Optimization

Màster de Fonaments de Ciència de Dades

Gerard Gómez

Lecture V. Convexity

V.1 Convex sets

Definitions:

- Given m points $u_1, u_2, \dots, u_m \in \mathbb{R}^n$, we define a **convex combination** of these points as
$$u = \sum_{i=1}^{m} \alpha_i u_i,$$
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 $u_1, u_2 \in C$, any convex conbination of these points verifies
$$\alpha u_1 + (1-\alpha) u_2 \in C.$$

Examples:

1. The empty set, the set containing a single point $x \in \mathbb{R}^n$ and $\mathbb{R}^n$.
2. The n-dimensional sphere with center at $x_0 \in \mathbb{R}^n$ and radius α
$$S_\alpha(x_0) = \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq \alpha\}.$$

V.1 Convex sets. Hyperplanes

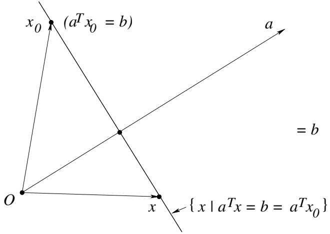
- Let $a \in \mathbb{R}^n$, $a \neq 0$ and b a given number, the set

$$\left\{x \,|\, a^T x = b\right\}$$

  is called a **hyperplane** of $\mathbb{R}^n$.
- A hyperplane is formed by all the vectors $x \in \mathbb{R}^n$ such that its scalar product with a is constant.
- If x_0 i x_1 are two points of the hyperplane, then

$$a^T(x_1 - x_0) = 0,$$

  by this reason the vector a is called the *normal* vector of the hyperplane.



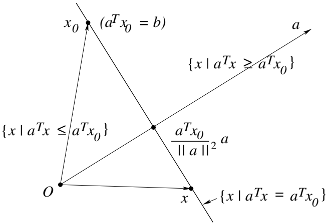
V.1 Hyperplanes. Properties

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

$$a^T(\lambda a) = \lambda a^T a = b, \quad \lambda = \frac{b}{\|a\|^2} = \frac{a^T x_0}{\|a\|^2},$$

   so

$$x = \frac{a^T x_0}{\|a\|^2} a$$



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:

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

V.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 s $x^1, \ldots, x^s \in C$. 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

$$x = \alpha_1 x^1 + \ldots + \alpha_s x^s + \alpha_{s+1} x^{s+1}.$$

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

$$x = (1 - \alpha_{s+1}) z + \alpha_{s+1} x^{s+1},$$

with

$$z = \frac{\alpha_1}{1 - \alpha_{s+1}} x^1 + \ldots + \frac{\alpha_s}{1 - \alpha_{s+1}} x^s.$$

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 $z \in C$, and by the convexity of C we have that $x \in C$.

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

V.1 Convex sets

Proposition

Given $u_1, u_2 \in \mathbb{R}^n$, all the points of the segment determined by them can be written as the convex combination

$$u = \alpha u_1 + (1-\alpha)u_2, \quad 0 \leq \alpha \leq 1$$

Theorem

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

Proof: Let x_1 and x_2 be points contained in the intersection. Then they are also contained in every member of the family and so is

$$x = \lambda x_1 + (1-\lambda)x_2, \quad 0 \leq \lambda \leq 1$$

□

V.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. According to the preceeding theorem, $C(A)$ is a convex set; it is actually the smallest convex set in $\mathbb{R}^n$ containing A.
3. The convex hull of A can also be defined as the set of all the convex combinations of points of 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

$$X + Y = \{x + y \mid x \in X, y \in Y\}\,.$$

   and if $\lambda \in \mathbb{R}$ and $X \subset \mathbb{R}^n$, we define the product λX as

$$\lambda X = \{\lambda x \mid x \in X\}\,.$$

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.

V.1 Convex sets and optimization

- Many important results of nonlinear optimization can be proves by using the so-called **separation theorems** of convex sets
- Let S and T be nonempty subsets of $\mathbb{R}^n$. Then, the hyperplane $H = \{x|a^T x = b\}$ is said to **separate** S and T if S is contained in one of the **closed** half spaces generated bt H and T is contained in the opposite **closed** half space.

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

- Such hyperplane is called a **separating hyperplane**
- A hyperplane **strictly separates** S and T if S is contained in one of the **open** half spaces generated bt H and T is contained in the opposite **open** half space.

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

Lemma

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

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.*

Lemma

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

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.*

V.1 An application: Farkas Lemma

Lemma

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

Proof: The statement

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

is equivalent to saying that

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

The second system has no solution if and only if the nonempty convex sets

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

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

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

V.1 Farkas Lemma (cont.)

Now:

- Letting $\rho = 0$, we conclude that $0 > \alpha$, so $c^T b = b^T c < 0$.
- The inequality $c^T A^T \geq 0$ holds, since for a certain k, $(c^T A^T)_k < 0$, and choosing $\rho = (0, ..., 0, \rho_k, 0, ..., 0)$ with $\rho_k \to +\infty$, we have that $c^T A^T \rho \to -\infty$, in contradiction with en $c^T A^T \rho \ > \alpha$.

So, we have found c such that $b^T c < 0$ and, furhermore

$$c^T A^T \geq 0 \ \Leftrightarrow \ Ac \geq 0,$$

so, we have found a solution of

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

Conversely, if $A^T \rho = b$, $\rho \geq 0$, and $Ay \geq 0$, then $b^T y = \rho^T Ay \geq 0$. $\square$

This lemma can be used to prove if a system like:

$$\left.\begin{array}{rcl} Ax & = & b, \\ x & \geq & 0, \end{array}\right\}$$

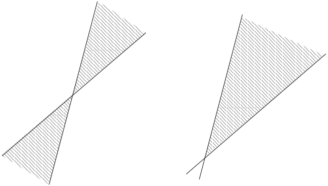
has or not solution

V.1 Convex polyhedra

- Let C be a convex set and $u \in C$, we say that 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 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

$$u \in S \;\Rightarrow\; \lambda u \in S, \;\; \forall \lambda \geq 0.$$

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



V.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.

- **Exercise 7:** Prove that the number of faces of dimension p of a n-dimensional simplex is equal to

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

- The **unitary simplex** are defined by the equations

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

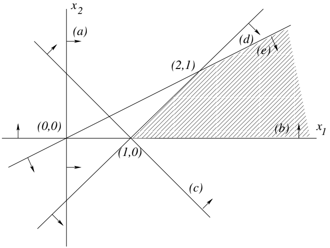
- If $n=3$ we get the tetrahedron with vertices at: $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(0,0,1)$

V.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



V.1 Convex sets and linear inequalities

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

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

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

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

V.1 Convex sets and linear inequalities

- In general, the set of points $(x_1, ..., x_n)$ 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 $\mathbb{R}^n$.

- The above set of inequalities can be also be written as a set of equalities.
- To do so, we must substract of each inequality an **unknown non-negative quantity**. These quatities 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

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

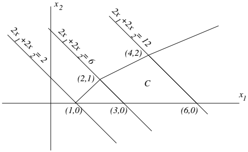
V.1 Convex sets and linear programming problems

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



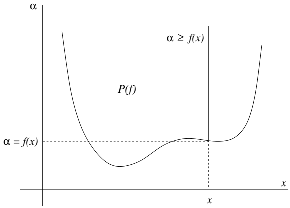
The figure displays the lines $2x_1 + 2x_2 = b$ for $b = 2, 6, 12$ The intersections of these lines with C is a point for $b = 2$. Of all the feasible points of C only one minimizes the objective function which, furthermore, is an extremal point of C.

V.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, $f(x)$ is either a real number or it is $\pm\infty$). The subset of $\mathbb{R}^{n+1}$

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

is called the **epigraf** of f.



Definition: We define f to be a **convex function** if $P(f)$ is a convex set.
We will see that this definition implies the one that we have already given:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2),\ \forall x_1, x_2 \in D,\ 0 \leq \lambda \leq 1.$$

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

V.2 Convex functions

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

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

  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 $a \in \mathbb{R}$, $b \in \mathbb{R}^n$. Then

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

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

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

$$DE(f) = \{x \in \mathbb{R}^n \mid f(x) < +\infty\},$$

  is called the **effective domain**, DE, of a function f. Note that $DE(f)$ is the projection of $P(f)$ on $\mathbb{R}^n$.

- If f is convex, then $DE(f)$ is also a convex set.
- **Exercise 8:** Prove that the converse of the above statement generally does not hold.

V.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. $f(x) > -\infty$,
3. $DE(f) \neq \emptyset$.

Theorem

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

$$f(q_1 x^1 + ... + q_s x^s) \leq q_1 f(x^1) + ... + q_s f(x^s). \tag{1}$$

Proof:

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

$$(q_1 x^1 + ... + q_s x^s, q_1 \alpha^1 + ... + q_s \alpha^s) \in P(f)$$

  this is

$$f(q_1 x^1 + ... + q_s x^s) \leq q_1 \alpha^1 + ... + q_s \alpha^s.$$

- Taking $\alpha^i = f(x^i)$ for $i = 1, ..., n$ the inequality follows.

$\square$

Theorem

Let f be a convex function and $\lambda \in \mathbb{R}$, $\lambda \geq 0$. Then λf is also convex. If f and g are convex functions, then $f + g$ 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.

V.2 Convex functions

Let Ψ be a function defined on $\mathbb{R}$ with values in the extended reals, it is said to be **nondecreasing** 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 on $\mathbb{R}^n$ and let Ψ be a nondecresing proper convex function on $\mathbb{R}$. Then $\Psi(f(x))$ 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

$$f(q_1 x^1 + q_2 x^2) \leq q_1 f(x^1) + q_2 f(x^2).$$

Since Ψ is nondecresing

$$\Psi(f(q_1 x^1 + q_2 x^2)) \leq \Psi(q_1 f(x^1) + q_2 f(x^2))$$

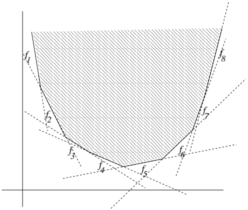
and by the convexity of Ψ

$$\Psi(f(q_1 x^1 + q_2 x^2)) \leq q_1 \Psi(f(x^1)) + q_2 \Psi(f(x^2)).$$

$\square$

V.2 Convex functions

Picewise linear functions, as the ones displayed in the figure, appear often in optimization. Its convexity can be proven by noting that a linear function is convex and by the following theorem



Theorem

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

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

The function f is convex.

V.2 Convex functions

Proof:

- The epigraphs of the 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

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

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

so the function f is convex. $\square$

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 $+\infty$.

V.2 Convex functions

Theorem

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

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

Then f is a convex function.

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

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

- Let (x^1, α^1), $(x^2, \alpha^2) \in P(f)$. From the definition of $P(f)$, it follows that there exist $\overline{\alpha}^1 \leq \alpha^1$, $\overline{\alpha}^2 \leq \alpha^2$ such that $(x^1, \overline{\alpha}^1)$, $(x^2, \overline{\alpha}^2) \in C$. Since C is convex
$$(q_1 x^1 + q_2 x^2, q_1 \overline{\alpha}^1 + q_2 \overline{\alpha}^2) \in C.$$
- According to the definition
$$f(q_1 x^1 + q_2 x^2) = \inf\{\alpha \in \mathbb{R} \mid (q_1 x^1 + q_2 x^2, \alpha) \in C\} \leq q_1 \overline{\alpha}^1 + q_2 \overline{\alpha}^2,$$
since $(q_1 x^1 + q_2 x^2, q_1 \overline{\alpha}^1 + q_2 \overline{\alpha}^2) \in C$ and $f(q_1 x^1 + q_2 x^2)$ is the infimum α such that $(q_1 x^1 + q_2 x^2, \alpha) \in C$. So
$$f(q_1 x^1 + q_2 x^2) \leq q_1 \overline{\alpha}^1 + q_2 \overline{\alpha}^2 \leq q_1 \alpha^1 + q_2 \alpha^2.$$

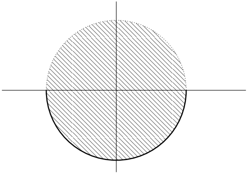
Thus

$$(q_1 x^1 + q_2 x^2, q_1 \alpha^1 + q_2 \alpha^2) \in P(f).$$

V.2 Convex functions. Example

Example: Let C be the open unit circle

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



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

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

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 x^1, $x^2 \in \mathbb{R}^n$, the function ϕ defined by

$$\phi(\lambda) = f(\lambda x^1 + (1-\lambda)x^2),$$

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

Proof. Suppose that f is convex on $\mathbb{R}^n$ and let x^1, $x^2 \in \mathbb{R}^n$ 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), $(\lambda^2, \alpha^2) \in P(\phi)$ and

$$z^1 = \lambda^1 x^1 + (1-\lambda^1)x^2, \quad z^2 = \lambda^2 x^1 + (1-\lambda^2)x^2.$$

We have then

$$f(z^1) = \phi(\lambda^1) \leq \alpha^1, \quad f(z^2) = \phi(\lambda^2) \leq \alpha^2.$$

Hence, (z^1, α^1), $(z^2, \alpha^2) \in P(f)$, and since $P(f)$ is convex we also have that $(q_1 z^1 + q_2 z^2, q_1\alpha^1 + q_2\alpha^2) \in P(f)$ for every q_1, q_2 such that $q_i \geq 0$, $q_1 + q_2 = 1$. Consequently

$$f(q_1 z^1 + q_2 z^2) \leq q_1\alpha^1 + q_2\alpha^2.$$

V.2 Convex functions (cont.)

According to the definitions

$$\begin{aligned}
\phi(q_1\lambda^1 + q_2\lambda^2) &= f\left[(q_1\lambda^1 + q_2\lambda^2)x^1 + (1 - q_1\lambda^1 - q_2\lambda^2)x^2\right] \\
&= f\left[(q_1\lambda^1 + q_2\lambda^2)x^1 + (q_1 + q_2 - q_1\lambda^1 - q_2\lambda^2)x^2\right] \\
&= f\left[(q_1[\lambda^1 x^1 + (1-\lambda^1)x^2] + q_2[\lambda^2 x^1 + (1-\lambda^2)x^2]\right] \\
&= f(q_1 z^1 + q_2 z^2).
\end{aligned}$$

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. □

V.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 $x^0 \in \mathbb{R}^n$, consider the collection of all the linear functions h of the form $h(x) = a^T x - b$ such that $h(x^0) \leq f(x^0)$.
- Define the **support set** by

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

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

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

  Clearly, $\mathrm{cl}f(x) \leq f(x)$ for all $x \in \mathbb{R}^n$.
- A convex function f is sdaid to be **closed** if $\mathrm{cl}f = f$.
- It can be shown that a proper convex function f is closed if and only if the convex level set $\{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$ is closed for every real number $\alpha \in \mathbb{R}$.

V.2 Continuity of convex functions

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

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

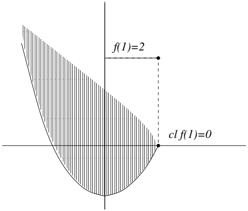
- This last relation can also be used as the definition of the closure operation for proper convex functions.
- For improper convex functions, the closure operation has a very simple meaning: if $f(x) = +\infty$ for every $x \in \mathbb{R}^n$, then $\mathrm{cl} f = f$.
- If, however, $f(x) = -\infty$ for some $x \in \mathbb{R}^n$, then the support set $L(f) = \emptyset$, and since, by convention, the supremum over the empty set is $-\infty$, we get $\mathrm{cl} f(x) = -\infty$ for all x.
- Consequently, the only closed improper convex functions are those that are identically $f(x) = \pm\infty$.

V.2 Continuity of convex functions

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 $DE(f)$.

V.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 $\text{cl}f$ 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 x^1, $x^2 \in B$ and for every $\alpha \in \mathbb{R}$, then $\alpha x^1 + (1 - \alpha)x^2 \in B$. 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.

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.

V.2 Continuity of convex functions

The proofs of the next two theorems require additional concepts from topology (and is not given). 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 $f(x) = -\infty$ for every x in the relative interior of its effective domain.

Theorem

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

From this theorem follows

Corollary

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

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

Theorem

Let $f_1,\dots,f_m$ be proper convex functions and let C be a nonempty convex set such that $C \subset \cap_{i=1}^{m} DE(f_i)$.
Then, exactly one of the alternatives holds:

1. *There exists an $x^0 \in C$ such that*
$$f_i(x^0) < 0, \quad i = 1, \dots, m$$
2. *There exist nonnegative numbers $\alpha_1,\dots,\alpha_m$ (not all zero) such that for every $x \in C$*
$$\sum_{i=1}^{m} \alpha_i f_i(x) \geq 0.$$

V.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 $C \subset \cap_{i=1}^{m} DE(f_i)$.
If the system of inequalities

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

has no solution $x \in C$, while there exists an $x^0 \in C$ such that

$$f_i(x^0) < 0, \quad i = 1, \ldots, m$$

then, either

$$f_0(x) \geq 0,$$

for every $x \in C$, or there exist nonnegative numbers $\alpha_1, \ldots, \alpha_m$ not all zero, such that

$$f_0(x) + \sum_{i=1}^{m} \alpha_i f_i(x) \geq 0,$$

for every $x \in C$.

V.2 Differential properties of convex functions

- Given $f : S \longrightarrow \mathbb{R}$ with $S \subset \mathbb{R}^n$, a point x^0 in the interior of S and any vector $y \in \mathbb{R}^n$, the derivative of f at x^0 in the direction of y is defined as

$$Df(x^0; y) = \lim_{t \to 0} \frac{f(x^0 + ty) - f(x^0)}{t}.$$

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

$$D^+f(x^0; y) = D^-f(x^0; y)$$

then

$$D^+f(x^0; y) = D^-f(x^0; y) = Df(x^0; y)$$

V.2 Differential properties of convex functions

- If y is a vector of tha canonical basis, then the directional derivatives are the partial derivatives.
- If f is differentiable at x^0, then the directional derivatives of f at x^0 in all directions y are finite and are given by

$$Df(x^0; y) = y^T \nabla f(x^0) = \nabla f(x^0)^T y.$$

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

$$f(tx) = t^k f(x).$$

Theorem

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

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

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

V.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 and to the directional derivatives in the more general case.

Definition

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

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

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

For a convex function f it is possible that, at some point x

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

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

V.2 Differential properties of convex functions

Some basic properties of subgradients are:

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

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

Theorem

A vector $\xi \in \mathbb{R}^n$ is a subgradient of a convex function f at a point x where $f(x)$ is finite if and only if

$$D^+ f(x; z) \geq \xi^T z, \tag{3}$$

for every direction z.

V.2 Proof of the theorem

Proof: If ξ is a subgradient of f at x, then it satisfies

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

Let $y = x + tz$, with $t > 0$, then

$$f(x + tz) \geq f(x) + t\xi^T z,$$

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

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

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

Conversely, if $D^+ f(x; z) \geq \xi^T z$ holds for every $z \in \mathbb{R}^n$, then $f(x + tz) \geq f(x) + t\xi^T z$ also holds by the same argument as above and, consequently, $f(y) \geq f(x) + \xi^T(y - x)$ holds. This, in turn, implies that ξ is a subgradient of f at x. $\square$

Corollary

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

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

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

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

Proof:

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

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

$\square$

V.2 Corollaries

Corollary

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

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

In particular, if f is differentiable at x and y, then

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

Proof: By the preceeding corollary

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

Thus

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

and using a previous theorem we get

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

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

V.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.

V.2 Differential properties of convex functions

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.

If f is convex, using the homogeneity of D^- and Corollary 3, 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

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

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

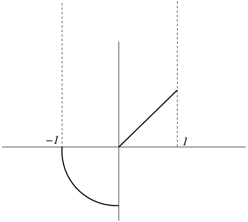
$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \geq D^+ f(x_1;1).$$

From these two inequalities follows the central one of the Theorem. $\square$

V.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

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

V.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 $-1 \leq x < 0$ 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

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

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

$$D^+f(x;1) \geq \xi \geq D^-f(x;1).$$

V.2 Example (cont.)

Consequently

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

V.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 nondecreasing function on D if and only if f is convex on D.

Proof: If f is convex, the results follows from the preceeding Theorem. 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 $q_1 + q_2 = 1$. 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 nondecreasing on $\mathbb{R}$, we have

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

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

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

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

V.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

$$f(x_2) \geq f(x_1) + (x_2 - x_1)^T \nabla f(x_1),$$

for every two points $x_1, x_2 \in \mathbb{R}^n$, then f is convex on $\mathbb{R}^n$.

Proof: Let $x_1, x_2 \in \mathbb{R}^n$ be any two points and $x_3 = q_1 x_1 + q_2 x_2$. Then

$$\begin{aligned} f(x_1) &\geq f(x_3) + q_2 (x_1 - x_2)^T \nabla f(x_3), && (7) \\ f(x_2) &\geq f(x_3) + q_1 (x_2 - x_1)^T \nabla f(x_3). && (8) \end{aligned}$$

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

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

so f is convex. $\square$

V.2 Some additional differential properties of convex functions

Corollary 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 x_0, that is, the linear function

$$f(x_0) + (x - x_0)^T \nabla f(x_0),$$

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

V.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 nondecreasing, that is, $f''(x) \geq 0$ for every $x \in D$. □

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 $x \in C$ is positive semidefinite. That is, for each $x \in C$

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

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"*.

V.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 x^* is a local minimum, then

$$f(x) \geq f(x^*),$$

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

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

Since f is a convex function

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

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

$$f(z) \geq f(x^*).$$

□

V.3 Extrema of convex functions

Consider the general problem

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

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, we can define a new objective function

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

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

V.3 Extrema of convex functions

Note that generalyy 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

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

are convex sets for any α.

Proof: Let x_1, $x_2 \in S(f, \alpha)$. It follows that for any q_1, q_2 such that $q_i \geq 0$, $q_1 + q_2 = 1$:

$$f(q_1 x_1 + q_2 x_2) \leq q_1 f(x_1) + q_2 f(x_2) \leq q_1 \alpha + q_2 \alpha = \alpha.$$

Hence, $S(f, \alpha)$ is convex. $\square$

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 $\{x \in \mathbb{R}^n \mid f(x) \leq \alpha^*\}$ is the set of points at which f attains its minimum; by the preceding Lemma, this is a convex set. $\square$

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, it is attained at a unique point of X.

Proof: Suppose that the minimum is attained at two distinct points , x_1, $x_2 \in X$ and let $f(x_1) = f(x_2) = \alpha$. From the preceding Theorem, it follows that for every q_1, q_2 ($q_i \geq 0$, $q_1 + q_2 = 1$), we have $f(q_1 x_1 + q_2 x_2) = \alpha$, contradicting that f is strictly convex. $\square$

V.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 $0 \in \partial f(x^*)$ if and only if f attains its minimum at x^*.

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

$$f(y) \geq f(x^*)$$

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

Corollary

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

$$\nabla f(x^*) = 0$$

if and only if f attains its minimum at x^*.

Proof: $\partial f(x) = \{\nabla f(x)\}$ if and only if f is differentiable at 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 $x^* \in X$.

V.4 Optimality conditions for convex problems

Consider the following nonlinear problem

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

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 $x \in \mathbb{R}^n$ satisfying the constraints is a convex set.

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

V.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, $g_1,\ldots,g_m$ are real-valued, differentiable, convex and concave functions on $\mathbb{R}^n$, respectively, and let $h_1,\ldots,h_p$ be linear. If there exist x^*, λ^*, μ^*, with x^* satisfying (10) and (11), together with

$$\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = 0, \tag{12}$$

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, \ldots, m, \tag{13}$$

$$\lambda^* \geq 0,$$

then $x*$ is a global optimum of the convex minimization problem.

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

Proof: Let x be any point satisfying (10) and (11). Then

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

Applying Corollary 3 to f, g_i and h_j and using (14), we obtain

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

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

$$f(x) \geq f(x^*).$$

$\square$

V.4 Optimality conditions for convex problems

A nonlinear programming problem is said to be **strongly consistent** if there exists a point $x_0 \in \mathbb{R}^n$ satisfying

$$g_i(x_0) > 0, \quad i = 1, \ldots, m,$$

$$h_j(x_0) = 0, \quad j = 1, \ldots, p.$$

We will also require that the vectors of coefficients $a_j = (a_{j1}, \ldots, a_{jn})$ in the linear functions

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

to be linearly independent.

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

V.4 Optimality conditions for convex problems

Theorem

Suppose that the functions f, $g_1,\ldots,g_m$ 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 $a^j = (a_{j1}, \ldots, a_{jn})$, $j = 1, \ldots, p$ are linearly independent. Suppose thet the nonlinear programming problem is strongly consistent. If x^* is a solution of the problem, then there exist vectors $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ i $\mu^* = (\mu_1^*, \ldots, \mu_p^*)$ such that

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

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, \ldots, m,$$

$$\lambda^* \geq 0,$$