible direction, but rather only those that—at least for sufficiently small step lengths—do not leave the set Ω we want to optimize over. These directions are called feasible:

Definition

Let $x \in \Omega$ and $p \in \mathbb{R}^n$. Then p is called a **feasible direction** at x if there exists $t > 0$ such that $x + tp \in \Omega$.

In other words, if we make a sufficiently small step in direction p starting at x, we still remain in the set Ω.

Feasible Directions

Remark

At this point, it is important to note that this definition of feasible directions is useful in the context of optimization only because the set Ω is assumed to be convex: The convexity of Ω implies that, given two points contained in Ω, the whole line segment connecting these points is itself completely contained in Ω. Thus, if p is a feasible direction at x and $t > 0$ is such that $x + tp$ is contained in Ω, then $x + \hat{t}p \in \Omega$ for all $0 < \hat{t} < t$ as well.

Moreover, we have the following characterisation of feasible directions:

Lemma 1

Let $x \in \Omega$. The direction $p \in \mathbb{R}^n$ is feasible if and only if $p = t(\hat{x} - x)$ for some $\hat{x} \in \Omega$ and $t > 0$.

Proof.

Let $p = t(\hat{x} - x)$ for some $\hat{x} \in \Omega$. Then $x + p/t = \hat{x} \in \Omega$, and therefore p is feasible. Conversely, let p be feasible-. Then $\hat{x} := x + tp \in \Omega$ for some $t > 0$, and we can write $p = (\hat{x} - x)/t$. $\square$

First Order Necessary Condition I

Proposition 1: First order necessary condition

Assume that x^* is a local solution of (P). Then

$$\nabla f(x^*)^T p \geq 0 \tag{1}$$

for all feasible directions p at x^*, or, equivalently,

$$\nabla f(x^*)^T (x - x^*) \geq 0 \tag{2}$$

for all $x \in \Omega$.

First Order Necessary Condition II

Proof.

Assume that p is a feasible direction. Then $x^* + tp \in \Omega$ for all sufficiently small $t > 0$. Since x^* is a local solution of (P), this implies that

$$f(x^*) \leq f(x^* + tp)$$

for all sufficiently small $t > 0$. Thus

$$\nabla f(x^*)^T p = \lim_{t \to 0^+} \frac{1}{t} \big(f(x^* + tp) - f(x^*) \big) \geq 0.$$

Due to Lemma 1, we can write any feasible direction as $p = t(\hat{x} - x^*)$ for some $t > 0$. Thus the two conditions (1) and (2) are equivalent. $\square$

Necessary and Sufficient Conditions for Convex Problems I

Proposition 2: Necessary and sufficient conditions for convex problems

Assume that f is convex. Then x^* is a global solution of (P) if and only if $x^* \in \Omega$ and

$$\nabla f(x^*)^T p \geq 0$$

for all feasible directions p at x^*, or, equivalently, that

$$\nabla f(x^*)^T (x - x^*) \geq 0 \tag{3}$$

for all $x \in \Omega$.

Necessary and Sufficient Conditions for Convex Problems II

Proof.

The necessity and equivalence of these conditions has already been shown in Proposition 1. It thus only remains to show that any one of them is sufficient. To that end, assume that (3) holds and let $x \in \Omega$. The convexity of f implies that

$$f(x) \geq f(x^*) + \nabla f(x^*)^T (x - x^*).$$

Since by assumption (3) is satisfied, the last term in this inequality is non-negative, and we obtain that $f(x) \geq f(x^*)$. Since this holds for every $x \in \Omega$, it follows that x^* is a global solution of (P). $\square$

Normal Cone I

Another possibility of formulating optimality conditions is based on the notion of the normal cone, which consists of all direction that form an obtuse angle with all feasible directions at a given point x:

Definition

Given $x \in \Omega$, we define the normal cone $N_\Omega(x)$ to Ω at x by

$$N_\Omega(x) = \left\{ q \in \mathbb{R}^n | q^T(\hat{x} - x) \leq 0 \text{ for all } \hat{x} \in \Omega \right\}.$$

Example 1

The normal cone of the single-valued set $\Omega = \{\hat{x}\}$ with $\hat{x} \in \mathbb{R}$ is $N_\Omega(\hat{x}) = \left\{ q \in \mathbb{R} | q\hat{x} \leq q\hat{x} \right\} = (-\infty, \infty)$.

Normal Cone II

Example 2

Let $\Omega = [0, 1]$. The normal cone to Ω is:

- $x = 0$: $N_\Omega(\hat{x}) = \left\{q \in \mathbb{R} | q\hat{x} \leq 0\right\} = (-\infty, 0]$
- $x = 1$: $N_\Omega(\hat{x}) = \left\{q \in \mathbb{R} | q\hat{x} \leq q\right\} = [0, +\infty]$
- $x \in (0, 1)$: $N_\Omega(x) = \left\{q \in \mathbb{R} | q(\hat{x} - x) \leq 0\right\} = \{0\}$

Sufficient Optimality Condition for Convex Problems based on Normal Cone I

Proposition 3

Assume that x^* is a local solution of (P). Then

$$-\nabla f(x^*) \in N_\Omega(x^*).$$

Conversely, if f is convex and $-\nabla f(x) \in N_\Omega(x)$, then x is a global solution of (P).

Proof.

This is an immediate consequence of the definition of the normal cone $N_\Omega(x)$ and Propositions 1 and 2. $\square$

Sufficient Optimality Condition for Convex Problems based on Tangent Cone I

Finally, it is possible to define the so called tangent cone to Ω at the point $x \in \Omega$ by

$$T_\Omega(x) := \{p \in \mathbb{R}^n | q^T p \le 0 \text{ for all } q \in N_\Omega(x)\}. \tag{4}$$

In the convex case, one can show that this tangent cone is the same tangent cone as defined in (N&W, Def. 12.2). Moreover, the latter can easily be seen to consist, only in the convex case, of all limits of feasible directions at x. The proof of these results relies on the notion of polar cones and some results from convex analysis and is quite a bit outside the scope.

Theorem

The tangent cone $T_\Omega(x)$ to the convex set Ω at the point $x \in \Omega$ is the closure of the set of all feasible directions at x.

In particular, this implies that the necessary optimality conditions as well as the sufficient optimality conditions for convex functions can be

Sufficient Optimality Condition for Convex Problems based on Tangent Cone II

formulated in terms of the tangent cone instead of the set of feasible directions. That is, if x^* is a local solution of (P), then (1) actually holds for all $p \in T_\Omega(x^*)$.

Remark

It is more common to start with defining the tangent cone as the closure of the cone of all feasible directions. Then one introduces the normal cone as

$$N_\Omega(x) = \{q \in \mathbb{R}^n | q^T p \leq 0 \text{ for all } p \in T_\Omega(x)\}.$$

Finally, one uses results from convex analysis in order to show that the tangent and normal cone defined in that manner also satisfy (4).

Projections I

Now we consider the special case where $f(x) = \frac{1}{2}\|x - z\|^2$, for some fixed $z \in \mathbb{R}^n$, that is, the problem

$$\min_{x \in \Omega} \frac{1}{2}\|x - z\|^2. \tag{5}$$

In other words, given $z \in \mathbb{R}^n$, we want to find the point $x^* \in \Omega$ for which the (squared Euclidean) distance to z is minimal.

Lemma

The problem (5) has a unique solution.

Proof.

The existence of a solution follows from the fact that the function $f(x) = \frac{1}{2}\|x - z\|^2$ is continuous and coercive, and the assumption that $\Omega \subset \mathbb{R}^n$ is non-empty and closed. The uniqueness of the solution follows from the strict convexity of f together with the convexity of Ω. $\square$

Projections II

Definition

Given $z \in \mathbb{R}^n$, we call the unique solution of (5) the *projection of z onto Ω* and denote it as $\pi_\Omega(z)$.

Proposition 4

The projection $\pi_\Omega(z)$ of z onto Ω is uniquely characterised by the conditions

$$\pi_\Omega(z) \in \Omega$$

and

$$\big(\pi_\Omega(z) - z\big)^T \big(x - \pi_\Omega(z)\big) \geq 0$$

for every $x \in \Omega$.

Projections III

Proof.

Denote $f(x) := \frac{1}{2}\|x - z\|^2$. Since f is convex, a necessary and sufficient condition for x^* to be a global solution of $\min_{x\in\Omega} f(x)$ is that $x^* \in \Omega$ and

$$\nabla f(x^*)^T(x - x^*) \geq 0$$

for all $x \in \Omega$ (see Proposition 2). Now $\nabla f(x^*) = x^* - z$, and thus the necessary and sufficient optimality condition reads as

$$(x^* - z)^T(x - x^*) \geq 0 \tag{6}$$

for every $x \in \Omega$. That is, $x^* = \pi_\Omega(z)$, if and only if $x^* \in \Omega$ and (6) holds, and the proof is complete. $\square$

Projections IV

We now return to the problem (P) of minimizing an arbitrary function f over a closed and convex set Ω and use the notion of a projection onto Ω to formulate yet another characterisation of the solutions.

Proposition 5

Assume that x^* is a solution of (P). Then

$$x^* = \pi_\Omega\big(x^* - \alpha \nabla f(x^*)\big) \tag{7}$$

for any $\alpha > 0$.

Projections V

Proof.

Since x^* solves (P), it follows that

$$\nabla f(x^*)^T(x - x^*) \geq 0$$

for every $x \in \Omega$. As a consequence,

$$\left(x^* - (x^* - \alpha\nabla f(x^*)^T)\right)^T(x - x^*) \geq 0$$

for every $x \in \Omega$. As shown in Proposition 4, however, this implies that x^* is the projection of $x^* - \alpha\nabla f(x^*)$ onto Ω, or
$x^* = \pi_\Omega(x^* - \alpha\nabla f(x^*))$. □

Remark

If f is convex, then it turns out that (7) is both a necessary and sufficient condition for a solution of (P). This readily follows from the fact that, in this case, the variational inequality (3) is a necessary and sufficient optimality condition according to Proposition 2.

Projections VI

Remark

Equation (7) implies that every local solution of (P) is a fixed point of the mapping $G\colon \mathbb{R}^n \to \mathbb{R}^n$,

$$G(x) = \pi_\Omega\big(x - \alpha \nabla f(x)\big).$$

As a consequence, it seems reasonable to try to solve (P) by means of the fixed point iteration (the *gradient projection method*)

$$\begin{aligned} z_{k+1} &\leftarrow x_k - \alpha \nabla f(x_k), \\ x_{k+1} &\leftarrow \pi_\Omega(z_{k+1}). \end{aligned}$$

Of course, this only makes sense if the projection on set Ω can be computed efficiently. In this case, however, this can be a viable, though possibly slow, algorithm for the solution of (P). Indeed, one can show that this algorithm converges if f is a strongly convex C^1-function, and the step length $\alpha > 0$ is chosen sufficiently small.