OPTIMIZATION A
Week 2: Thursday, November 7, 2024

Elena del Mercato

Outline

- Inequality constraints
- Karush-Kuhn-Tucker (KKT) conditions
- KKT necessary conditions
- KKT sufficient conditions

Inequality constraints

Let U be an **open** subset of $\mathbb{R}^n$. The functions f and $h_1, \ldots, h_j, \ldots, h_m$ are defined on U.

We study the **maximization** problem $(\mathcal{I})$ with the following **inequality constraints** (i.e., ≤ 0).

$$(\mathcal{I}) \begin{cases} \max_{x \in U} f(x) \\ h_j(x) \leq 0, \, j = 1, \ldots, m \end{cases}$$

The adaptation of the following study to minimization problems of a function g or optimization problems with inequality constraints described by the inequality $g_j(x) \geq 0$ is left to the reader, by remarking that :

1. $\min g(x) = \max f(x)$, with $f(x) = -g(x)$.
2. $g_j(x) \geq 0$ if and only if $h_j(x) \leq 0$, with $h_j(x) = -g_j(x)$.

Binding constraints

Definition

Let $x^* \in U$, we say that the constraint j is **binding** at x^* if $h_j(x^*) = 0$. We denote :

1. $J(x^*)$ the set of all binding constraints at x^*, that is :

$$J(x^*) := \{j = 1, ..., m : h_j(x^*) = 0\},$$

2. $m^* \leq m$ the number of elements of $J(x^*)$, and
3. $h^* := (h_j)_{j \in J(x^*)}$ the following mapping :

$$h^* : x \in U \subseteq \mathbb{R}^n \longrightarrow h^*(x) = (h_j(x))_{j \in J(x^*)} \in \mathbb{R}^{m^*}$$

Karush-Kuhn-Tucker (KKT) conditions

From now on, f and $h_1, \ldots, h_j, \ldots, h_m$ are $\mathcal{C}^1$ on U.

KKT conditions associated with the maximization problem $(\mathcal{I})$:

$$(KKT) \begin{cases} \nabla f(x) = \sum_{j=1}^{m} \mu_j \nabla h_j(x), \\ \forall j = 1, ..., m,\ \mu_j \in \mathbb{R}_+ \text{ and } h_j(x) \leq 0, \\ \forall j = 1, ..., m,\ \mu_j h_j(x) = 0 \text{ (complementary slackness).} \end{cases}$$

That is, at x :
1) The gradient of the objective function is a linear combination of the gradients of the constraint functions, with positive coefficients $\mu_j \geq 0$.
2) All the constraints are satisfied.
3) If $\mu_j > 0$, then the constraint j is **binding** at x. If x belongs to the **interior** of constraint j, i.e., $h_j(x) < 0$, then $\mu_j = 0$.

Linearized problem

Let $x^* \in U$ be a solution of problem $(\mathcal{I})$.

The main idea to prove that KKT conditions are necessary to solve problem $(\mathcal{I})$ is to replace problem $(\mathcal{I})$ with the ***linearized*** problem $(\mathcal{L}^*)$:

$$(\mathcal{L}^*) \begin{cases} \max\limits_{x \in \mathbb{R}^n} \nabla f(x^*) \cdot (x - x^*) \\ \nabla h_j(x^*) \cdot (x - x^*) \leq 0, \, j \in J(x^*) \end{cases}$$

Notice that, in problem $(\mathcal{L}^*)$, what really matters is the use of the **binding constraints** at x^*.

Generalized constraint qualification condition

Definition

Let $x^* \in U$ be a solution of problem $(\mathcal{I})$ such that $h_j(x^*) = 0$ for all $j \in J(x^*)$. The generalized constraint qualification (GCQ) condition is satisfied at x^* if x^* is **also** a solution of problem $(\mathcal{L}^*)$.

Remark that condition GCQ is not always satisfied.

One can easily find examples where $x^* \in \mathcal{S}ol(\mathcal{I})$, but $x^* \notin \mathcal{S}ol(\mathcal{L}^*)$.

Existence of Lagrange multipliers

Theorem

Assume that f and $h_1, \ldots, h_j, \ldots, h_m$ are $\mathcal{C}^1$ on U.

If $x^* \in U$ is a solution of problem $(\mathcal{I})$ and x^* satisfies condition GCQ, then there exists $\mu^* = (\mu_1^*, ..., \mu_j^*, ..., \mu_m^*) \in \mathbb{R}_+^m$ such that the vector $(x^*, \mu^*) \in U \times \mathbb{R}_+^m$ satisfies the KKT conditions associated with problem $(\mathcal{I})$.

Sketch of the proof

If $\nabla f(x^*) = 0$, take $\mu_j^* = 0$ for all $j = 1, ..., m$.

Assume now that $\nabla f(x^*) \neq 0$. Since x^* solves problem $(\mathcal{L}^*)$, there is no $x \neq x^*$ such that :

$$\nabla f(x^*) \cdot (x - x^*) > 0 = \nabla f(x^*) \cdot (x^* - x^*),$$

and

$$\nabla h_j(x^*) \cdot (x - x^*) \leq 0, \ \forall j \in J(x^*).$$

Take $b = \nabla f(x^*)$, and $a^j = \nabla h_j(x^*)$ for all $j \in J(x^*)$.

By **Farkas' Lemma**, there exists $\mu^* = (\mu_j^*)_{j \in J(x^*)} \in \mathbb{R}_+^{m^*}$ such that :

$$b = \sum_{j \in J(x^*)} \mu_j^* a^j.$$

For all $j \notin J(x^*)$, take $\mu_j^* = 0$.

By construction, we get $\mu_j^* h_j(x^*) = 0$ for all $j = 1, ..., m$, and $\nabla f(x^*) = \sum_{j=1}^{m} \mu_j^* \nabla h_j(x^*)$.

Further, $h_j(x^*) \leq 0$ for all $j = 1, ..., m$, because x^* is a solution of problem $(\mathcal{I})$.

Hence, (x^*, μ^*) satisfies the KKT conditions associated with problem $(\mathcal{I})$. ■

Sufficient conditions for generalized constraint qualification

Theorem

Assume that f and $h_1, \ldots, h_j, \ldots, h_m$ are $\mathcal{C}^1$ on U.

1. *If h_j **is linear or affine** for all $j = 1, ..., m$, then condition GCQ is satisfied.*
2. **(Slater's condition)** *Assume that U is also convex and :*
   - *the constraint functions h_j is **convex** for all $j = 1, ..., m$,*
   - *there exists $\widetilde{x} \in U$ such that $h_j(\widetilde{x}) < 0$ for all $j = 1, ..., m$.*

   Then, condition GCQ is satisfied.
3. **(Rank condition)** *If all the gradients $(\nabla h_j(x^*))_{j \in J(x^*)}$ are **linearly independent**, i.e., the rank of the Jacobian matrix $Dh^*(x^*)$ is equal to m^* (full row rank), then condition GCQ is satisfied.*

Two remarks

Remark 1 In the Rank condition, one easily recognizes the classical constraint qualification condition given for optimization problems with **equality constraints**.

Remark 2 In Slater's condition, the convexity of h_j can be weakened by another assumption, that is, h_j is "pseudo-convex".

It is well know that :

1. A $\mathcal{C}^1$ convex function is pseudo-convex.
2. A $\mathcal{C}^1$ quasi-convex function with gradient different from zero everywhere is pseudo-convex. Hence, in Slater's condition, the convexity of h_j can be replaced with the following assumption :
   - h_j is quasi-convex with $\nabla h_j(x) \neq 0$ for all $x \in U$.

KKT necessary conditions

As a consequence of the previous two theorems one gets the following theorem.

Theorem

Assume that f and $h_1, \dots, h_j, \dots, h_m$ are $\mathcal{C}^1$ on U.
Let $x^* \in U$ be a solution of problem $(\mathcal{I})$.
*Assume that **one** of the following three conditions is satisfied.*

1. *If h_j **is linear or affine** for all $j = 1, ..., m$.*
2. **Slater's condition.**
3. **Rank condition.**

Then, there exists $\mu^* = (\mu_1^*, ..., \mu_j^*, ..., \mu_m^*) \in \mathbb{R}_+^m$ such that $(x^*, \mu^*) \in U \times \mathbb{R}_+^m$ satisfies the KKT conditions associated with problem $(\mathcal{I})$.

KKT sufficient conditions

Let U be an **open and convex** subset of $\mathbb{R}^n$.

Theorem

Assume that f and $h_1, \ldots, h_j, \ldots, h_m$ are $\mathcal{C}^1$ on U.

If there exists $\mu^* = (\mu_1^*, ..., \mu_j^*, ..., \mu_m^*) \in \mathbb{R}_+^m$ such that $(x^*, \mu^*) \in U \times \mathbb{R}_+^m$ satisfies the KKT conditions associated with problem $(\mathcal{I})$, and the following condition (C) holds true, then x^* is a solution of problem $(\mathcal{I})$.

Condition (C) : *The function $\mathcal{L}(x) = f(x) - \sum_{j=1}^{m} \mu_j^* h_j(x)$ is concave in x.*

Let U be an **open and convex** subset of $\mathbb{R}^n$. Assume that f and $h_1, \ldots, h_j, \ldots, h_m$ are $\mathcal{C}^1$ on U.

Proposition

The previous theorem still holds true if ***Condition (C)*** *is replaced by* ***one*** *of the following two conditions.*

1. *The objective function f is* ***concave*** *and the constraint functions h_j are* ***quasi-convex*** *for all $j = 1, ..., m$.*
2. *The objective function f is* ***quasi-concave*** *with $\nabla f(x) \neq 0$ for all $x \in U$, and the constraint functions h_j are* ***quasi-convex*** *for all $j = 1, ..., m$.*

Hence, in order to check if KKT conditions are sufficient to solve problem $(\mathcal{I})$, we have to verify also some properties of the **objective function** f.

Sketch of the proof

Without loss of generality, f is pseudo-concave on U.

Assume that there exists $\mu^* = (\mu_1^*, ..., \mu_j^*, ..., \mu_m^*) \in \mathbb{R}_+^m$ such that $(x^*, \mu^*) \in U \times \mathbb{R}_+^m$ satisfies the KKT conditions associated with problem $(\mathcal{I})$.

If $\nabla f(x^*) = 0$, then $f(x) \leq f(x^*)$ for all $x \in U$ (because U is **open** and f is pseudo-concave on U). Hence, $f(x) \leq f(x^*)$ for all $x \in U$ such that $h_j(x) \leq 0$ for all $j = 1, ..., m$. Further, $h_j(x^*) \leq 0$ for all $j = 1, ..., m$. Then, x^* solves problem $(\mathcal{I})$.

Assume now that $\nabla f(x^*) \neq 0$.

By contradiction, if x^* is not a solution of problem $(\mathcal{I})$, then there is $x \in U$, $x \neq x^*$, such that $h_j(x) \leq 0$ for all $j = 1, ..., m$, and $f(x) > f(x^*)$. By pseudo-concavity of f, one gets :

$$\nabla f(x^*) \cdot (x - x^*) > 0.$$

Since h_j is quasi-convex and $h_j(x) \leq 0 = h_j(x^*)$ for all $j \in J(x^*)$, we have that $\nabla h_j(x^*) \cdot (x - x^*) \leq 0$ for all $j \in J(x^*)$. Then, we get $\mu_j^* \nabla h_j(x^*) \cdot (x - x^*) \leq 0$ for all $j \in J(x^*)$, because $\mu_j^* \geq 0$.

If $j \notin J(x^*)$, then $\mu_j^* = 0$, because of complementary slackness.

Hence, we get :

$$\mu_j^* \nabla h_j(x^*) \cdot (x - x^*) \leq 0, \ \forall j = 1, ..., m.$$

Summing over $j = 1, ..., m$, we have :

$$\sum_{j=1}^{m} \mu_j^* \nabla h_j(x^*) \cdot (x - x^*) < \nabla f(x^*) \cdot (x - x^*).$$

That is impossible, because $\sum_{j=1}^{m} \mu_j^* \nabla h_j(x^*) = \nabla f(x^*)$. We then conclude that x^* must be a solution of problem $(\mathcal{I})$. ■

Farkas' Lemma is a consequence of one of the Separation Theorems, and it is often used in mathematical programming.

Let $A = \{a^1, ..., a^j, ..., a^m\}$ be a set of m points of $\mathbb{R}^n$.

$K(A)$ denotes the set of all linear combinations of elements of A with positive coefficients :

$$K(A) = \left\{ z = \sum_{j=1}^{m} \mu_j a^j \in \mathbb{R}^n : \mu_j \geq 0 \text{ and } a^j \in A, \ \forall j = 1, ..., m \right\}.$$

That is, $K(A)$ is the smallest (in the sense of inclusion) **convex cone** of vertex 0 generated by $a^1, ..., a^j, ..., a^m$.

Theorem (Farkas' Lemma)

Let $A = \{a^1, ..., a^j, ..., a^m\}$ be a set of m points of $\mathbb{R}^n$. Consider any point $b \in \mathbb{R}^n$.

*Then, **only one** of the following two alternatives holds true.*

1. *There exists $\mu = (\mu_1, ..., \mu_j, ..., \mu_m) \in \mathbb{R}^m_+$ such that :*

$$b = \sum_{j=1}^{m} \mu_j a^j.$$

2. *There exists $p \in \mathbb{R}^n$ with $p \neq 0$ such that :*

$$p \cdot b > 0 \ \text{ and } \ p \cdot a^j \leq 0, \ \forall\, j = 1, ..., m.$$

Let U be an **open and convex** subset of $\mathbb{R}^n$, f is a $\mathcal{C}^1$ on C.

Definition (Pseudo-concavity)

f is pseudo-concave on U if for all x and $\overline{x}$ in U with $x \neq \overline{x}$,

$$f(x) > f(\overline{x}) \Longrightarrow \nabla f(\overline{x}) \cdot (x - \overline{x}) > 0$$

A function g is pseudo-convex on U if and only if the function $f = -g$ is pseudo-concave on U.

Proposition

1. *If f is concave on U, then f is pseudo-concave on U.*
2. *If f is quasi-concave on U and $\nabla f(x) \neq 0$ for all $x \in U$, then f is pseudo-concave on U.*