TMA4180 Optimization: Optimization with convex constraints III – Concave Inequality Constraints

Elisabeth Köbis

NTNU, TMA4180, Optimering, Spring 2021

February 15th, 2021

Optimization with Concave Inequality Constraints I

Finally, we will discuss the situation where the convex set Ω is the solution set of a number of inequalities. That is, we are given functions $c_i\colon \mathbb{R}^n \to \mathbb{R}$, $i \in \mathcal{I}$, for some (finite) index set $\mathcal{I}$ and define

$$\Omega = \{x \in \mathbb{R}^n | c_i(x) \geq 0,\ i \in \mathcal{I}\}.$$

The problem reads

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

Lemma

Assume that the functions $c_i\colon \mathbb{R}^n \to \mathbb{R}$ are concave. Then the set Ω is convex.

Optimization with Concave Inequality Constraints II

Proof.

Assume that x, $y \in \Omega$, and that $0 < \lambda < 1$. Then the concavity of the functions c_i implies that

$$c_i(\lambda x + (1-\lambda)y) \geq \lambda c_i(x) + (1-\lambda)c_i(y) \tag{1}$$

for all $i \in \mathcal{I}$. Now the assumption that x, $y \in \Omega$ implies that $c_i(x)$, $c_i(y) \geq 0$. Moreover, we have that λ, $1-\lambda \geq 0$. As a consequence, the right hand side in (1) is non-negative, which in turn shows that

$$c_i(\lambda x + (1-\lambda)y) \geq 0$$

for all $i \in \mathcal{I}$. This, however, shows that $\lambda x + (1-\lambda)y \in \Omega$, and thus Ω is convex. □

Slater's Constraint Qualification

In order to obtain reasonable optimality conditions for optimization problems with concave inequality constraints, we have to impose additional restrictions on the constraints that guarantee that the tangent and normal cones to Ω can be easily described by means of the gradients of the constraints. In the general context of constrained optimization, such conditions are called "constraint qualifications."

Definition

We say that *Slater's constraint qualification* is satisfied, if there exists $\hat{x} \in \mathbb{R}^n$ such that

$$c_i(\hat{x}) > 0 \qquad \text{for all } i \in \mathcal{I}.$$

Recall that the set of *active constraints* at $x \in \Omega$ is given by

$$\mathcal{A}(x) = \{i \in \mathcal{I} | c_i(x) = 0\}$$

(note that here, $\mathcal{E} = \emptyset$).

Optimization with Concave Inequality Constraints

Theorem

Assume that Slater's constraint qualification holds and that $x \in \Omega$. Then

$$T_\Omega(x) = \{p \in \mathbb{R}^n | p^T \nabla c_i(x) \geq 0 \text{ for all } i \in \mathcal{A}(x)\}. \tag{2}$$

Proof.

Let $p \in T_\Omega(x)$. Theorem[1](and the definition of feasible directions) implies that there exist convergent sequences $x_k \subset \Omega$ and $t_k > 0$ such that

$$p = \lim_{k \to \infty} t_k (x_k - x).$$

Moreover, due to the concavity of c_i[2], we note that for every $i \in \mathcal{A}(x)$ we have

$$0 \leq c_i(x_k) \leq c_i(x) + \nabla c_i(x)^T (x_k - x) = \nabla c_i(x)^T (x_k - x).$$

---

[1] 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.

[2] c_i is concave $\iff$ for any $x, x_k \in \Omega$ (Ω convex): $c_i(x_k) - c_i(x) \leq \nabla c_i(x)^T (x_k - x)$

Optimization with Concave Inequality Constraints I

Thus

$$p^T \nabla c_i(x) = \lim_{k \to \infty} t_k (x_k - x)^T \nabla c_i(x) \geq 0$$

for all $i \in \mathcal{A}(x)$. That is, every vector $p \in T_\Omega(x)$ has the form given in (2).

Auxiliary result: Now, let $p \in \mathbb{R}^n$ be such that $p^T \nabla c_i(x) > 0$ for all $i \in \mathcal{A}(x)$. Then there exists $t > 0$ such that for all $i \in \mathcal{I}$

$$c_i(x + tp) = c_i(x) + tp^T \nabla c_i(x) + o(t) > 0.$$

As a consequence, we can write such a vector p as $p = (\tilde{x} - x)/t$ with $\tilde{x} = x + tp \in \Omega$. This shows that all vectors $p \in \mathbb{R}^n$ with $p^T \nabla c_i(x) > 0$ for all $i \in \mathcal{A}(x)$ are feasible directions at x.

Finally, let $p \in \mathbb{R}^n$ be such that $p^T \nabla c_i(x) \geq 0$ for all $i \in \mathcal{A}(x)$.

Optimization with Concave Inequality Constraints I

Let moreover $\hat{x} \in \Omega$ be such that $c_i(\hat{x}) > 0$ for all $i \in \mathcal{I}$ (such a point exists because of Slater's constraint qualification) and define

$$p_k := p + \frac{1}{k}(\hat{x} - x).$$

Then

$$\nabla c_i(x)^T p_k = \nabla c_i(x)^T p + \frac{1}{k}\nabla c_i(x)^T(\hat{x} - x) \geq \frac{1}{k}\nabla c_i(x)^T(\hat{x} - x). \quad (3)$$

However, because of the concavity of c_i we have that

$$0 < c_i(\hat{x}) \leq \underbrace{x_i(x)}_{=0,\ \text{as } i \in \mathcal{A}(x)} + \nabla c_i(x)^T(\hat{x} - x) = \nabla c_i(x)^T(\hat{x} - x)$$

for all $i \in \mathcal{A}(x)$.

Together with (3), this shows that

$$\nabla c_i(x)^T p_k > 0$$

for all $i \in \mathcal{A}(x)$, which in turn, by the auxiliary result, shows that all the vectors p_k are feasible directions at x. As a consequence, p is the limit of a sequence of feasible directions at x. Using Theorem[3], we obtain that $p \in T_\Omega(x)$. □

[3] 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.

Farkas' Lemma

Theorem: Farkas' Lemma

Let s_j, $j \in \mathcal{J}$ finite, be vectors in $\mathbb{R}^n$, and let $g \in \mathbb{R}^n$. Then exactly one of the following statements is true:

1. There exist $\lambda_j \geq 0$, $j \in \mathcal{J}$, such that
$$\sum_{j \in \mathcal{J}} \lambda_j s_j = g.$$
2. There exists $p \in \mathbb{R}^n$ such that $g^T p < 0$ and $s_j^T p \geq 0$ for all $j \in \mathcal{J}$.

Proof.

See e.g.: Lemma 12.4 in Nocedal&Wright or Sec. III.4.3 in *J.-B. Hiriart-Urruty and C. Lemaréchal: Convex analysis and minimization algorithms. I, volume 305 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1993.* □

Optimization with Concave Inequality Constraints: Optimality Condition

Theorem

Assume that Slater's constraint qualification holds and that x^* is a local minimizer of the problem (P). Then there exists a *Lagrange multiplier* $\lambda^* \in \mathbb{R}^{|\mathcal{I}|}$ such that

$$\begin{aligned} \nabla f(x^*) &= \sum_{i \in \mathcal{I}} \lambda_i^* \nabla c_i(x^*), & \\ \lambda_i &\geq 0, & i \in \mathcal{I}, \\ \lambda_i &= 0, & i \notin \mathcal{A}(x^*). \end{aligned} \tag{4}$$

Conversely, if additionally f is convex and (4) holds, then x^* is a global solution of (P).

Optimization with Concave Inequality Constraints: Optimality Condition

Proof.

Since x^* is a local minimizer of the problem (P), it follows that $\nabla f(x^*)^T p \geq 0$ for all $p \in T_\Omega(x^*)$. (2) in the previous theorem implies that this is equivalent to stating that $\nabla f(x^*)^T p \geq 0$ for all $p \in \mathbb{R}^n$ with $\nabla c_i(x^*)^T p \geq 0$, $i \in \mathcal{A}(x^*)$. In other words, there does not exist a vector $p \in \mathbb{R}^n$ with $\nabla f(x^*)^T p < 0$ and $\nabla c_i(x^*)^T p \geq 0$ for all $i \in \mathcal{A}(x^*)$. Thus, Farkas' Lemma implies that we can write

$$\nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*)$$

for some $\lambda_i^* \geq 0$, $i \in \mathcal{A}(x^*)$. Setting $\lambda_i^* = 0$ for $i \in \mathcal{I} \setminus \mathcal{A}(x^*)$, we obtain the desired representation of $\nabla f(x^*)$.

The converse direction follows from the fact that the condition $\nabla f(x^*)^T p \geq 0$ for all $p \in T_\Omega(x^*)$ is a sufficient optimality condition in the convex case. □

Optimization with Concave Inequality Constraints: Optimality Condition

Remark

One can generalize these results to the case where Ω is given by concave inequality constraints and linear equality constraints:

$$\Omega = \{x \in \mathbb{R}^n | c_i(x) \geq 0, i \in \mathcal{I}, \text{ and } Ax = b\},$$

with $c_i \colon \mathbb{R}^n \to \mathbb{R}$ concave, and $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. In such a case, Slater's constraint qualification reads: There exists $\hat{x} \in \mathbb{R}^n$ with $c_i(\hat{x}) > 0$, $i \in \mathcal{I}$, and $A\hat{x} = b$. If this condition is satisfied, one can show (by essentially following the same argumentation as above) that a necessary optimality condition is the existence of $\lambda^* \in \mathbb{R}^{\mathcal{I}}$ and $\mu^* \in \mathbb{R}^m$ such that

$$\nabla f(x^*) = A^T \mu^* + \sum_{i \in \mathcal{I}} \lambda_i^* \nabla c_i(x^*)$$

with $\lambda_i^* \geq 0$ for all $i \in \mathcal{I}$, and $\lambda_i^* = 0$ for all $i \notin \mathcal{A}(x^*)$.