Chapter 6

A brief introduction to find zeros of functions

Definition of convex sets in $\mathbb{R}^n$

Definition. We say that $\mathcal{C} \subset \mathbb{R}^n$ is a **convex set** if for any two points $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathcal{C}$ we have

$$\alpha \boldsymbol{u}_1 + (1-\alpha)\boldsymbol{u}_2 \in \mathcal{C}, \;\; \alpha \in [0,1]$$

Exercise. Any (open and closed) hyper-cube and any (open and closed) hyper-ball in $\mathbb{R}^n$ are convex sets.

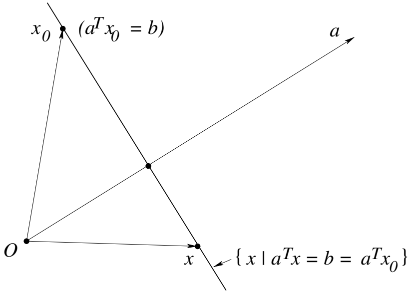
$$Q := \{\boldsymbol{x} \in \mathbb{R}^n \mid \; 0 < |x_j| \; < 1, \; j = 1, \ldots n\}$$
$$B := \{\boldsymbol{x} \in \mathbb{R}^n \mid ||x|| < 1\}$$

Hyperplanes

Definition. Let $\boldsymbol{a} \in \mathbb{R}^n \setminus \{\vec{0}\}$, and let $b \in \mathbb{R}$. The set

$$\mathbb{H}_a := \mathbb{H} = \left\{ \boldsymbol{x} \,|\, \boldsymbol{a}^T \boldsymbol{x} = b \right\}$$

is called a **hyperplane** of $\mathbb{R}^n$. Alternatively, $\mathbb{H}$ is the set of all all the vectors $\boldsymbol{x} \in \mathbb{R}^n$ such that its scalar product with $\boldsymbol{a} \in \mathbb{R}^n \setminus \{\vec{0}\}$ is constant.



Hyperplanes

Exercise. If $\boldsymbol{x}_0$ i $\boldsymbol{x}_1$ are two points in $\mathbb{H}_a$, then

$$\boldsymbol{a}^T(\boldsymbol{x}_1 - \boldsymbol{x}_0) = 0$$

Definition. The vector $\boldsymbol{a}$ is called the **normal vector** of $\mathbb{H}_a$.

Exercise. The set $\mathbb{H}_a$ is a convex set. Moreover the set $\mathbb{H}_a$ defines two convex open half-spaces and two closed half-spaces given by

$$\mathbb{H}_a^+ = \{\boldsymbol{x} \mid \boldsymbol{a}^T x > b\},\ \mathbb{H}_a^- = \{\boldsymbol{x} \mid \boldsymbol{a}^T x < b\}, \quad \text{and}$$
$$\overline{\mathbb{H}}_a^+ = \{\boldsymbol{x} \mid \boldsymbol{a}^T x \geq b\},\ \overline{\mathbb{H}}_a^- = \{\boldsymbol{x} \mid \boldsymbol{a}^T x \leq b\}$$

The convex hull

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

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

Corollary. For any given $G \subset \mathbb{R}^n$, the set $\mathcal{C}(G)$ is a convex set.

Exercise. Compute $\mathcal{C}(G)$ for

$$G = \{\cup_{n=1}^{3}(x_n, y_n)\} \subset \mathbb{R}^2.$$

Separating hyperplanes

Let G_1 and G_2 be nonempty subsets of $\mathbb{R}^n$.

Definition. We say that $\mathbb{H}_a$ separates G_1 from/and G_2 if

$$G_1 \subset \left\{ \boldsymbol{x} \mid \boldsymbol{a}^T \boldsymbol{x} \geq b \right\} \quad \text{and} \quad G_2 \subset \left\{ \boldsymbol{x} \mid \boldsymbol{a}^T \boldsymbol{x} \leq b \right\}$$

The set $\mathbb{H}_a$ **strictly separates G_1 and G_2** if the inequalities are strict.

Theorem (separation theorem). Let $G_j \in \mathbb{R}^n$, $j = 1, 2$ be two **disjoint** nonempty convex sets. Then there exists a hyperplane that separates them. Moreover, if we assume that C_2 is compact then there exists a hyperplane that strictly separates them.

Farkas Lemma

Theorem (Farkas' Lemma). Let A be an $m \times n$ real matrix and let $\boldsymbol{b} \in \mathbb{R}^n$. 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 $\boldsymbol{\rho} \in \mathbb{R}^m$ with $\boldsymbol{\rho} \geq 0$, such that $A^T \boldsymbol{\rho} = \boldsymbol{b}$

Proof. The statement is equivalent to

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

$\Leftarrow$) Then, the nonempty convex sets

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

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

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

Farkas Lemma

Proof (continue). This is

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

(a) Claim: $\boldsymbol{c}^T\boldsymbol{b} = \boldsymbol{b}^T\boldsymbol{c} < 0$. To see this claim, take $\boldsymbol{\rho} = 0$ above. Then $\alpha < 0$.

(b) Claim: $\boldsymbol{c}^T A^T \geq 0$. To this this claim notice that 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 $\boldsymbol{c}^T A^T \boldsymbol{\rho} > \alpha$.

Accordingly the vector $\boldsymbol{c}$ is a solution of

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

as desired.

Farkas Lemma

Proof (continue).

$\Rightarrow$) We should prove that

$$\left.\begin{array}{rcc} A\boldsymbol{y} & \geq & 0 \\ \boldsymbol{b}^T\boldsymbol{y} & < & 0 \end{array}\right\} \text{has a solution implies } \left.\begin{array}{rcc} A^T\boldsymbol{\rho} & = & \boldsymbol{b} \\ \boldsymbol{\rho} & \geq & 0 \end{array}\right\} \text{has no solution}$$

(We prove the negative version.) Assume 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

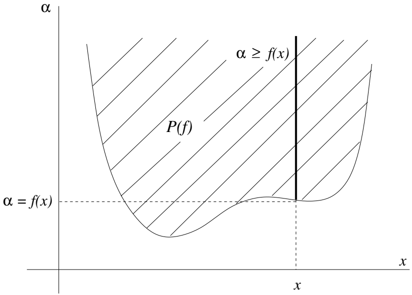
$$\left.\begin{array}{rcc} A\boldsymbol{y} & \geq & 0 \\ \boldsymbol{b}^T\boldsymbol{y} & < & 0 \end{array}\right\} \text{has no solution.}$$

Convex functions: The epigraf of f.

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

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

is called the epigraf of f. We say f is a convex function if $P(f)$ is a convex set.



Convex functions: The epigraf of f.

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 $f|_D$ is identical to the one of $f_1|_{\mathbb{R}^n}$. Hence we can always extend a convex function f (over D), to be a convex function defined throughout all $\mathbb{R}^n$.

Remark. 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 (not continuous) function defined over all $\mathbb{R}^n$.

Convex functions: The effective domain of f.

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}$. The effective domain of f is the set

$$\mathrm{ED}(f) = \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) < +\infty\}$$

Exercises.

(a) Show that $\mathrm{ED}(f)$ is the projection of $P(f)$ over $\mathbb{R}^n$ (the first component).

(b) If f is a convex function, then $\mathrm{ED}(f)$ is a convex set.

(c) Show that the converse (of statement (b)) is not necessarily true.

Definition. We say that f is a proper convex function if f is convex, $f(\boldsymbol{x}) > -\infty$ for every $\boldsymbol{x}$, and $\mathrm{ED}(f) \neq \emptyset$.

An equivalent definition for convexity

Theorem. Let $q_1, \ldots, q_s \in \mathbb{R}$ with $q_j \geq 0$, $j = 1, \ldots, s$ and $\sum_{j=1}^{s} q_j = 1$.

Then, f is a (proper) convex function on $\mathbb{R}^n$ if and only if for all $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s \in \mathbb{R}^n$ we have

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

Proof ($\Rightarrow$)).

(a) If $f(\boldsymbol{x}_j) = +\infty$ for some $j = 1, \ldots, s$, then (1) trivially holds.

(b) Assume now that $f(\boldsymbol{x}_j) < +\infty$ for all $j = 1, \ldots, s$. Since f is convex, then $P(f)$ is a convex set. That is,

$$(\boldsymbol{x}_1, \alpha_1) \in P(f), \ldots, (\boldsymbol{x}_s, \alpha_s) \in P(f) \;\Rightarrow\; (q_1\boldsymbol{x}_1 + ... + q_s\boldsymbol{x}_s, q_1\alpha_1 + ... + q_s\alpha_s) \in P(f).$$

This is to say that

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

(c) Since $(\boldsymbol{x}_i, \alpha_i) \in P(f) \;\Rightarrow\; f(\boldsymbol{x}_i) \leq \alpha_i$, we can take $\alpha_i = f(\boldsymbol{x}_i)$, for $i = 1, ..., n$, and (1) follows.

Linear combinations of convex functions

Lemma. Let f and g be convex functions. Let $\lambda \in \mathbb{R}_+$. Then the functions λf and $f + g$ are also convex functions (provided that the operation $+\infty + (-\infty)$ is avoided).

In particular, every linear combination $\lambda_1 f_1 + \cdots + \lambda_k f_k$ of convex functions with $\lambda_j \geq 0$ for all $j = 1, \ldots, k$ is also a convex function.

Exercise. Prove the above statements.

Composition and convex functions

Definition. Let $\Psi : \mathbb{R} \to \overline{\mathbb{R}}$ be a function defined on $\mathbb{R}$ with values in the extended reals. We say that Ψ is non-decreasing if for every $x_1 < x_2$ we have $\Psi(x_1) \leq \Psi(x_2)$.

Theorem. Let f be a real convex function defined on $\mathbb{R}^n$, and let Ψ be a non-decreasing proper convex function defined on $\mathbb{R}$. Then $\Psi \circ f$ is convex on $\mathbb{R}^n$.

Proof. Since f is convex and Ψ is non-decresing we have $(0 \leq q_1 \leq 1)$

$$f\left(q_1 x_1 + (1-q_1)x_2\right) \leq q_1 f(x_1) + (1-q_1)f(x_2), \text{ and}$$
$$\Psi\left(f\left(q_1 x_1 + (1-q_1)x_2\right)\right) \leq \Psi\left(q_1 f(x_1) + (1-q_1)f(x_2)\right)$$

Finally by the convexity of Ψ we have

$$\Psi\left(f(q_1 x_1 + (1-q_1)x_2)\right) \leq \Psi\left(q_1 f(x_1) + (1-q_1)f(x_2)\right) \leq q_1 \Psi\left(f\left(x_1\right)\right) + (1-q_1)\Psi\left(f\left(x_2\right)\right).$$

The maximum of convex functions

Theorem. Let $f_j,\ j = 1, \dots m$ be a finite collection of convex functions on $\mathbb{R}^n$. Then the function

$$F(x) := \max_j f_j(x)$$

is a convex function (i.e., $P(F)$ is a convex set).

Proof. The sets $P(f_j),\ j = 1, \dots m$ (epigrafs) are convex sets and so their intersection is convex as well. By definition

$$\bigcap_j P(f_j) = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid \max_j f_j(x) \le \alpha,\ \text{for all } j = 1, \dots m\} =$$

$$= \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid \max_j f_j(x) = F(x) \le \alpha\} = P(F).$$

Two important results

Theorem A. A real valued 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 $\phi : [0, 1] \to \mathbb{R}$ defined by

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

is convex.

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

Convex differentiable functions

Definition. Let $D \in \mathbb{R}^n$ an open set and let $\boldsymbol{x}_0 \in D$. Let $f : D \to \mathbb{R}$. Let $\boldsymbol{v} \in \mathbb{R}^n$ a unitary vector. We define the $\boldsymbol{v}$-directional derivative of f at the point $\boldsymbol{x}_0$ by

$$Df(\boldsymbol{x}_0, \boldsymbol{v}) := \lim_{t \to 0} \frac{f(\boldsymbol{x}_0 + t\boldsymbol{v}) - f(\boldsymbol{x}_0)}{t}.$$

When we consider the above limit with $t \to 0^+$ and $t \to 0^-$ we denote them by $D^+ f(\boldsymbol{x}_0, \boldsymbol{v})$ and $D^- f(\boldsymbol{x}_0, \boldsymbol{v})$ and we called them right-sided (left-sided) $\boldsymbol{v}$-directional derivative of f at the point $\boldsymbol{x}_0$, respectively.

Remark. According to previous notation and results we have

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

Convex differentiable functions

Definition. A function f is said to be positively homogeneous of degree $k \geq 1$ if for every $\boldsymbol{x} \in \mathbb{R}^n$ and every $t \in \mathbb{R}^+$ we have

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

Theorem. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex (finite) function. Then

(a) For any unitary $\boldsymbol{v} \in \mathbb{R}^n$ there exist the right-sided and left-sided derivatives of f at every $\boldsymbol{x}$.

(b) $D^+ f$ and $D^- f$ are positively homogeneous convex functions of $\boldsymbol{v}$ of degree one; i.e., $D^{\pm} f(\boldsymbol{x}, \lambda\boldsymbol{v}) = \lambda D^{\pm} f(\boldsymbol{x}, \boldsymbol{v})$.

(c) The following inequality holds:

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

Convex differentiable functions: Subgradients

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. A subgradient of f at a point $\boldsymbol{x}_0 \in \mathbb{R}^n$, is a vector $\boldsymbol{\xi} \in \mathbb{R}^n$ such that

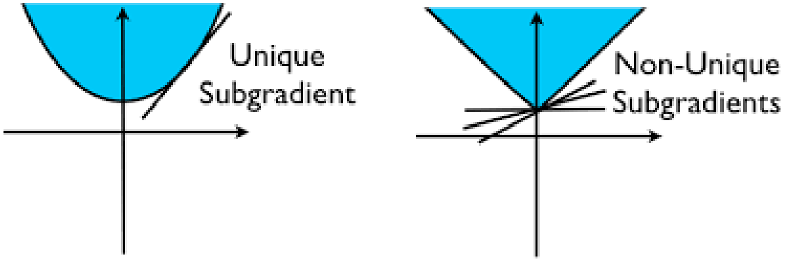
$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}_0) + \boldsymbol{\xi}^T(\boldsymbol{y} - \boldsymbol{x}) \tag{2}$$

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

Remark. A subgradient of a convex function $f : \mathbb{R}^n \to \mathbb{R}$ at a $x_0 \in \mathbb{R}^n$ may be a unique vector or several (infinitely many) vectors.

Notation and definition. We denote by $\partial f(\boldsymbol{x})$ the set of all subgradients of a convex function f at a given point $\boldsymbol{x}$. In some books $\partial f(\boldsymbol{x})$ is called subdifferential.

Convex differentiable functions: Subgradients



Theorem. Let f be a convex function. A vector $\boldsymbol{\xi} \in \partial f(\boldsymbol{x})$ if and only if

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

Convex differentiable functions: Subgradients

Proof. If $\boldsymbol{\xi} \in \partial f(\boldsymbol{x})$, then it satisfies $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \boldsymbol{\xi}^T(\boldsymbol{y} - \boldsymbol{x})$ for all $\boldsymbol{y} \in \mathbb{R}^n$. If we write $\boldsymbol{y} = \boldsymbol{x} + t\boldsymbol{z}$, with $t > 0$, then the previous inequality writes as

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

for every $\boldsymbol{z} \in \mathbb{R}^n$ and $t > 0$. We deduce from above that $D^+ f(\boldsymbol{x}; \boldsymbol{z}) \geq \boldsymbol{\xi}^T\boldsymbol{z}$ since $D^+ f(\boldsymbol{x}; \boldsymbol{z})$ is the right-sidded limit of the incremental quotients $(t > 0)$.

The other implication follows similarly.

Convex differentiable functions: Subgradients

Lemma. Let f be a convex function on $\mathbb{R}^n$. 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. (We use the notion of inf).

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

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

where the inequality follows from f being convex.

Convex differentiable functions: Subgradients

Remark. From Theorem A (above) we may study the convexity of a function f in $\mathbb{R}^n$ by studding the convexity of its restriction to any line segment in $\mathbb{R}^n$.

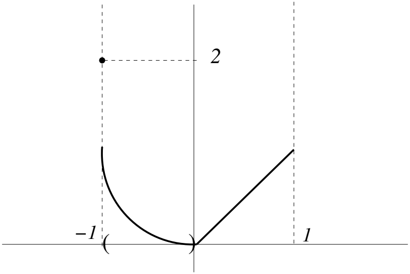
So, in some cases it is sufficient to study the behaviour of convex functions on $\mathbb{R}$. In particular because of the homogeneity of $D^{\pm}f$ with respect to v it is enough to consider $D^{\pm}f(x, 1)$.

Proposition. 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

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

An example

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



Using the definitions we can compute

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

Final comments on differentiable convex functions

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 $\mathbb{R}^n$.

Final comments on differentiable convex functions

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

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

Optimality of convex functions

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

Proof. We have $f(x) \geq f(x^\star)$ for all $x \in B(\boldsymbol{x}, \varepsilon)$. Let $z \in \mathbb{R}^n$. Then

$$((1-\lambda)x^\star + \lambda z) \in B(\boldsymbol{x}, \varepsilon)$$

if $0 < \lambda < 1$ is small enough. Moreover for those small λ's we have

$$\begin{aligned} & f((1-\lambda)x^\star + \lambda z) \geq f(x^\star) && (x^\star \text{ is local minimum}) \\ & (1-\lambda)f(x^\star) + \lambda f(z) \geq f((1-\lambda)x^\star + \lambda z) && (f \text{convex}) \end{aligned}$$

Direct computations give

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

Optimality of convex functions

Theorem. Let f be a convex function on $\mathbb{R}^n$ and let α be a real number. Then, the sets

$$S\left(f,\alpha\right) := \left\{\boldsymbol{x} \in \mathbb{R}^n \middle| \left| f\left(\boldsymbol{x}\right) \leq \alpha\right.\right\}$$

are convex sets for any α.

Proof. Let $x_j \in S\left(f,\alpha\right),\ j = 1,2$. Let $q_1 \in [0,1]$. We have

$$f\left(q_1 x_1 + (1-q_1)x_2\right) \leq q_1 f\left(x_1\right) + (1-q_1) f\left(x_2\right) \leq q_1\alpha + (1-q_1)\alpha = \alpha,$$

where the first inequality follows from convexity and the second from $x_j \in S\left(f,\alpha\right),\ j = 1,2$. So, $S\left(f,\alpha\right)$ is convex.

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

Optimality of convex functions

Lemma. Let f be a convex function on $\mathbb{R}^n$. Then, $0 \in \partial f(x^\star)$ if and only if f attains its minimum at $x^\star$.

Proof. By definition $0 \in \partial f(x^\star)$ if and only if $f(y) \geq f(x^\star)$ for all $y \in \mathbb{R}^n$.

Corollary. Let f be a convex differentiable function on $\mathbb{R}^n$. Then, $\nabla f(x^\star) = 0$ if and only if f attains its global minimum at $x^\star$.