Optimization

Màster de Fonaments de Ciència de Dades

Gerard Gómez

Lecture IX. Convexity

IX.1 Convex sets

IX.2 Convex functions. Continuity and differentiability

IX.3 Extrema of convex functions

IX.4 Optimality conditions for convex problems

IX.1 Convex sets

Definitions:

- Given m points $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m \in \mathbb{R}^n$, we define a **convex combination** of these points as
$$\boldsymbol{u} = \sum_{i=1}^{m} \alpha_i \boldsymbol{u}_i,$$
with $\alpha_i \geq 0$ and $\sum_{i=1}^{m} \alpha_i = 1$.

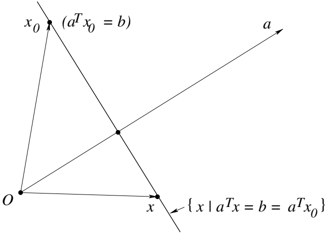
- A subset $C \subset \mathbb{R}^n$ is a **convex set** if and only if for any two points $\boldsymbol{u}_1, \boldsymbol{u}_2 \in C$, any convex conbination of these points verifies
$$\alpha \boldsymbol{u}_1 + (1-\alpha)\boldsymbol{u}_2 \in C.$$

Examples:

1. If $\boldsymbol{u}_1 = (1,0)^T$, $\boldsymbol{u}_2 = (0,1)^T$ then, the convex conbinations of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are all the points in the segment joining $(1,0)$ and $(0,1)$.
2. The empty set $\emptyset$, the set containing a single point $\boldsymbol{x} \in \mathbb{R}^n$, and $\mathbb{R}^n$.
3. The n-dimensional sphere with center at $\boldsymbol{x}_0 \in \mathbb{R}^n$ and radius α
$$S_\alpha(\boldsymbol{x}_0) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{x}_0\| \leq \alpha\}.$$

IX.1 Convex sets. Hyperplanes

- Let $\boldsymbol{a} \in \mathbb{R}^n$, $\boldsymbol{a} \neq 0$ and b a given number, the set
$$\left\{ \boldsymbol{x} \,|\boldsymbol{a}^T \boldsymbol{x} = b \right\},$$
is called a **hyperplane** of $\mathbb{R}^n$.
- A hyperplane is formed by all the vectors $\boldsymbol{x} \in \mathbb{R}^n$ such that its scalar product with $\boldsymbol{a}$ is constant.
- If $\boldsymbol{x}_0$ i $\boldsymbol{x}_1$ are two points of the hyperplane, then
$$\boldsymbol{a}^T (\boldsymbol{x}_1 - \boldsymbol{x}_0) = 0,$$
by this reason the vector $\boldsymbol{a}$ is called the **normal vector of the hyperplane**.



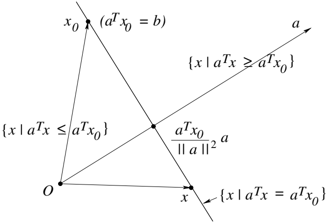
IX.1 Hyperplanes. Properties

1. If $\boldsymbol{x} = \lambda \boldsymbol{a}$ is the point of the hyperplane **closest to the origin**, then

$$\boldsymbol{a}^T(\lambda \boldsymbol{a}) = \lambda \boldsymbol{a}^T \boldsymbol{a} = b, \quad \lambda = \frac{b}{\|\boldsymbol{a}\|_2} = \frac{\boldsymbol{a}^T \boldsymbol{x}_0}{\|\boldsymbol{a}\|_2},$$

so

$$\boldsymbol{x} = \frac{\boldsymbol{a}^T \boldsymbol{x}_0}{\|\boldsymbol{a}\|_2} \boldsymbol{a}.$$



2. **Hyperplanes are convex** sets.
3. A hyperplane in $\mathbb{R}^n$ defines **four convex sets**: two closed half-spaces and two open half-spaces:

$$\left\{\boldsymbol{x} | \boldsymbol{a}^T \boldsymbol{x} \geq b\right\}, \quad \left\{\boldsymbol{x} | \boldsymbol{a}^T \boldsymbol{x} \leq b\right\}, \quad \left\{\boldsymbol{x} | \boldsymbol{a}^T \boldsymbol{x} > b\right\}, \quad \left\{\boldsymbol{x} | \boldsymbol{a}^T \boldsymbol{x} < b\right\}.$$

IX.1 Convex sets

Theorem

The set $C \subset \mathbb{R}^n$, is convex if and only if every convex combination of any finite number of points of C is contained in C.

Proof: Supose that C is a convex set. The proof is **by induction** on the number of points $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_s \in C$. **For** $s = 1$, the theorem **is clearly true**. **Assume that it is true for a certain** $s > 1$, and let us prove it for $s + 1$. Let

$$\boldsymbol{u} = \alpha_1 \boldsymbol{u}_1 + \ldots + \alpha_s \boldsymbol{u}_s + \alpha_{s+1} \boldsymbol{u}_{s+1}, \quad \text{with} \quad \sum_{i=1}^{s+1} \alpha_i = 1.$$

Without loss of generality, we can assume that $\alpha_{s+1} \neq 1$. We write

$$\boldsymbol{u} = (1 - \alpha_{s+1})\boldsymbol{z} + \alpha_{s+1} \boldsymbol{u}_{s+1},$$

with

$$\boldsymbol{z} = \frac{\alpha_1}{1 - \alpha_{s+1}} \boldsymbol{u}_1 + \ldots + \frac{\alpha_s}{1 - \alpha_{s+1}} \boldsymbol{u}_s.$$

Since

$$\frac{\alpha_1}{1 - \alpha_{s+1}} \geq 0, \ldots, \frac{\alpha_s}{1 - \alpha_{s+1}} \geq 0, \quad \sum_{i=1}^{s} \frac{\alpha_i}{1 - \alpha_{s+1}} = 1,$$

according to the induction hypothesis, it follows that $\boldsymbol{z} \in C$, and by the convexity of C we have that $\boldsymbol{u} \in C$.

To proof the converse, we take $s = 2$ and use the definition of a convex set. $\square$

Proposition

Given $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^n$, all the points of the **segment** *determined by them can be written as the convex combination*

$$\boldsymbol{u} = \alpha \boldsymbol{u}_1 + (1-\alpha)\boldsymbol{u}_2, \quad 0 \leq \alpha \leq 1.$$

Theorem

The intersection of an arbitrary family of convex sets is also a convex set.

Proof: Let $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ be points contained in the intersection. Then they are also contained in every member of the family and so is

$$\boldsymbol{x} = \lambda \boldsymbol{u}_1 + (1-\lambda)\boldsymbol{u}_2, \quad 0 \leq \lambda \leq 1.$$

□

IX.1 The convex hull

1. The **intersection of all the convex sets containing an arbitrary set** $A \subset \mathbb{R}^n$, is called the **convex hull of** A, and will be denoted by $C(A)$.

2. The convex hull of A can also be defined as **the set of all the convex combinations of points of** A.

3. According to the preceeding theorem, $C(A)$ **is a convex set**; it is actually the smallest convex set in $\mathbb{R}^n$ containing A.

4. **Examples:**
   - If A is the set defined by the 8 vertices of a cube, then $C(A)$ if the full cube.
   - If A is a circumference, then $C(A)$ is the cercle determined by it.

5. If $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^n$, we define the **sum of the two sets** as
$$X + Y = \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in X, \boldsymbol{y} \in Y\},$$
and if $\lambda \in \mathbb{R}$ and $X \subset \mathbb{R}^n$, we define the **product** λX as
$$\lambda X = \{\lambda \boldsymbol{x} \mid \boldsymbol{x} \in X\}.$$

6. **If** X **and** Y **are two convex subsets** of $\mathbb{R}^n$ and $\lambda \in \mathbb{R}$, then $X \pm Y$ **and** λX **are also convex sets.**

IX.1 Separating hyperplanes

- Many important results of nonlinear optimization (including Farkas Lemma) can be proved by using the so-called **separation theorems** of convex sets.

- Let S and T be nonempty subsets of $\mathbb{R}^n$. Then, the hyperplane $H = \left\{\boldsymbol{x} | \boldsymbol{a}^T \boldsymbol{x} = b\right\}$ is said to **separate** S and T if:

$$S \subset \left\{\boldsymbol{x} \mid \boldsymbol{a}^T \boldsymbol{x} \geq b\right\}, \; T \subset \left\{\boldsymbol{x} \mid \boldsymbol{a}^T \boldsymbol{x} \leq b\right\},$$

  this is, S is **contained in one of the closed** half spaces generated by H and T is **contained in the opposite closed** half space.

- A hyperplane **strictly separates** S and T if:

$$S \subset \left\{\boldsymbol{x} \mid \boldsymbol{a}^T \boldsymbol{x} > b\right\}, \; T \subset \left\{\boldsymbol{x} \mid \boldsymbol{a}^T \boldsymbol{x} < b\right\},$$

  this is, S is **contained in one of the open** half spaces generated by H and T is **contained in the opposite open** half space.

- Such hyperplanes are called a **separating hyperplanes.**

Lemma

Let C be a nonempty **closed convex** *set in $\mathbb{R}^n$* **not containing the origin** *($\mathbf{0} \notin C$). Then there exists a* **hyperplane that strictly separates** *C and $\{\mathbf{0}\}$.*

Lemma

Let C be a nonempty **convex** *set in $\mathbb{R}^n$* **not containing the origin** *($\mathbf{0} \notin C$). Then there exists a* **hyperplane that separates** *C and $\{\mathbf{0}\}$.*

Theorem

*(***Strict Separation Theorem***) Let C_1 and C_2 be two* **disjoint** *nonempty* **closed convex** *sets in $\mathbb{R}^n$, and suppose that C_2* **is compact***. Then there exists a* **hyperplane that strictly separates** *them.*

Theorem

*(***Separation Theorem***) Let C_1 and C_2 be two* **disjoint** *nonempty* **convex** *sets in $\mathbb{R}^n$. Then there exists a* **hyperplane that separates** *them.*

IX.1 An application: Farkas Lemma

Lemma

Let A be a given $m \times n$ real matrix and $\boldsymbol{b} \in \mathbb{R}^n$ a given vector. The inequality $\boldsymbol{b}^T \boldsymbol{y} \geq 0$ holds for all vectors $\boldsymbol{y} \in \mathbb{R}^n$ satisfying $A\boldsymbol{y} \geq 0$ if and only if there exists a vector $\boldsymbol{\rho} \in \mathbb{R}^m$, $\boldsymbol{\rho} \geq 0$, such that $A^T \boldsymbol{\rho} = \boldsymbol{b}$.

Proof: The statement

$$\forall \boldsymbol{y} \mid A\boldsymbol{y} \geq 0 \;\Rightarrow\; \boldsymbol{b}^T \boldsymbol{y} \geq 0 \quad \text{if and only if} \quad \exists \boldsymbol{\rho} \in \mathbb{R}^m,\, \boldsymbol{\rho} \geq 0 \mid A^T \boldsymbol{\rho} = \boldsymbol{b},$$

is equivalent to saying that

$$\left.\begin{array}{rcl} A\boldsymbol{y} & \geq & \boldsymbol{0} \\ \boldsymbol{b}^T \boldsymbol{y} & < & 0 \end{array}\right\} \text{has a solution if and only if} \quad \left.\begin{array}{rcl} A^T \boldsymbol{\rho} & = & \boldsymbol{b} \\ \boldsymbol{\rho} & \geq & 0 \end{array}\right\} \text{has no solution.}$$

Assume first that the second system has no solution. This is true if and only if the nonempty convex sets

$$C_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = A^T \boldsymbol{\rho},\, \boldsymbol{\rho} \geq 0 \right\}, \quad C_2 = \{\boldsymbol{b}\},$$

are disjoint. Note that C_2 is compact. According to the Strict Separation Theorem, there exist $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{c} \neq 0$ and $\alpha \in \mathbb{R}$ such that the hyperplane $H = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{c}^T \boldsymbol{x} = \alpha\}$ separates them, this is

$$\left\{\begin{array}{rrcl} & \boldsymbol{c}^T \boldsymbol{b} & < & \alpha, \\ \forall \boldsymbol{x} \in C_1, & \boldsymbol{c}^T \boldsymbol{x} & > & \alpha, \end{array}\right. \Leftrightarrow \quad \forall \boldsymbol{\rho} \geq 0, \;\; \boldsymbol{c}^T A^T \boldsymbol{\rho} > \alpha, \quad .$$

IX.1 Farkas Lemma (cont.)

- Take $\boldsymbol{\rho} = 0$ (since the only restriction is $\boldsymbol{\rho} \geq 0$). From $c^T A^T \boldsymbol{\rho} > \alpha$, we conclude that $0 > \alpha$, and since $\boldsymbol{c}^T \boldsymbol{b} < \alpha$, we get $\boldsymbol{c}^T \boldsymbol{b} = \boldsymbol{b}^T \boldsymbol{c} < 0$.

- If for a certain k we have that $(\boldsymbol{c}^T A^T)_k < 0$, then, choosing $\boldsymbol{\rho} = (0, ..., 0, \rho_k, 0, ..., 0)$ with $\rho_k \to +\infty$, we have that $\boldsymbol{c}^T A^T \boldsymbol{\rho} \to -\infty$, in contradiction with en $\boldsymbol{c}^T A^T \boldsymbol{\rho} > \alpha$, so we must have $\boldsymbol{c}^T A^T \geq 0$.

- According to the above two items, $\boldsymbol{c}$ satisfies the inequalities

$$\left\{ \begin{array}{rcl} \boldsymbol{b}^T \boldsymbol{c} & < & 0, \\ \boldsymbol{c}^T A^T & \geq & 0, \end{array} \right. \Leftrightarrow \quad A\boldsymbol{c} \geq 0.$$

So, we have found a solution of

$$\left. \begin{array}{rcl} A\boldsymbol{y} & \geq & 0, \\ \boldsymbol{b}^T \boldsymbol{y} & < & 0. \end{array} \right\}$$

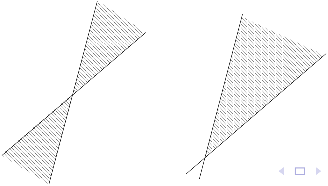
- Assume now the converse, this is: there are $\boldsymbol{\rho}$ and $\boldsymbol{y}$ such that: $A^T \boldsymbol{\rho} = \boldsymbol{b}$, $\boldsymbol{\rho} \geq 0$ and $A\boldsymbol{y} \geq 0$, then $\boldsymbol{b}^T \boldsymbol{y} = \boldsymbol{\rho}^T A \boldsymbol{y} \geq 0$, so the first system of inequalities has no solution. $\square$

IX.1 Convex polyhedra

- Let C be a convex set and $\boldsymbol{u} \in C$, we say that $\boldsymbol{u}$ is an **extremal point** if it cannot be written as a convex combination of points of C.
- **Examples:**
  - The points of a circumference around a circle.
  - The vertices of a triangle.
- The convex hull of a finite set of points S is called a **convex polyhedra**.
- If C **is a convex, bounded and closed set with a finite number of extremal points** (for instance, a convex polyhedra) then the points of C **can be written as a convex combination of the extremal points**, this is, C **is the convex hull of its extremal points.**
- A subset $S \subset \mathbb{R}^n$ is said to be a **cone**, if

$$\boldsymbol{u} \in S \;\Rightarrow\; \lambda \boldsymbol{u} \in S, \quad \forall \lambda \geq 0.$$

- The origin is always a point of a cone (since λ can be zero) but **not all the cones are convex.**



IX.1 Convex polyhedra

- A n-dimensional convex polyhedra with $n+1$ vertices is called **simplex.**
- A 0-dimensional simplex is a **point**, a 1-dimensional simplex is a **segment**, in dimension 2 a **triangle**, and in dimension 3 a **tetrahedron**.
- The faces of a simplex are also lower dimensional simplex.
- The n-**unitary simplex** is a subset of $\mathbb{R}^{n+1}$ defined by all the points $\boldsymbol{x} = (x_0, x_1, ..., x_n) \in \mathbb{R}^{n+1}$, such that

$$x_i \geq 0, \quad \sum_{i=1}^{n} x_i \leq 1.$$

- If $n = 3$ we get the tetrahedron with vertices at the points: $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(0,0,1)$.

- Obviously, the set of points (x_1, x_2) defined by a set of inequalities can be empty. For instance:

$$\begin{array}{rcl} x_1 + x_2 & \leq & 1, \\ 2x_1 + 2x_2 & \geq & 3. \end{array}$$

- **Exercise 10.** To be delivered before 16-XII-2019 (`Ex10-YourSurname.pdf`)

  Prove that the number of faces of dimension p of a n-dimensional simplex is equal to

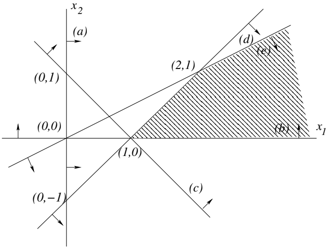
$$\left(\begin{array}{c} n+1 \\ p+1 \end{array} \right) = \frac{(n+1)!}{(p+1)!(n-p)!}.$$

IX.1 Convex sets and linear inequalities

Consider the following 2-dimensional linear inequalities

$$\begin{array}{lrcl}
(a) & x_1 & \geq & 0, \\
(b) & x_2 & \geq & 0, \\
(c) & x_1 + x_2 & \geq & 1, \\
(d) & x_1 - x_2 & \geq & 1, \\
(e) & -x_1 + 2x_2 & \leq & 0.
\end{array}$$

As the figure shows, each inequality can be used to define a separating hyperplane in $\mathbb{R}^2$ and they determine the shaded region



Changing the sign of the last inequality, the above system can also be written as

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

Using $\mathbf{p}_1 = (1, 0, 1, 1, 1)^T$, $\mathbf{p}_2 = (0, 1, 1, -1, -2)^T$, $\mathbf{p}_0 = (0, 0, 1, 1, 0)^T$, the above system becomes

$$x_1\mathbf{p}_1 + x_2\mathbf{p}_2 \geq \mathbf{p}_0.$$

Note that $x_1 + x_2 \neq 1$.

- In general, the set of points $(x_1, ..., x_n) \in \mathbb{R}^n$ that verify a linear inequality as:

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \geq b_1,$$

define a semi-space of $\mathbb{R}^n$.

- It can be shown that **the non-negative solutions** of:

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & \geq & b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & \geq & b_2, \\ \vdots & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & \geq & b_m, \end{array}$$

define a convex set of $\mathbb{R}^n$.

IX.1 Convex sets and linear inequalities

- The set of inequalities

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & \geq & b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & \geq & b_2, \\ \vdots & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & \geq & b_m, \end{array}$$

  can be also be written as a set of equalities.

- To do so, we must substract of each inequality an **unknown non-negative quantity**: x_{n+i}.

  These quatities, $x_{n+1}, ..., x_{n+m}$, are called **slack variables**, and with them the system becomes:

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n - x_{n+1} & = & b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n - x_{n+2} & = & b_2, \\ \vdots & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n - x_{n+m} & = & b_m, \end{array}$$

  with $x_{n+i} \geq 0$, $i = 1, ..., m$.

Since any number can be written as the difference between two non-negative numbers, the above system can also be written as

$$\begin{array}{rcl}
a_{11}(x_1' - x_1'') + a_{12}(x_2' - x_2'') + \cdots + a_{1n}(x_n' - x_n'') - x_{n+1} & = & b_1, \\
a_{21}(x_1' - x_1'') + a_{22}(x_2' - x_2'') + \cdots + a_{2n}(x_n' - x_n'') - x_{n+2} & = & b_2, \\
\vdots & & \vdots \\
a_{m1}(x_1' - x_1'') + a_{m2}(x_2' - x_2'') + \cdots + a_{mn}(x_n' - x_n'') - x_{n+m} & = & b_m,
\end{array}$$

with

$$\begin{array}{rcll}
x_j & = & x_j' - x_j'', & \\
x_j' & \geq & 0, & j = 1, \ldots, n, \\
x_j'' & \geq & 0, & j = 1, \ldots, n, \\
x_{n+i} & \geq & 0, & i = 1, \ldots, n.
\end{array}$$

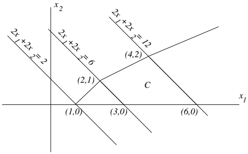
IX.1 Convex sets and linear programming problems

Definition: The general **linear programming problem** can be described as follows: Given a **convex set** defined by a **linear ser of constraints**, determine in which subset (that eventually can be a point) a certain **linear function (objective function**) has its maximum or minimum.

Example. Let C be the **convex set** defined by

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

Consider the **linear function** $2x_1 + 2x_2$ as the **objective function** to be minimised



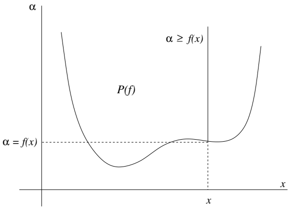
The figure shows the lines $2x_1 + 2x_2 = b$ for $b = 2, 6, 12$. Their intersections with C is a point for $b = 2$. Of all the feasible points of C only **one minimizes the objective function** which, furthermore, **is an extremal point of** C**.**

IX.2 Convex functions

Definition: Let f be a function defined on a subset $D \subset \mathbb{R}^n$ with values in the extended reals (this is, $f(\boldsymbol{x})$ is either a real number or it is $\pm\infty$). The subset of $\mathbb{R}^{n+1}$

$$P(f) = \{(\boldsymbol{x}, \alpha) \in D \times \mathbb{R} \mid f(\boldsymbol{x}) \leq \alpha\} \subset \mathbb{R}^{n+1},$$

is called the **epigraf** of f.



Definition: We define f to be a **convex function** if $P(f)$ is a convex set.

Examples: The function $f(\boldsymbol{x}) = +\infty$, $\boldsymbol{x} \in \mathbb{R}^n$ is a convex function, since $P(f) = \emptyset$. Similarly, $f(\boldsymbol{x}) = -\infty$ is also convex, since $P(f) = \mathbb{R}^{n+1}$.

Remark. We will see that this definition implies the one that we have already given in Lecture - 2:

$$f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \leq \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2), \ \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in D, \ 0 \leq \lambda \leq 1.$$

IX.2 Convex functions

- Consider a convex function f defined in a subset $D \subset \mathbb{R}^n$. Let

$$f_1(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & \text{if} \quad \boldsymbol{x} \in D, \\ +\infty & \text{if} \quad \boldsymbol{x} \notin D. \end{cases}$$

  The epigraph of the function f defined on D is idetical to the one of f_1. In this way, **we can always construct convex functions defined throughout $\mathbb{R}^n$.**

- In particular, let $a \in \mathbb{R}$, $\boldsymbol{b} \in \mathbb{R}^n$. Then

$$f_1(\boldsymbol{x}) = \begin{cases} a & \text{if} \quad \boldsymbol{x} = \boldsymbol{b}, \\ +\infty & \text{if} \quad \boldsymbol{x} \neq \boldsymbol{b}, \end{cases}$$

  is a convex function on $\mathbb{R}^n$.

- As a result of the above, we shall assume that, unless mentioned explicitly, **any convex function is defined on all $\mathbb{R}^n$**.

Remark. Note that all the above functions are not continuous.

IX.2 Convex functions

- The set
$$ED(f) = \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) < +\infty\},$$
is called the **effective domain**, ED, of a function f.
Note that $ED(f)$ is the projection of $P(f)$ on $\mathbb{R}^n$.

- If f is convex, then $ED(f)$ is also a convex set.

- The converse of this last statement generally does not hold.

IX.2 Convex functions

We shall be concerned mainly with **proper convex functions** defined as convex functions that:

- nowhere have the value $-\infty$, and
- are not identically equal to $+\infty$.

This is:

Definition

The function f is a proper convex function if:

1. *f is convex,*
2. *$f(\boldsymbol{x}) > -\infty$ for every $\boldsymbol{x}$,*
3. *$ED(f) \neq \emptyset$, this is: $f(\boldsymbol{x}) < +\infty$ for at least one $\boldsymbol{x}$.*

IX.2 Convex functions

Theorem

Let f be a proper convex function on $\mathbb{R}^n$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_s \in \mathbb{R}^n$ and $q_1, \ldots, q_s \in \mathbb{R}$ be numbers such that $q_i \geq 0$, $i = 1, \ldots, s$ and $q_1 + \ldots + q_s = 1$. Then

$$f(q_1\mathbf{x}_1 + \ldots + q_s\mathbf{x}_s) \leq q_1 f(\mathbf{x}_1) + \ldots + q_s f(\mathbf{x}_s). \tag{1}$$

Proof:

- If $f(\mathbf{x}_i) = +\infty$ for some i, then (1) trivially holds.
- Assume now that $f(\mathbf{x}_i) < +\infty$ for all i. Since f is convex, then the epigraph of f is a convex set and, according to a previous theorem (pag. 6), it contains every convex combination of its points.
- Hence, if $(\mathbf{x}_1, \alpha_1) \in P(f)$,..., $(\mathbf{x}_s, \alpha_s) \in P(f)$ and $q_1, \ldots, q_s \in \mathbb{R}$ are such that $q_i \geq 0$, $q_1 + \ldots + q_s = 1$, then, since $P(f)$ is convex, it follows that

  $$(q_1\mathbf{x}_1 + \ldots + q_s\mathbf{x}_s, q_1\alpha_1 + \ldots + q_s\alpha_s) \in P(f),$$

  this is

  $$f(q_1\mathbf{x}_1 + \ldots + q_s\mathbf{x}_s) \leq q_1\alpha_1 + \ldots + q_s\alpha_s.$$

- Since $(\mathbf{x}_i, \alpha_i) \in P(f) \ \Rightarrow \ f(\mathbf{x}_i) \leq \alpha_i$, we can take $\alpha_i = f(\mathbf{x}_i)$, for $i = 1, \ldots, n$, anf the inequality follows.

$\square$

Theorem

Let f and g be convex functions and $\lambda \in \mathbb{R}$, $\lambda \geq 0$. Then:

- λf *is also convex.*
- $f + g$ *is also convex, provided that the indefined operation* $+\infty + (-\infty)$ *is avoided.*

Corollary

Under the hypotheses of the above theorem, every linear combination $\lambda_1 f_1 + ... + \lambda_k f_k$ of convex functions with $\lambda_1 \geq 0, ..., \lambda_k \geq 0$, is also a convex function.

IX.2 Convex functions

Let Ψ be a function defined on $\mathbb{R}$ with values in the extended reals, it is said to be **non-decreasing** if for every $x_1 < x_2$ we have $\Psi(x_1) \leq \Psi(x_2)$.

The following theorem is useful in identifying convex functions or in constructing new convex functions from existing ones.

Theorem

Let f be a real convex function defined on $\mathbb{R}^n$, and let Ψ be a non-decresing proper convex function defined on $\mathbb{R}$. Then $\Psi(f(\boldsymbol{x}))$ is convex on $\mathbb{R}^n$

Proof: Since f is convex, given q_1, q_2 ($q_i \geq 0$, $q_1 + q_2 = 1$) we have that

$$f(q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2) \leq q_1 f(\boldsymbol{x}_1) + q_2 f(\boldsymbol{x}_2).$$

Since Ψ is non-decresing

$$\Psi(f(q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2)) \leq \Psi(q_1 f(\boldsymbol{x}_1) + q_2 f(\boldsymbol{x}_2))$$

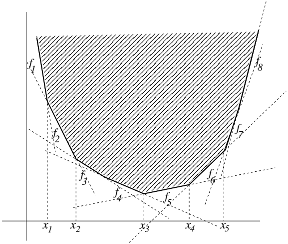
and by the convexity of Ψ

$$\Psi(f(q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2)) \leq \Psi(q_1 f(\boldsymbol{x}_1) + q_2 f(\boldsymbol{x}_2)) \leq q_1\Psi(f(\boldsymbol{x}_1)) + q_2\Psi(f(\boldsymbol{x}_2)),$$

so $\Psi(f(\boldsymbol{x}))$ is convex. $\square$

IX.2 Convex functions

- **Picewise linear functions**, as the one displayed in the figure (together with its epigraph) defined by f_1, f_2,... appear often in optimization.



- The convexity of picewise linear functions

$$f(\boldsymbol{x}) = \sup_{i \in I} f_i(\boldsymbol{x})$$

can be proven by noting that a linear function is convex and by the theorem that follows.

Remark: The **supremum of an ordered set** A is the **least upper bound of** A, this is: is the quantity S such that no member of the set exceeds S, but if ϵ is any positive quantity, however small, there is a member that exceeds $S - \epsilon$. The **infimum of an ordered set** A is the **greatest lower bound of** A

IX.2 Convex functions

Theorem

Let $\{f_i\}$, $i \in I$ be a finite or infinite collection of real convex functions defined on $\mathbb{R}^n$. For every $\boldsymbol{x} \in \mathbb{R}^n$, define

$$f(\boldsymbol{x}) = \sup_{i \in I} f_i(\boldsymbol{x}).$$

The function f is convex.

Proof:

- Since the f_i are convex, their epigraphs $P(f_i)$ are convex sets.
- Since the intersection of convex sets is also convex, we have that the intersection of the sets $P(f_i)$ is also convex.
- By definition

$$\bigcap_{i \in I} P(f_i) = \{(\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f_i(\boldsymbol{x}) \leq \alpha, \forall i \in I\} =$$

$$= \left\{(\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid \sup_{i \in I} f_i(\boldsymbol{x}) \leq \alpha\right\} = \{(\boldsymbol{x}, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(\boldsymbol{x}) \leq \alpha\},$$

so the intersection of the sets $P(f_i)$ is the epigraph of f, and the function f is convex.

$\square$

IX.2 Convex functions

We have seen that with every convex function f on $\mathbb{R}^n$ one can associate a convex set $P(f) \subset \mathbb{R}^{n+1}$. The next result deals with the converse statement, this is, **constructing convex functions on $\mathbb{R}^n$ from convex sets in $\mathbb{R}^{n+1}$.**

By **convention**, the infimum taken over the empty set is $+\infty$.

Theorem

Let C be a convex set in $\mathbb{R}^{n+1}$ and f the function defined on $\mathbb{R}^n$ by

$$f(\boldsymbol{x}) = \inf\{\alpha \in \mathbb{R} \mid (\boldsymbol{x}, \alpha) \in C\}.$$

Then, f is a convex function[1].

Proof: We need to show that $P(f)$ is convex.

$$P(f) = \{(\boldsymbol{x}, \beta) \mid f(\boldsymbol{x}) \leq \beta\} = \{(\boldsymbol{x}, \beta) \mid \inf\{\alpha \in \mathbb{R} \mid (\boldsymbol{x}, \alpha) \in C\} \leq \beta\}.$$

This is, if

$$(\boldsymbol{x}_1, \beta_1),\ (\boldsymbol{x}_2, \beta_2) \in P(f),$$

then

$$(q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2, q_1\beta_1 + q_2\beta_2) \in P(f),$$

for $q_1 + q_2 = 1,\ q_1,\ q_2 \geq 0$.

[1] Recall that the $\inf(A)$ is the greatest lower bound of A.

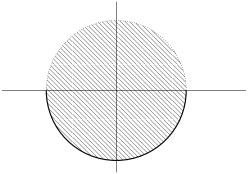
- Let $(\boldsymbol{x}_1, \beta_1)$, $(\boldsymbol{x}_2, \beta_2) \in P(f)$. From the definition of $P(f)$, it follows that there exist $\alpha_1 \le \beta_1$, $\alpha_2 \le \beta_2$ such that $(\boldsymbol{x}_1, \alpha_1)$, $(\boldsymbol{x}_2, \alpha_2) \in C$. Since C is convex
$$(q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2, q_1\alpha_1 + q_2\alpha_2) \in C,$$
for $q_1 + q_2 = 1$, q_1, $q_2 \ge 0$.
- According to the definition of the function f, and since $(q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2, q_1\alpha_1 + q_2\alpha_2) \in C$:
$$f(q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2) = \inf\{\alpha \in \mathbb{R} \mid (q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2, \alpha) \in C\} \le q_1\alpha_1 + q_2\alpha_2,$$
So, recalling that $\alpha_1 \le \beta_1$, $\alpha_2 \le \beta_2$:
$$f(q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2) \le q_1\alpha_1 + q_2\alpha_2 \le q_1\beta_1 + q_2\beta_2.$$
Thus
$$(q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2, q_1\beta_1 + q_2\beta_2) \in P(f).$$

$\square$

IX.2 Convex functions. Example

Example: Let C be the open unit circle

$$C = \{(x_1, x_2) \mid x_1^2 + x_2^2 < 1\}.$$



We construct f as the convex function defined on $\mathbb{R}$ by

$$f(x) = \begin{cases} -(1-x^2)^{1/2} & \text{if} \quad |x| < 1, \\ +\infty & \text{if} \quad |x| \geq 1. \end{cases}$$

The next theorem states that **a function is convex** on a convex set C if and only **if the restriction of f to each line segment in the set C is a convex function.**

Theorem

The function f defined on $\mathbb{R}^n$ is convex if and only if for every $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in \mathbb{R}^n$, the function ϕ defined by

$$\phi(\lambda) = f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2),$$

is convex for all $0 \leq \lambda \leq 1$.

Proof. Suppose that f is convex on $\mathbb{R}^n$ and let $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in \mathbb{R}^n$ be two arbitrary points. We must show that the epigraph of ϕ

$$P(\phi) = \{(\lambda, \alpha) \in [0,1] \times \mathbb{R} \mid \phi(\lambda) \leq \alpha\},$$

is a convex set.

Let (λ_1, α_1), $(\lambda_2, \alpha_2) \in P(\phi)$, we mush show that for every q_1, q_2 such that $q_i \geq 0$, $q_1 + q_2 = 1$:

$$(q_1\lambda_1 + q_2\lambda_2, q_1\alpha_1 + q_2\alpha_2) \in P(\phi) \quad \Leftrightarrow \quad \phi(q_1\lambda_1 + q_2\lambda_2) \leq q_1\alpha_1 + q_2\alpha_2.$$

IX.2 Convex functions (cont.)

Let

$$\boldsymbol{z}_1 = \lambda_1 \boldsymbol{x}_1 + (1-\lambda_1)\boldsymbol{x}_2, \quad \boldsymbol{z}_2 = \lambda_2 \boldsymbol{x}_1 + (1-\lambda_2)\boldsymbol{x}_2$$

then, according to the definition of ϕ, and that (λ_1, α_1), $(\lambda_2, \alpha_2) \in P(\phi)$:

$$f(\boldsymbol{z}_1) = \phi(\lambda_1) \le \alpha_1, \quad f(\boldsymbol{z}_2) = \phi(\lambda_2) \le \alpha_2.$$

Hence, $(\boldsymbol{z}_1, \alpha_1)$, $(\boldsymbol{z}_2, \alpha_2) \in P(f)$ and, since $P(f)$ is convex, we also have that $(q_1\boldsymbol{z}_1 + q_2\boldsymbol{z}_2, q_1\alpha_1 + q_2\alpha_2) \in P(f)$ for every q_1, q_2 such that $q_i \ge 0$, $q_1 + q_2 = 1$.

Since $(q_1\boldsymbol{z}_1 + q_2\boldsymbol{z}_2, q_1\alpha_1 + q_2\alpha_2) \in P(f)$, it follows that

$$f(q_1\boldsymbol{z}_1 + q_2\boldsymbol{z}_2) \le q_1\alpha_1 + q_2\alpha_2.$$

According to the definitions

$$\begin{aligned}
\phi(q_1\lambda_1 + q_2\lambda_2) &= f\left[(q_1\lambda_1 + q_2\lambda_2)\boldsymbol{x}_1 + (1 - q_1\lambda_1 - q_2\lambda_2)\boldsymbol{x}_2\right] \\
&= f\left[(q_1\lambda_1 + q_2\lambda_2)\boldsymbol{x}_1 + (q_1 + q_2 - q_1\lambda_1 - q_2\lambda_2)\boldsymbol{x}_2\right] \\
&= f\left[(q_1[\lambda_1\boldsymbol{x}_1 + (1-\lambda_1)\boldsymbol{x}_2] + q_2[\lambda_2\boldsymbol{x}_1 + (1-\lambda_2)\boldsymbol{x}_2]\right] \\
&= f(q_1\boldsymbol{z}_1 + q_2\boldsymbol{z}_2).
\end{aligned}$$

And so $(q_1\lambda_1 + q_2\lambda_2, q_1\alpha_1 + q_2\alpha_2) \in P(\phi)$, that is: ϕ is convex.

The proof of the converse statement is similar. $\square$

IX.2 Continuity of convex functions

- Roughly speaking, discontinuities in convex functions can occur only at some boundary points of their effective domain.
- Some of these discontinuities can be eliminated by the **closure** operation for convex functions.
- Let f be a convex function on $\mathbb{R}^n$. For a given point $\boldsymbol{x}_0 \in \mathbb{R}^n$, consider linear functions h, $h(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x} - b$, such that $h(\boldsymbol{x}_0) \leq f(\boldsymbol{x}_0)$.
- The **support set** of a convex function f is defined by

$$L(f) = \left\{(\boldsymbol{a}, b) \in \mathbb{R}^n \times \mathbb{R} \mid \boldsymbol{a}^T \boldsymbol{x} - b \leq f(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathbb{R}^n\right\}.$$

- Define the **closure of a convex function**, denoted by $\mathrm{cl} f$, as

$$\mathrm{cl} f(\boldsymbol{x}) = \sup_{(\boldsymbol{a}, b) \in L(f)} \{\boldsymbol{a}^T \boldsymbol{x} - b\}.$$

  Clearly, $\mathrm{cl} f(\boldsymbol{x}) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$.
- A convex function f is said to be **closed** if $\mathrm{cl} f = f$.

IX.2 Continuity of convex functions

- It can be shown that a proper convex function f is closed if and only if the convex level set $\{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \leq \alpha\}$ is closed for every real number $\alpha \in \mathbb{R}$.

- The closure operation for proper convex functions is related to the closure operation for sets. **The epigraph of** $\text{cl} f$ **is the closure of the epigraph of** f

$$P(\text{cl} f) = \overline{P(f)}.$$

- This **last relation can also be used as the definition** of the closure operation for proper convex functions.

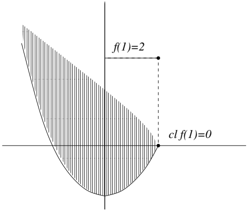
- For improper convex functions, the closure operation has a simple meaning:
  - If $f(\boldsymbol{x}) = +\infty$ for every $\boldsymbol{x} \in \mathbb{R}^n$, then $\text{cl} f = f$.
  - If $f(\boldsymbol{x}) = -\infty$ for some $\boldsymbol{x} \in \mathbb{R}^n$, then the support set $L(f) = \emptyset$, and since, by convention, the supremum over the empty set is $-\infty$, we get $\text{cl} f(\boldsymbol{x}) = -\infty$ for all $\boldsymbol{x}$.

- Consequently, **the only closed improper convex functions are those that are identically $f(x) = \pm\infty$.**

Example: Let us apply the closure operation to the convex function defined by

$$f(x) = \begin{cases} x^2 - 1, & x < 1, \\ 2, & x = 1, \\ +\infty, & x > 1. \end{cases}$$

Since this is a proper convex function **we simply close the epigraph of f,** as is shown in the figure



In this way, we get

$$\mathrm{cl} f(x) = \begin{cases} x^2 - 1, & x \leq 1, \\ +\infty, & x > 1. \end{cases}$$

In other words, the closure operation in this example **lowered the value of f at the boundary of its effective domain until the function became continuous over $ED(f)$.**

IX.2 Continuity of convex functions

- As we just saw in the example, **the closure operation** for convex functions **eliminates certain discontinuities** at boundary points of the effective domain.
- Since **convex functions are continuous on the interior of their effective domains**, the function $\mathrm{cl}f$ agrees with f at every point in the interior of $ED(f)$.
- A slightly more general result holds that requires the notion of **relative interior of a nonempty convex set**.
- Let $B \subset \mathbb{R}^n$ such that for every two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in B$ and for every $\alpha \in \mathbb{R}$, then $\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2 \in B$. A set satisfying this property is called an **affine set**.
- Single points, lines, and hyperplanes are examples of affine sets in $\mathbb{R}^n$.
- Given a convex set $C \subset \mathbb{R}^n$, the intersection of all affine sets containing C is called the **affine hull** of C.
- The **relative interior** of C, denoted by $ri(C)$, is defined as the interior of C viewed as a subset of its affine hull.

Example: Let

$$C = \{(x, 0) \mid a \leq x \leq b\} \subset \mathbb{R}_2.$$

This set has no interior if we view C as a subset of $\mathbb{R}_2$, since one cannot find an open disc in $\mathbb{R}_2$ contained in C. The affine hull of C is the hole x axis, and the relative interior of C is

$$\mathrm{ri}(C) = \{(x, 0) \mid a < x < b\},$$

which is the interior of C viewed as a subset of the x axis.

Remark: Note that if the convex set $C \subset \mathbb{R}^n$ is n-dimensional, such as a square in $\mathbb{R}_2$ or a cube in $\mathbb{R}^3$, then the affine hull of C is $\mathbb{R}^n$, so the relative interior of C coincides with the interior of C.

IX.2 Continuity of convex functions

The proofs of the next two theorems require additional concepts from topology (and is not given here[2]). The first theorem deals with certain improper convex functions. Such a function has the value $+\infty$ everywhere or it takes on the value $-\infty$ at some points of its effective domain.

Theorem

If f is an improper function, then $f(\boldsymbol{x}) = -\infty$ for every $\boldsymbol{x}$ in the relative interior of its effective domain.

Theorem

A convex function is continuous on the relative interior of its effective domain.

From this theorem follows

Corollary

A real-valued convex function on $\mathbb{R}^n$ is continuous everywhere.

[2] For the proofs see, for instance, R.T. Rockafellar: Convex Analysis. Princenton University Press, 1970.

IX.2 Continuity of convex functions

The following results deal with the existence of solutions to systems of inequalities involving convex functions.

Theorem

Let $f_1,\ldots,f_m$ be proper convex functions and let C be a nonempty convex set such that

$$C \subset \bigcap_{i=1}^{m} ED(f_i).$$

Then, exactly one of the alternatives holds:

1. *There exists an $\boldsymbol{x}_0 \in C$ such that*

$$f_i(\boldsymbol{x}_0) < 0, \quad i = 1, \ldots, m.$$

2. *There exist nonnegative numbers $\alpha_1,\ldots,\alpha_m$ (not all zero) such that for every $\boldsymbol{x} \in C$*

$$\sum_{i=1}^{m} \alpha_i f_i(\boldsymbol{x}) \geq 0.$$

IX.2 Continuity of convex functions

Theorem

Let $f_0, f_1, \ldots, f_m$ be proper convex functions and let C be a nonempty convex set such that $C \subset \bigcap_{i=1}^m ED(f_i)$.
If the system of inequalities

$$\begin{aligned} f_0(\mathbf{x}) &< 0, \\ f_i(\mathbf{x}) &\leq 0, \quad i = 1, \ldots, m, \end{aligned}$$

has no solution $\mathbf{x} \in C$, while there exists an $\mathbf{x}_0 \in C$ such that

$$f_i(\mathbf{x}_0) < 0, \quad i = 1, \ldots, m,$$

then, either

$$f_0(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in C,$$

or there exist non-negative numbers $\alpha_1, \ldots, \alpha_m$ not all zero, such that

$$f_0(\mathbf{x}) + \sum_{i=1}^m \alpha_i f_i(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in C.$$

IX.2 Differential properties of convex functions

- Given $f : S \longrightarrow \mathbb{R}$ with $S \subset \mathbb{R}^n$, a point $\boldsymbol{x}_0$ in the interior of S, and any vector $\boldsymbol{y} \in \mathbb{R}^n$, the derivative of f at $\boldsymbol{x}_0$ in the direction of $\boldsymbol{y}$ is defined as

$$D_{\boldsymbol{y}} f(\boldsymbol{x}_0) \equiv Df(\boldsymbol{x}_0; \boldsymbol{y}) = \lim_{t \to 0} \frac{f(\boldsymbol{x}_0 + t\boldsymbol{y}) - f(\boldsymbol{x}_0)}{t}.$$

- When $t \to 0^+$ or $t \to 0^-$ we talk about the **right-sided** and **left-sided** derivatives of f at $\boldsymbol{x}_0$, they are denoted by $D^+ f(\boldsymbol{x}_0; \boldsymbol{y})$ and $D^- f(\boldsymbol{x}_0; \boldsymbol{y})$.
- If $\boldsymbol{y} = \boldsymbol{0}$, then $D^+ f(\boldsymbol{x}_0; \boldsymbol{0}) = D^- f(\boldsymbol{x}_0; \boldsymbol{0}) = 0$.
- One can easily verify that $D^+ f(\boldsymbol{x}_0; -\boldsymbol{y}) = -D^- f(\boldsymbol{x}_0; \boldsymbol{y})$.
- If for some $\boldsymbol{x}_0, \ \boldsymbol{y} \in \mathbb{R}^n$

$$D^+ f(\boldsymbol{x}_0; \boldsymbol{y}) = D^- f(\boldsymbol{x}_0; \boldsymbol{y}),$$

  then

$$D^+ f(\boldsymbol{x}_0; \boldsymbol{y}) = D^- f(\boldsymbol{x}_0; \boldsymbol{y}) = Df(\boldsymbol{x}_0; \boldsymbol{y}).$$

IX.2 Differential properties of convex functions

- If $\boldsymbol{y}$ is a vector of the canonical basis, then the directional derivatives are the partial derivatives.
- If f is differentiable at $\boldsymbol{x}_0$, then the directional derivatives of f at $\boldsymbol{x}_0$ in all directions $\boldsymbol{y}$ are finite and are given by

$$Df(\boldsymbol{x}_0; \boldsymbol{y}) = \boldsymbol{y}^T \nabla f(\boldsymbol{x}_0) = \nabla f(\boldsymbol{x}_0)^T \boldsymbol{y}.$$

- A function f is said to be **positively homogeneous of degree** k if for every $\boldsymbol{x} \in \mathbb{R}^n$ and every $t \in \mathbb{R}^+$

$$f(t\boldsymbol{x}) = t^k f(\boldsymbol{x}).$$

Theorem

Let f be a convex function and let $\boldsymbol{x} \in \mathbb{R}^n$ be a point such that $f(\boldsymbol{x}) < \infty$. Then:

- *For any $\boldsymbol{y} \in \mathbb{R}^n$ there exist the right-sided and left-sided derivatives of f at $\boldsymbol{x}$: $D^+ f(\boldsymbol{x}; \boldsymbol{y})$, $D^- f(\boldsymbol{x}; \boldsymbol{y})$.*
- *$D^+ f$ and $D^- f$ are positively homogeneous convex functions of $\boldsymbol{y}$ of degree one.*
- *The following inequality holds:*

$$D^+ f(\boldsymbol{x}; \boldsymbol{y}) \geq D^- f(\boldsymbol{x}; \boldsymbol{y}).$$

IX.2 Differential properties of convex functions

The next concept to be introduced is the **subgradient**, which is related to:

- the ordinary gradient in the case of differentiable convex functions,
- to the directional derivatives in the more general case.

Definition

A **subgradient** *of a convex function f at a point $\boldsymbol{x} \in \mathbb{R}^n$, is a vector $\boldsymbol{\xi} \in \mathbb{R}^n$ such that*

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \boldsymbol{\xi}^T(\boldsymbol{y} - \boldsymbol{x}), \tag{2}$$

for every $\boldsymbol{y} \in \mathbb{R}^n$.

For a convex function f it is possible that, at some point $\boldsymbol{x}$

1. No vector $\boldsymbol{\xi} \in \mathbb{R}^n$ satisfying (2) exists.
2. There is a unique vector $\boldsymbol{\xi} \in \mathbb{R}^n$ satisfying (2).
3. There is more than one vector $\boldsymbol{\xi} \in \mathbb{R}^n$ satisfying (2).

We denote by $\partial f(\boldsymbol{x})$ the set of all subgradients of a convex function f at $\boldsymbol{x}$.

IX.2 Differential properties of convex functions

Some basic properties of subgradients are:

- The set $\partial f(\boldsymbol{x})$, also called **subdifferential** of f, is a closed convex set.
- The set $\partial f(\boldsymbol{x})$ contains a single vector $\boldsymbol{\xi} \in \mathbb{R}^n$ if and only if the convex function f is differentiable in the ordinary sense at $\boldsymbol{x}$ and then $\boldsymbol{\xi} = \nabla f(\boldsymbol{x})$, that is
$$\xi_j = \frac{\partial f(\boldsymbol{x})}{\partial x_j}, \quad j = 1, \ldots, n.$$

Subgradients can be characterized by the directional derivatives, as the following theorem shows.

Theorem

A vector $\boldsymbol{\xi} \in \mathbb{R}^n$ is a subgradient of a convex function f at a point $\boldsymbol{x}$ where $|f(\boldsymbol{x})| < +\infty$ (this is: $\boldsymbol{\xi} \in \partial f(\boldsymbol{x})$) if and only if

$$D^+ f(\boldsymbol{x}; \boldsymbol{y}) \geq \boldsymbol{\xi}^T \boldsymbol{y}, \quad \forall \boldsymbol{y} \in \mathbb{R}^n. \tag{3}$$

IX.2 Proof of the theorem

Proof: If $\boldsymbol{\xi} \in \partial f(\boldsymbol{x})$, then it satisfies

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \boldsymbol{\xi}^T(\boldsymbol{y} - \boldsymbol{x}), \quad \forall \boldsymbol{y} \in \mathbb{R}^n.$$

Let $\boldsymbol{y} = \boldsymbol{x} + t\boldsymbol{z}$, with $t > 0$, then

$$f(\boldsymbol{x} + t\boldsymbol{z}) \geq f(\boldsymbol{x}) + t\boldsymbol{\xi}^T\boldsymbol{z},$$

for every $\boldsymbol{z} \in \mathbb{R}^n$ and $t > 0$.
Dividing both sides by t and rearranging, we get

$$\frac{f(\boldsymbol{x} + t\boldsymbol{z}) - f(\boldsymbol{x})}{t} \geq \boldsymbol{\xi}^T\boldsymbol{z}.$$

The inequality $D^+f(\boldsymbol{x};\boldsymbol{z}) \geq \boldsymbol{\xi}^T\boldsymbol{z}$ then holds by noting that $D^+f(\boldsymbol{x};\boldsymbol{z})$ is the infimum of the left hand side quotient.

Conversely, if $D^+f(\boldsymbol{x};\boldsymbol{z}) \geq \boldsymbol{\xi}^T\boldsymbol{z}$ holds for every $\boldsymbol{z} \in \mathbb{R}^n$, then

$$f(\boldsymbol{x} + t\boldsymbol{z}) \geq f(\boldsymbol{x}) + t\boldsymbol{\xi}^T\boldsymbol{z},$$

also holds by the same argument as above and, consequently

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \boldsymbol{\xi}^T(\boldsymbol{y} - \boldsymbol{x}).$$

This, in turn, implies that $\boldsymbol{\xi}$ is a subgradient of f at $\boldsymbol{x}$. $\square$

Corollary

Let f be a convex function on $\mathbb{R}^n$ and suppose that $f(\boldsymbol{x})$ is finite. Then

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + D^+ f(\boldsymbol{x}; \boldsymbol{y} - \boldsymbol{x}),$$

for every $\boldsymbol{y} \in \mathbb{R}^n$. In particular, if f is differentiable at $\boldsymbol{x}$, then

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + (\boldsymbol{y} - \boldsymbol{x})^T \nabla f(\boldsymbol{x}).$$

Proof:

$$D^+ f(\boldsymbol{x}; \boldsymbol{y} - \boldsymbol{x}) = \inf_{t \geq 0} \frac{f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{t} = \inf_{t \geq 0} \frac{f(t\boldsymbol{y} + (1-t)\boldsymbol{x}) - f(\boldsymbol{x})}{t} \leq$$

$$\leq \inf_{t \geq 0} \frac{tf(\boldsymbol{y}) + (1-t)f(\boldsymbol{x}) - f(\boldsymbol{x})}{t} = \inf_{t \geq 0} \frac{t(f(\boldsymbol{y}) - f(\boldsymbol{x}))}{t} = f(\boldsymbol{y}) - f(\boldsymbol{x}).$$

$\square$

Corollary

Let f be a convex function on $\mathbb{R}^n$ and suppose that $f(\mathbf{x})$ and $f(\mathbf{y})$ are finite. Then

$$\begin{aligned} D^+ f(\mathbf{y}; \mathbf{y}-\mathbf{x}) &\geq D^+ f(\mathbf{x}; \mathbf{y}-\mathbf{x}), \\ D^- f(\mathbf{y}; \mathbf{y}-\mathbf{x}) &\geq D^- f(\mathbf{x}; \mathbf{y}-\mathbf{x}). \end{aligned}$$

In particular, if f is differentiable at $\mathbf{x}$ and $\mathbf{y}$, then

$$(\mathbf{y}-\mathbf{x})^T \left(\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\right) \geq 0.$$

Proof: By the preceeding corollary

$$\begin{aligned} f(\mathbf{y}) \geq f(\mathbf{x}) + D^+ f(\mathbf{x}; \mathbf{y}-\mathbf{x}) &\quad\Rightarrow\quad f(\mathbf{y}) - f(\mathbf{x}) \geq D^+ f(\mathbf{x}; \mathbf{y}-\mathbf{x}), \\ f(\mathbf{x}) \geq f(\mathbf{y}) + D^+ f(\mathbf{y}; \mathbf{x}-\mathbf{y}) &\quad\Rightarrow\quad -D^+ f(\mathbf{x}; \mathbf{x}-\mathbf{y}) \geq f(\mathbf{y}) - f(\mathbf{x}). \end{aligned}$$

Thus

$$-D^+ f(\mathbf{y}; \mathbf{x}-\mathbf{y}) \geq D^+ f(\mathbf{x}; \mathbf{y}-\mathbf{x}),$$

and using a previous Theorem (see page 43), together with the relation between D^+ and D^-, we get

$$D^+ f(\mathbf{y}; \mathbf{y}-\mathbf{x}) \geq D^- f(\mathbf{y}; \mathbf{y}-\mathbf{x}) = -D^+ f(\mathbf{y}; \mathbf{x}-\mathbf{y}) \geq D^+ f(\mathbf{x}; \mathbf{y}-\mathbf{x}).$$

The proof of the similar result for the left-sided derivatives is identical. $\square$

IX.2 Differential properties of convex functions

- We have already seen that the convexity of a function f on $\mathbb{R}^n$ is equivalent to the convexity of its restriction to any line segment in $\mathbb{R}^n$.
- Therefore, in some cases, it is sufficient to study the behavior of convex functions on the real line $\mathbb{R}$ where, often, the results are considerably simpler.
- For example, in the one-dimensional case all the right- and left-sided derivatives of f at a point x can be computed from $D^+(x, 1)$ and $D^-(x, 1)$, respectively, because of the homogeneity of the derivatives.
- Next, we will see that these derivatives are monotone non-decreasing functions of x.

Theorem

Let f be a convex function on $\mathbb{R}$ and let $x_2 > x_1$ be two points such that $f(x_1)$ and $f(x_2)$ are both finite. Then

$$D^+ f(x_2;1) \geq D^- f(x_2;1) \geq D^+ f(x_1;1) \geq D^- f(x_1;1).$$

Proof: The first and last inequalities have been already proved in a previous Theorem (see page 43).
If f is convex, using the the Corollary in page 47, the relation between D^+ and D^-, and the homogeneity of D^-, it follows that

$$\begin{aligned} f(x_1) &\geq f(x_2) + D^+ f(x_2; x_1 - x_2) = f(x_2) - D^- f(x_2; x_2 - x_1) \\ &= f(x_2) - (x_2 - x_1) D^- f(x_2;1), \end{aligned}$$

thus

$$f(x_1) - f(x_2) \geq -(x_2 - x_1) D^- f(x_2;1) = (x_1 - x_2) D^- f(x_2;1).$$

Since $x_1 - x_2 < 0$ we get

$$D^- f(x_2;1) \geq \frac{f(x_1) - f(x_2)}{x_1 - x_2}.$$

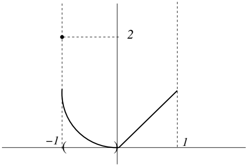
Analogously, we can prove that

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \geq D^+ f(x_1;1).$$

IX.2 Example

Consider the convex function f defined on $\mathbb{R}$ by

$$f(x) = \begin{cases} +\infty & x < -1, \\ 2 & x = -1, \\ x^2 & -1 < x \leq 0, \\ x & 0 \leq x \leq 1, \\ +\infty & 1 < x. \end{cases}$$



Using the definitions we can compute

$$D^+ f(x;1) = \begin{cases} \text{undefined} & x < -1 \\ -\infty & x = -1 \\ 2x & -1 < x < 0 \\ 1 & 0 \leq x < 1 \\ +\infty & x = 1 \\ \text{undefined} & 1 < x \end{cases} \qquad D^- f(x;1) = \begin{cases} \text{undefined} & x < -1 \\ -\infty & x = -1 \\ 2x & -1 < x \leq 0 \\ 1 & 0 < x \leq 1 \\ \text{undefined} & 1 < x \end{cases}$$

IX.2 Example (cont.)

For instance:

$$D^+f(-1;1) = \lim_{t\to 0^+} \frac{f(-1+t)-f(-1)}{t} = \lim_{t\to 0^+} \frac{(t-1)^2-2}{t} = \lim_{t\to 0^+} (t-2-\frac{1}{t}) = -\infty.$$

We can see that $D^+f(x;1) = D^-f(x;1)$ for $-1 \le x < 0$ and $0 < x < 1$, and that $D^+f(x;1) > D^-f(x;1)$ for $x = 0,\ 1$.

Recall that ξ is a subgradient of f if and only if

$$D^+f(x;z) \ge \xi z,$$

for every $z \in \mathbb{R}$. Since one-sided derivatives are positively homogeneous, we get

$$D^+f(x;z) = \begin{cases} zD^+f(x;1) & z > 0, \\ 0 & z = 0, \\ -zD^+f(x;-1) & z < 0. \end{cases}$$

Thus, $\xi \in \partial f(x)$ if and only if

$$D^+f(x;1) \ge \xi \ge D^-f(x;1),$$

since $D^-(x;1) = -D^+f(x;-1)$

IX.2 Example (cont.)

Consequently

$$\partial f(x) = \begin{cases} \emptyset & x \le -1, \\ 2x & -1 < x < 0, \\ \{\xi \mid 0 \le \xi \le 1\} & x = 0, \\ 1 & 0 < x < 1, \\ \{\xi \mid \xi \ge 1\} & x = 1, \\ \emptyset & x > 1. \end{cases}$$

IX.2 Some additional differential properties of convex functions

Theorem

Let f be a real-valued differentiable function on an open interval $D \subset \mathbb{R}$. Then the first derivative of f, f', is a non-decreasing function on D if and only if f is convex on D.

Proof: If f is convex, the result follows from the Theorem in page 53.
Let $x_1,\ x_2 \in D$ be such that $x_2 > x_1$, and let $x_3 = q_1 x_1 + q_2 x_2$ with $q_1 + q_2 = 1$, $q_1,\ q_2 \geq 0$. By the Mean value Theorem

$$\begin{aligned} f(x_2) &= f(x_3) + q_1(x_2 - x_1) f'(x^*), \quad x_2 \geq x^* \geq x_3, & (4) \\ f(x_3) &= f(x_1) + q_2(x_2 - x_1) f'(x^{**}), \quad x_3 \geq x^{**} \geq x_1. & (5) \end{aligned}$$

If f' is non-decreasing on $\mathbb{R}$, we have that $f'(x^*) \geq f'(x^{**})$, thus

$$f(x_3) \leq f(x_1) + q_2(x_2 - x_1) f'(x^*). \tag{6}$$

Multiplying (6) and (4) by q_1 and $-q_2$, respectively, and adding up, we get

$$q_1 f(x_3) - q_2 f(x_2) \leq q_1 f(x_1) - q_2 f(x_3),$$

so

$$q_1 f(x_3) + q_2 f(x_3) = f(x_3) = f(q_1 x_1 + q_2 x_2) \leq q_1 f(x_1) + q_2 f(x_2),$$

that is, f is a convex function. $\square$

IX.2 Some additional differential properties of convex functions

Most differential results established so far were necessary conditions for convex functions. The next Theorem shows that, in the special case of differentiable convex functions, these conditions are also quite easily established as sufficient ones.

Theorem

Let f be a real-valued differentiable function on $\mathbb{R}^n$. If

$$f(\boldsymbol{x}_2) \geq f(\boldsymbol{x}_1) + (\boldsymbol{x}_2 - \boldsymbol{x}_1)^T \nabla f(\boldsymbol{x}_1),$$

for every two points $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in \mathbb{R}^n$, then f is convex on $\mathbb{R}^n$.

Proof: Let $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in \mathbb{R}^n$ be any two points and $\boldsymbol{x}_3 = q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2$. Then

$$\begin{aligned} f(\boldsymbol{x}_1) &\geq f(\boldsymbol{x}_3) + q_2(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T \nabla f(\boldsymbol{x}_3), && (7) \\ f(\boldsymbol{x}_2) &\geq f(\boldsymbol{x}_3) + q_1(\boldsymbol{x}_2 - \boldsymbol{x}_1)^T \nabla f(\boldsymbol{x}_3). && (8) \end{aligned}$$

Multiplying (7) and (8) by q_1 and q_2, respectively, and adding up, we get

$$q_1 f(\boldsymbol{x}_1) + q_2 f(\boldsymbol{x}_2) \geq f(\boldsymbol{x}_3),$$

so f is convex. $\square$

IX.2 Some additional differential properties of convex functions

Corollary in page 50 [3] and the preceeding Theorem have a simple geometric meaning: a differentiable function f is convex on $\mathbb{R}^n$ if and only if the first two terms of the Taylor expansion of f at a point $\boldsymbol{x}_0$, that is, the linear function

$$f(\boldsymbol{x}_0) + (\boldsymbol{x} - \boldsymbol{x}_0)^T \nabla f(\boldsymbol{x}_0),$$

has values less than or equal to $f(\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathbb{R}^n$.

[3] f convex and differentiable at $\boldsymbol{x}_0 \Rightarrow f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0) + (\boldsymbol{x} - \boldsymbol{x}_0)^T \nabla f(\boldsymbol{x}_0)$ for all $\boldsymbol{x}$

IX.2 Some results involving twice differentiable convex functions

Theorem

Let f be a real-valued convex function on an open convex set $C \subset \mathbb{R}^n$. If f is differentiable on C, then f has continuous first partial derivatives on C.

Theorem

Let f be a real-valued twice differentiable function on an open interval $D \subset \mathbb{R}$. Then f is convex on D if and only if the second derivative of f, f'', evaluated at every $x \in D$ is nonnegative.

Proof: We have already seen that f is convex on D if and only if f' is non-decreasing (Theorem in page 57), that is, $f''(x) \geq 0$ for every $x \in D$. $\square$

The above theorem can be extended to the multidimentional case:

Theorem

Let f be a real-valued function on an open convex set $C \subset \mathbb{R}^n$ of class C^2. Then f is convex on C if and only if the Hessian of f evaluated at every $\boldsymbol{x} \in C$ is positive semidefinite. That is, for each $\boldsymbol{x} \in C$

$$\boldsymbol{y}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{y} \geq 0, \quad \forall \boldsymbol{y} \in \mathbb{R}^n.$$

This Theorem **cannot** be sharpened in the case of strictly convex functions by replacing *"positive semidefinite"* in the statement of the Theorem by *"positive definite"*.

IX.3 Extrema of convex functions

The main importance of convex functions lies in some basic properties that are summarized below.

Theorem

Let f be a proper convex function on $\mathbb{R}^n$. Then every local minimum of f is a global minimum of f on $\mathbb{R}^n$.

Proof: If $\boldsymbol{x}^*$ is a local minimum, then

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*),$$

for all $\boldsymbol{x} \in N_\delta(\boldsymbol{x}^*)$, if δ is small enough. Let $\boldsymbol{z} \in \mathbb{R}^n$ and $\lambda \in (0, 1)$ be such that $(1-\lambda)\boldsymbol{x}^* + \lambda\boldsymbol{z} \in N_\delta(\boldsymbol{x}^*)$. Then

$$f((1-\lambda)\boldsymbol{x}^* + \lambda\boldsymbol{z}) \geq f(\boldsymbol{x}^*).$$

Since f is a convex function

$$(1-\lambda)f(\boldsymbol{x}^*) + \lambda f(\boldsymbol{z}) \geq f((1-\lambda)\boldsymbol{x}^* + \lambda\boldsymbol{z}).$$

Using the last two inequalities, and dividing the result by λ, we get

$$f(\boldsymbol{z}) \geq f(\boldsymbol{x}^*).$$

$\square$

IX.3 Extrema of convex functions

- Consider the general problem

$$\begin{aligned} & \min f(\boldsymbol{x}), \\ \text{subject to:} \quad & g_i(\boldsymbol{x}) \geq 0, \quad i = 1, ..., m, \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, ..., p. \end{aligned}$$

- Suppose that the functions g_i are all convex and the h_j are all linear functions. Then, the feasible set X is convex set.
- If the objective function f to be minimized is a proper convex function in X, we can define a new objective function

$$\hat{f}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}), & \text{if} \quad \boldsymbol{x} \in X, \\ \infty, & \text{if} \quad \boldsymbol{x} \notin X, \end{cases} \tag{9}$$

  which is a proper convex function on $\mathbb{R}^n$ coinciding with f on X.
- From the previous Theorem we conclude that, if X is nonempty, every local minimum of f at some point $\boldsymbol{x} \in X$ is also a global minimum of f on all X. Formally:

Theorem

Let f be a proper convex function on $\mathbb{R}^n$ and $X \subset \mathbb{R}^n$ be a convex set. Then, every local minimum of f at $\boldsymbol{x} \in X$ is a global minimum of f over all X.

IX.3 Extrema of convex functions

Note that generaly the minimal value of a convex function can be attained at more than one point. Next, we shall see that the set of minimizing points of a proper convex function is a convex set.

Lemma

Let f be a convex function on $\mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Then, the level sets of f, given by

$$S(f,\alpha) = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \leq \alpha\},$$

are convex sets for any α.

Proof: Let $\mathbf{x}_1$, $\mathbf{x}_2 \in S(f,\alpha)$. It follows that for any q_1, q_2 such that $q_i \geq 0$, $q_1 + q_2 = 1$:

$$f(q_1\mathbf{x}_1 + q_2\mathbf{x}_2) \leq q_1 f(\mathbf{x}_1) + q_2 f(\mathbf{x}_2) \leq q_1\alpha + q_2\alpha = \alpha.$$

Hence, $S(f,\alpha)$ is convex. □

Theorem

Let f be a convex function on $\mathbb{R}^n$. The set of points at which f attains its minimum is convex.

Proof: Let α^* be the value of f at the minimizing points. Then, the set $\{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \leq \alpha^*\}$ is the set of points at which f attains its minimum; by the preceding Lemma, this is a convex set. □

Corollary

Let f be a strictly convex function defined on a convex set $X \subset \mathbb{R}^n$. If f attains its minimum on X, then it is attained at a unique point of X.

Proof: Suppose that the minimum is attained at two distinct points , $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in X$ and let $f(\boldsymbol{x}_1) = f(\boldsymbol{x}_2) = \alpha$. From the preceding Theorem, it follows that for every q_1, q_2 ($q_i \geq 0$, $q_1 + q_2 = 1$), we have $f(q_1\boldsymbol{x}_1 + q_2\boldsymbol{x}_2) = \alpha$, contradicting that f is strictly convex. $\square$

IX.3 Extrema of convex functions

In many applications, when the minimum of a differentiable function is sought, one looks for the points at which the gradient vanishes. This situation can be justified in the case of convex functions by the following

Theorem

Let f be a convex function. Then $0 \in \partial f(\boldsymbol{x}^*)$ if and only if f attains its minimum at $\boldsymbol{x}^*$.

Proof: By the definition of subgradients, $0 \in \partial f(\boldsymbol{x}^*)$ if and only if

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}^*)$$

for every $\boldsymbol{y} \in \mathbb{R}^n$; that is, $\boldsymbol{x}^*$ is a minimum of f. $\square$

Corollary

Let f be a differentiable convex function on $\mathbb{R}^n$. Then

$$\nabla f(\boldsymbol{x}^*) = 0$$

if and only if f attains its minimum at $\boldsymbol{x}^*$.

Proof: $\partial f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}$ if and only if f is differentiable at $\boldsymbol{x}$. Then the Corollary is a direct consequence of the preceding Theorem. $\square$

The Corollary will generally remain valid if we replace $\mathbb{R}$ by some open convex subset $X \subset \mathbb{R}^n$, such that $\boldsymbol{x}^* \in X$.

IX.4 Optimality conditions for convex problems

Consider the following nonlinear problem

$$\begin{aligned} & \min f(\boldsymbol{x}), \\ \text{subject to:} \quad & g_i(\boldsymbol{x}) \geq 0, \quad i = 1, ..., m, && (10) \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, ..., p, && (11) \end{aligned}$$

where f is a proper convex function on $\mathbb{R}^n$, the functions g_i are proper and concave and the h_j are linear functions of the form

$$h_j = \sum_{k=1}^{n} a_{jk} x_k - b_j.$$

Such a problem is called convex because the objective function is a convex function, and the set of all $\boldsymbol{x} \in \mathbb{R}^n$ satisfying the constraints is a convex set.

Note that generally the set of $\boldsymbol{x} \in \mathbb{R}^n$ satisfying satisfying an equation $h(\boldsymbol{x}) = 0$, where h is a nonlinear convex or concave function, is not a convex set.

IX.4 Optimality conditions for convex problems

Under an appropiate assumption of differentiability, the Kuhn-Tucker necessary conditions of optimality are also sufficient when applied to a convex problem.

Theorem

Suppose that the functions f, g_1,...,g_m are real-valued, differentiable, convex and concave functions on $\mathbb{R}^n$, respectively, and let h_1,...,h_p be linear. If there exist $\boldsymbol{x}^*$, $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$, with $\boldsymbol{x}^*$ satisfying (10) and (11), together with

$$\nabla_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \nabla f(\boldsymbol{x}^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) - \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*) = 0, \qquad (12)$$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, \quad i = 1, ..., m, \qquad (13)$$

$$\boldsymbol{\lambda}^* \geq 0,$$

then $\boldsymbol{x}^*$ is a global optimum of the convex minimization problem.

Remark: *In this Theorem we do not require the condition $\boldsymbol{z}^T \nabla^2 L \boldsymbol{z} \geq 0$ but it still guarantees the global character of the minimum.*

Proof: Let $\boldsymbol{x}$ be any point satisfying (10) and (11). Then

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}) - \sum_{i=1}^{m} \lambda_i^* g_i(\boldsymbol{x}) - \sum_{j=1}^{p} \mu_j^* h_j(\boldsymbol{x}). \qquad (14)$$

Applying Corollary in pag. 50[4] to f, g_i and h_j and using (14), we obtain

$$\begin{aligned}
f(\boldsymbol{x}) \quad & \geq \quad f(\boldsymbol{x}^*) + (\boldsymbol{x} - \boldsymbol{x}^*)^T \nabla f(\boldsymbol{x}^*) \\
& \quad - \textstyle\sum_{i=1}^{m} \lambda_i^* g_i(\boldsymbol{x}^*) - \sum_{i=1}^{m} \lambda_i^* (\boldsymbol{x} - \boldsymbol{x}^*)^T \nabla g_i(\boldsymbol{x}^*) \\
& \quad - \textstyle\sum_{j=1}^{p} \mu_j^* h_j(\boldsymbol{x}^*) - \sum_{j=1}^{p} \mu_j^* (\boldsymbol{x} - \boldsymbol{x}^*)^T \nabla h_j(\boldsymbol{x}^*) \\
& \geq \quad f(\boldsymbol{x}^*) - \textstyle\sum_{i=1}^{m} \lambda_i^* g_i(\boldsymbol{x}^*) - \sum_{j=1}^{p} \mu_j^* h_j(\boldsymbol{x}^*) \\
& \quad + (\boldsymbol{x} - \boldsymbol{x}^*)^T \left(\nabla f(\boldsymbol{x}^*) - \textstyle\sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) - \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*) \right),
\end{aligned}$$

And by (13), (11) and (12):

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*).$$

$\square$

[4] $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + (\boldsymbol{x} - \boldsymbol{x}^*)^T \nabla f(\boldsymbol{x}^*), \ \forall \boldsymbol{x} \in \mathbb{R}^n$

IX.4 Optimality conditions for convex problems

A nonlinear programming problem is said to be **strongly consistent** if there exists a point $\boldsymbol{x}_0 \in \mathbb{R}^n$ satisfying

$$g_i(\boldsymbol{x}_0) > 0, \quad i = 1, ..., m,$$

$$h_j(\boldsymbol{x}_0) = 0, \quad j = 1, ..., p.$$

We will also require that the vectors of coefficients $\boldsymbol{a}_j = (a_{j1}, ..., a_{jn})$ in the linear functions

$$h_j(x) = \sum_{k=1}^{n} a_{jk} x_k - b_j = \boldsymbol{a}_j^T \boldsymbol{x} - b_j,$$

to be linearly independent.

We can now prove the Kuhn-Tucker necessary conditions for optimality in a convex problem

IX.4 Optimality conditions for convex problems

Theorem

Suppose that the functions f, g_1,...,g_m are real-valued, differentiable, convex and concave functions on $\mathbb{R}^n$, respectively, and let

$$h_j = \sum_{k=1}^{n} a_{jk} x_k - b_j,$$

be such that the vectors $\boldsymbol{a}_j = (a_{j1}, ..., a_{jn})$, $j = 1, ..., p$ are linearly independent. Suppose thet the nonlinear programming problem is strongly consistent. If $\boldsymbol{x}^*$ is a solution of the problem, then there exist vectors $\boldsymbol{\lambda}^* = (\lambda_1^*, ..., \lambda_m^*)$ i $\boldsymbol{\mu}^* = (\mu_1^*, ..., \mu_p^*)$ such that

$$\nabla f(\boldsymbol{x}^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) - \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*) = 0,$$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, \quad i = 1, ..., m,$$

$$\boldsymbol{\lambda}^* \geq 0,$$