Convex Optimization

Adrien Taylor

Class notes by Antoine Groudiev

Last modified 29th October 2024

Contents

Abstract

This document is Antoine Groudiev's class notes while following the class *Convex Optimization* (Optimisation Convexe) at the Computer Science Department of ENS Ulm. It is freely inspired by the lectures of Adrien Taylor.

1 Introduction

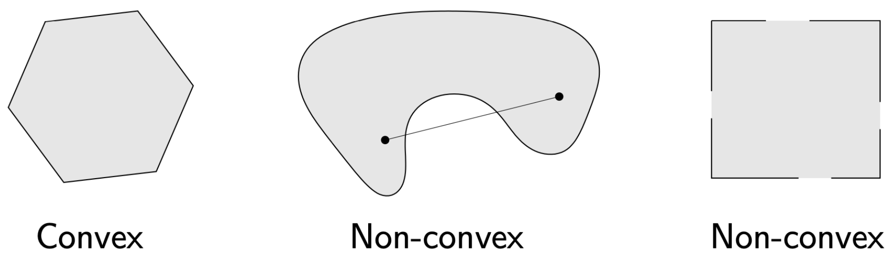
2 Convex sets

2.1 Definitions

Definition (Convex set)**.** A set C is a *convex set* if every segment that connects two points in C is in C. Formally:

$$\forall x, y \in C, \forall \theta \in [0,1], \quad \theta x + (1-\theta) y \in C$$

Example. Here are some examples of convex and non-convex sets:



In many cases, we will use proper (i.e. non-empty) convex sets, and closed convex sets.

Definition (Convex hull)**.** The *convex hull* of S, denoted $\mathrm{Conv}(S)$, is the smallest convex set that contains S.

Definition (Convex combinations)**.** The *convex combinations* of $x_1, \dots, x_k$ are all the point x of the form:

$$x = \theta_1 x_1 + \cdots + \theta_k x_k$$

with $\theta_1, \dots, \theta_k \geqslant 0$ and $\sum_{i=1}^k \theta_i = 1$.

Property 2.1. The convex hull of a set S is the set of all convex combinations of points in S:

$$\mathrm{Conv}(S) = \left\{ \sum_{i=1}^k \theta_i x_i \,\middle|\, (x_i) \in S^k, (\theta_i) \in \mathbb{R}_+^k, \sum_{i=1}^k \theta_i = 1 \right\}$$

2.2 Examples

2.2.1 Hyperplanes and halfspaces

Definition (Hyperplane)**.** A *hyperplane* is the set of the form:

$$H = \left\{ x \,\middle|\, a^\top x = b \right\}$$

for some $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$. a is called the *normal vector* of H. Hyperspaces are affine and convex.

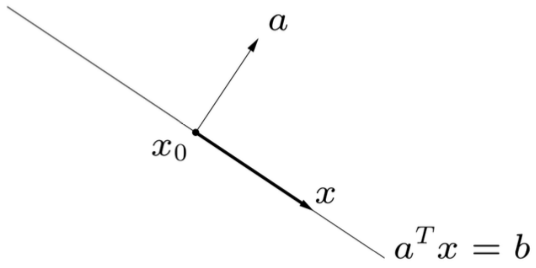


Figure 2.1: Hyperplane

Definition (Halfspace)**.** A *halfspace* is the set of the form:

$$H = \left\{ x \,\middle|\, a^\top x \leqslant b \right\}$$

for some $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$. a is called the *normal vector* of H. Halfspaces are convex.

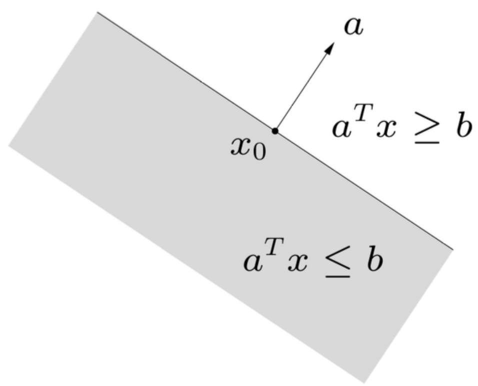


Figure 2.2: Halfspace

2.2.2 Euclidian balls and ellipsoids

Definition (Euclidian ball)**.** The *Euclidian ball* of center x_c and radius r is the set:

$$B(x_c, r) = \{ x \mid \|x - x_c\|_2 \leqslant r \} = \{ x_c + ru \mid \|u\|_2 \leqslant 1 \}$$

Euclidian balls are convex.

Definition (Ellipsoid)**.** An *ellipsoid* is the set of the form:

$$E = \left\{ x \,\middle|\, (x - x_c)^\top P^{-1} (x - x_c) \leqslant 1 \right\}$$

with $P \in \mathbb{S}^n_{++}$[1] and $x_c \in \mathbb{R}^n$. Ellipsoids are convex.

[1] $\mathbb{S}^n_{++}$ denotes the set of symmetric positive definite matrices of size n

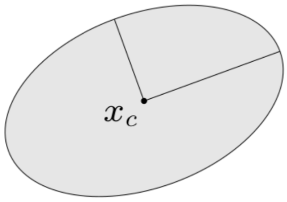


Figure 2.3: Ellipsoid

An alternative representation of an ellipsoid is:

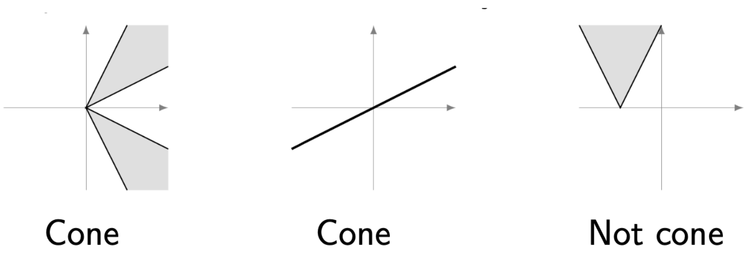
$$E = \{\, x_c + Au \mid \|u\|_2 \leqslant 1 \,\}$$

for some nonsingular matrix $A \in \mathrm{GL}_n(\mathbb{R})$. We can choose A symmatric and positive definite without loss of generality, for instance by choosing $A = P^{1/2}$.

2.2.3 Cones

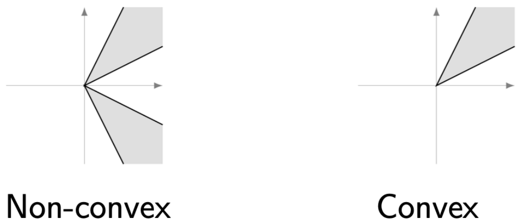
Definition (Cones)**.** A set K is a *cone*, or a *nonnegative homogeneous set*, if:

$$\forall x \in K, \forall \theta \in \mathbb{R}_+^*, \quad \theta x \in K$$



Definition (Convex cone)**.** A set K is a *convex cone* if:

$$\forall x_1, x_2 \in K, \forall \theta_1, \theta_2 \in \mathbb{R}_+^*, \quad \theta_1 x_1 + \theta_2 x_2 \in K$$



In the followings, we will denote by:

- $\mathbb{S}^n \subset \mathbb{R}^{n \times n}$ the set of symmetric matrices of size n
- $\mathbb{S}_+^n$ the set of positive semidefinite matrices of size n, that is matrices verifying:

  $$\forall z \in \mathbb{R}^n, z^\top X z \geqslant 0$$

  also denoted $X \succcurlyeq 0$.

- $\mathbb{S}^n_{++}$ the set of positive definite matrices of size n, that is matrices verifying:

$$\forall z \in \mathbb{R}^n, z^\top X z > 0$$

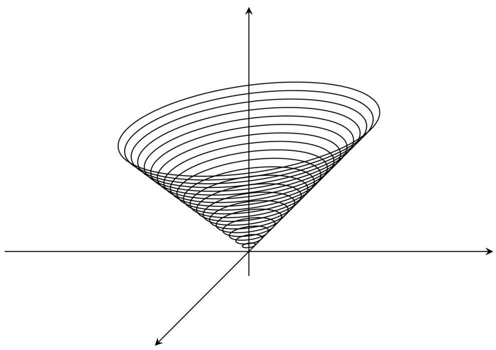
also denoted $X \succ 0$.

$\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ are convex cones.

Special cases of cones include:

Positive orthant $K = \mathbb{R}^n_+ = \{\, x \in \mathbb{R}^n \mid x_i \geqslant 0, \forall i \,\}$

Norm cones $K = \{\, (x,t) \in \mathbb{R}^n \times \mathbb{R} \mid \|x\| \leqslant t \,\}$. A particular case is the second-order cone (SOC), based on the ℓ_2 norm.



Positive polynomials $K_n = \{\, x \in \mathbb{R}^{n+1} \mid \forall t \in \mathbb{R}, \sum_{i=0}^n x_i t^i \geqslant 0 \,\}$

Positive semidefinite cone $\mathbb{S}^n_+ = \left\{ X \in \mathbb{S}^n \;\middle|\; \forall z \in \mathbb{R}^n, z^\top X z \geqslant 0 \right\}$

Co-positive cone $\mathbb{S}^n_+ = \left\{ X \in \mathbb{S}^n \;\middle|\; \forall z \in \mathbb{R}^n_+, z^\top X z \geqslant 0 \right\}$

Exponential cone $\left\{ (x,y,z) \in \mathbb{R} \times \mathbb{R}^*_+ \times \mathbb{R} \;\middle|\; z \geqslant y e^{x/y} \right\}$

Definition (Dual cones)**.** The *dual cone* to a convex cone K is the set:

$$K^* = \left\{ y \;\middle|\; \forall x \in K, \quad y^\top x \geqslant 0 \right\}$$

Convex cones and their duals are particularly useful for convex duality. A convex cone that satisfies $K = K^*$ is called *self-dual*.

Definition (Polar cones)**.** The *polar cone* to a convex cone K is the set:

$$K^\diamond = \left\{ y \;\middle|\; \forall x \in K, \quad y^\top x \leqslant 0 \right\}$$

We have the identity $K^\diamond = -K^*$.

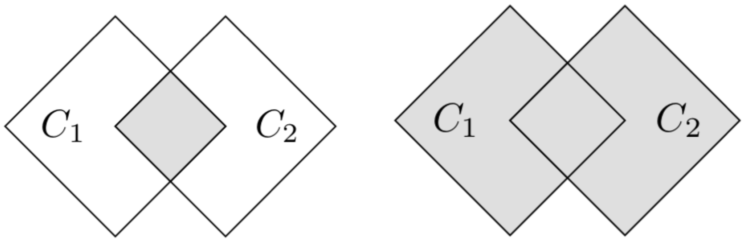
2.3 Convexity-preserving operations

To establish the convexity of a set C, the most basic approach is to apply the definition by proving that every segment that connects two points in C is in C. However, this can be tedious in practice. Instead, we can use operations that preserve convexity.

2.3.1 Intersection and union

Property 2.2 (Convexity is preserved by intersection)**.** For any convex sets C_1 and C_2, the intersection $C_1 \cap C_2$ is convex.

Likewise, the intersection of an arbitrary number of convex sets is convex.

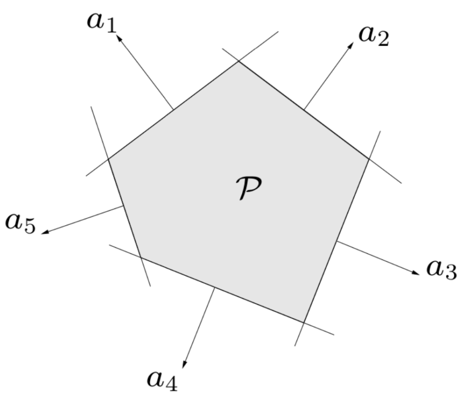


Remark. *The union of convex sets is not necessarily convex. For instance in* $\mathbb{R}$*, both* $[0,1]$ *and* $[2,3]$ *are convex, but their union* $[0,1] \cup [2,3]$ *is not.*

Definition (Polyhedron)**.** A *polyhedron* is the solution set of finitely many linear inequalities and equalities:

$$S = \{\, x \in \mathbb{R}^n \mid Ax \leqslant b, Cx = d \,\}$$

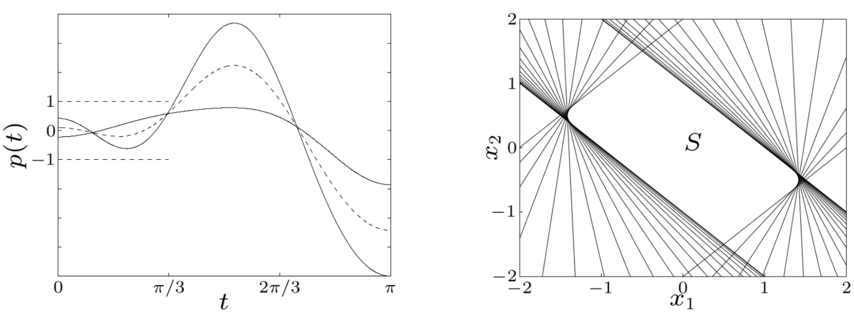
for $A \in \mathscr{M}_{m,n}(\mathbb{R})$ and $C \in \mathscr{M}_{p,n}(\mathbb{R})$. Polyhedra are convex, since they are the intersection of halfspaces and hyperplanes which are convex.



Example. Let:

$$S = \left\{ x \in \mathbb{R}^m \;\middle|\; \forall t \in \mathbb{R}, \quad |t| \leqslant \frac{\pi}{3} \implies \left| \sum_{k=1}^{m} x_k \cos(kt) \right| \leqslant 1 \right\}$$

S is convex, since it can be written as the intersection of convex sets.



Example. $\mathbb{S}_+^n$ is convex since it is the intersection of convex sets:

$$\mathbb{S}_+^n = \left\{ X \in \mathbb{S}^n \,\middle|\, \forall z \in \mathbb{R}^n, z^\top X z \geqslant 0 \right\} = \mathbb{S}^n \cap \bigcap_{z \in \mathbb{R}^n} \left\{ X \in \mathscr{M}_n(\mathbb{R}) \,\middle|\, z^\top X z \geqslant 0 \right\}$$

Each set $\left\{ X \in \mathscr{M}_n(\mathbb{R}) \,\middle|\, z^\top X z \geqslant 0 \right\}$ being convex, their intersection is convex. In particular, it doesn't matter if the number of sets is finite, countable or uncountable.

2.3.2 Affine functions

Property 2.3 (The image of a convex set by an affine function is convex)**.** If $L : \mathbb{R}^n \to \mathbb{R}^m$ is an affine function, then if C is convex, $L(C)$ is convex.

More explicitly, let $A \in \mathscr{M}_{m,n}(\mathbb{R})$ and $b \in \mathbb{R}^m$. The affine function $L(x) = Ax + b$ maps C to $L(C) = \{ y \in \mathbb{R}^m \mid \exists x \in C, y = Ax + b \}$, which is convex if C is convex.

Property 2.4 (The pre-image of a convex set by an affine function is convex)**.** If $L : \mathbb{R}^n \to \mathbb{R}^m$ is an affine function, then $L^{-1}(C)$, the pre-image of C by L defined by:

$$L^{-1}(C) = \{ x \in \mathbb{R}^n \mid L(x) \in C \}$$

is convex if C is convex.

Example (Linear matrix inequalities)**.** Let $A_1, \ldots, A_m \in \mathbb{S}^n(\mathbb{R})$. The set:

$$\left\{ x \in \mathbb{R}^m \,\middle|\, \sum_{i=1}^m x_i A_i \succcurlyeq 0 \right\}$$

is an affine pre-image of $\mathbb{S}_+^n$ for the mapping $L : \mathbb{R}^m \to \mathbb{S}^n$ defined by:

$$L(x) = \sum_{i=1}^m x_i A_i$$

$\mathbb{S}_+^n$ being convex, the set is convex. $\sum_{i=1}^m x_i A_i \succcurlyeq 0$ is called a *linear matrix inequality*.

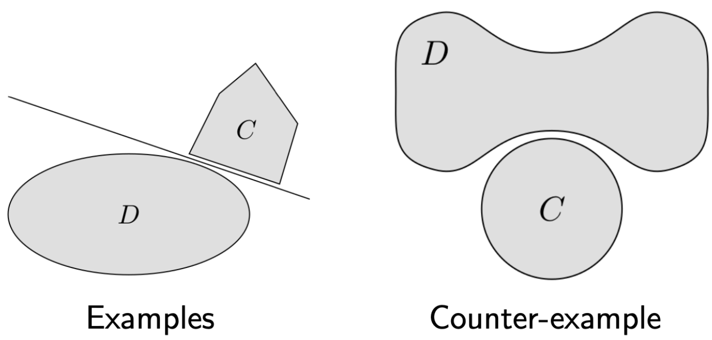
2.4 Geometric elements

2.4.1 Separating and supporting hyperplanes

Property 2.5 (Separating hyperplanes)**.** Soppose that $C, D \subseteq \mathbb{R}^n$ are two non-intersecting convex sets (that is $C \cap D = \varnothing$). Then there exists a hyperplane that separates C and D, that is:

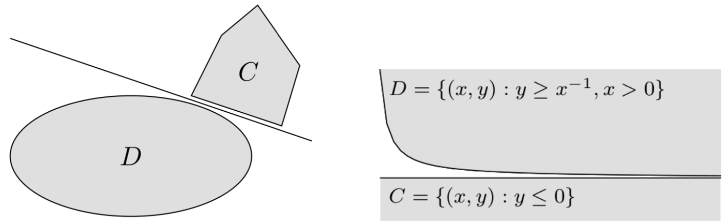
$$\exists s \in \mathbb{R}^n \setminus \{0\}, \exists r \in \mathbb{R}, \quad \forall x \in C, s^\top x \leqslant r \quad \text{and} \quad \forall x \in D, s^\top x \geqslant r$$

where $\left\{ x \in \mathbb{R}^n \,\middle|\, s^\top x = t \right\}$ is called the *separating hyperplane*.



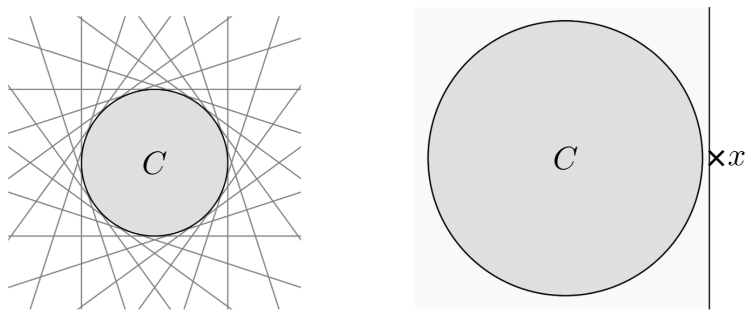
Property 2.6 (Strict separating hyperplanes)**.** Suppose that $C, D \subseteq \mathbb{R}^n$ are two non-intersecting **closed** convex sets, and that one of them is compact (closed and bounded in finite dimension). Then there exists a hyperplane that strictly separates C and D, that is:

$$\exists s \in \mathbb{R}^n \setminus \{0\}, \exists r \in \mathbb{R}, \quad \forall x \in C, s^\top x < r \quad \text{and} \quad \forall x \in D, s^\top x > r$$



Examples Counter-example

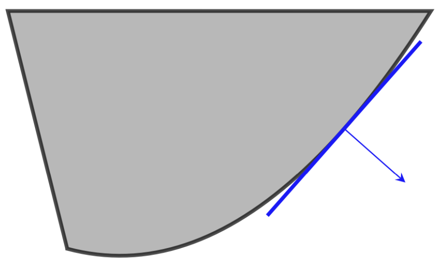
Note that a closed convex set C is the intersection of all halfspaces that contain it.



Definition (Supporting hyperplanes)**.** Supporting hyperplanes touch the boundary of a convex set, and have the entire set on one side. Formally, a hyperplane $H = \left\{ y \,\middle|\, s^\top y = r \right\}$ is a *supporting hyperplane* to a convex set C at a point $x \in \partial C$ if:

$$x^\top x = r \qquad \text{and} \qquad \forall y \in C, \quad a^\top y \leqslant r = s^\top x$$

We also say that H *supports* C at x.



Property 2.7. Let $C \subseteq \mathbb{R}^n$ be a non-empty convex set, and let $x \in \partial C$. Then there exists a supporting hyperplane to C at x.

2.4.2 Cone operators

Definition (Normal cone operator)**.** The *normal cone operator* to a set C at a point x is the set:

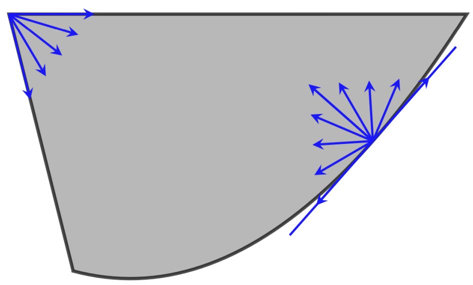
$$N_C(x) = \begin{cases} \left\{ g \in \mathbb{R}^n \,\middle|\, \forall y \in C, \quad g^\top(y-x) \leqslant 0 \right\} & \text{if } x \in C \\ \varnothing & \text{if } x \notin C \end{cases}$$

Intuitively, it is the set of vectors g that form obtuse angles for all $y - x$ with $y \in C$.

For $x \in \mathring{C}$, we have $N_c(x) = \{0\}$. For $x \in \partial C$, $N_C(x)$ is the set of the normal vectors to the supporting hyperplanes to C at x. If $x \notin C$, $N_C(x)$ is empty.

Definition (Tangent vector)**.** Let $C \subseteq \mathbb{R}^n$ be a convex set. A vector $d \in \mathbb{R}^n$ is tangent to C at x if:

$$\exists \{x_k\}_k \subseteq C, \exists \{\lambda_k\}_k \subset \mathbb{R}_+, \quad \lim_{k \to +\infty} \lambda_k(x_k - x) = d$$



Definition (Tangent cone)**.** The tangent cone of a convex set C at x is:

$$T_C(x) = N_C^\circ(x)$$

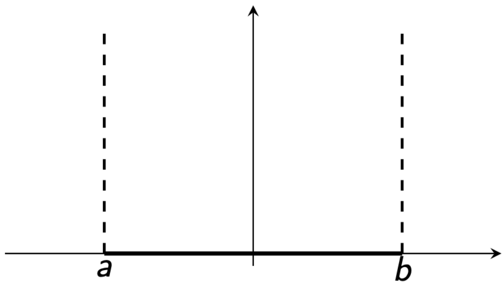
3 Convex functions

3.1 Extended-valued functions

Definition (Extended-valued function)**.** A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is *extended-valued* if its domain is $\mathbb{R}^n$ and its range is $\bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$.

Example (Indicator function)**.** We consider the indicator function of interval $[a, b]$:

$$\mathbb{1}_{[a,b]}(x) := \begin{cases} 0 & \text{if } x \in [a, b] \\ +\infty & \text{otherwise} \end{cases}$$



Definition (Effective domain)**.** The *effective domain* of $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ is the set of points where f is finite:

$$\operatorname{dom} f := \{\, x \in \mathbb{R}^n \mid f(x) < +\infty \,\} \tag{3.1.1}$$

A function is said to be *proper* if its effective domain is non-empty: $\operatorname{dom} f \neq \varnothing$.

3.2 Definition and first properties

Definition (Convex function)**.** A function $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ is *convex* if its graph is below any line connecting two points of the graph $(x, f(x))$ and $(y, f(y))$. That is:

$$\forall x, y \in \mathbb{R}^n, \forall \theta \in [0, 1], \quad f(\theta \cdot x + (1 - \theta) \cdot y) \leqslant \theta \cdot f(x) + (1 - \theta) \cdot f(y) \tag{3.2.1}$$

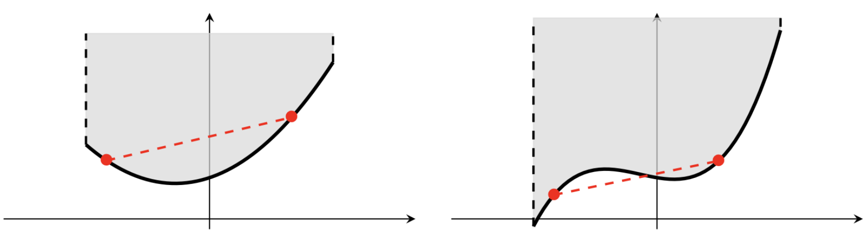
Definition (Concave function)**.** A function $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ is *concave* if $-f$ is convex. That is:

$$\forall x, y \in \mathbb{R}^n, \forall \theta \in [0, 1], \quad f(\theta \cdot x + (1 - \theta) \cdot y) \geqslant \theta \cdot f(x) + (1 - \theta) \cdot f(y)$$

Definition (Epigraph)**.** The *epigraph* of a function $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ is the set of points lying above the graph of f:

$$\operatorname{epi} f := \{\, (x, t) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leqslant t \,\} \tag{3.2.2}$$

Property 3.1 (Convexity and epigraph)**.** A function $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ is convex if and only if its epigraph is a convex set.



The following property allows to check the convexity of a multivariate function f by checking the convexity of functions of one variable.

Property 3.2. Let $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ be a function, and let $x \in \operatorname{dom} f$. We define:

$$\begin{aligned} g_{x,v} : \mathbb{R} &\longrightarrow \bar{\mathbb{R}} \\ t &\longmapsto f(x+tv) \end{aligned}$$

with $\operatorname{dom} g_{x,v} = \{\, t \in \mathbb{R} \mid x + tv \in \operatorname{dom} f \,\}$. Then, f is convex if and only if $g_{x,v}$ is convex in t for all $x \in \operatorname{dom} f$ and all $v \in \mathbb{R}^n$.

Definition (Sublevel sets)**.** Let $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ be a function. The *sublevel set* of f at level $\alpha \in \mathbb{R}$ is the set of points lying below the level α:

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

Property 3.3. If f is convex, then its sublevel sets are convex:

$$f \text{ is convex} \implies \forall \alpha \in \mathbb{R}, \quad S_\alpha(f) \text{ is convex}$$

The converse is not true.

3.3 First-order conditions

Property 3.4 (First-order condition for convexity)**.** Let $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ be a differentiable function, that is that $\nabla f(x)$ exists for all $x \in \operatorname{dom} f$. Then, f is convex if and only if $\operatorname{dom} f$ is convex and:

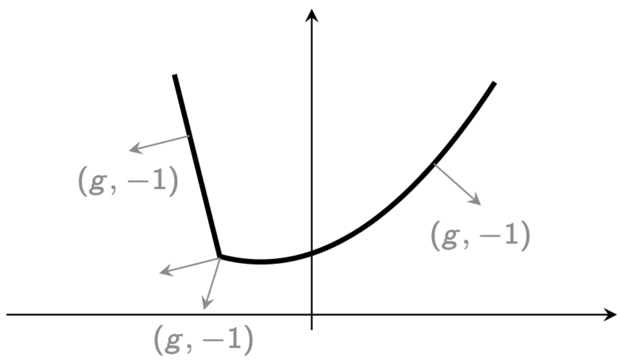
$$\forall x, y \in \operatorname{dom} f, \quad f(y) \geqslant f(x) + \nabla f(x)^\top (y - x)$$

In general, the function f might not be differentiable. In this case, we can use the subdifferential, a generalization of the local variation of a function, to characterize the convexity of f.

Recall that a supporting hyperplane $(g, -1)$ of $\operatorname{epi} f$ at $(x, f(x))$ is a hyperplane such that:

$$\forall y \in \mathbb{R}^n, \quad f(y) \geqslant f(x) + g^\top (y - x)$$

This motivates the following definition.



Definition (Subdifferential)**.** The *subdifferential* of a function $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ is the function associating to each point x the set of all supporting hyperplanes of $\operatorname{epi} f$ at $(x, f(x))$:

$$\begin{aligned} \partial f(x) : \mathbb{R}^n &\longrightarrow \mathcal{P}(\mathbb{R}^n) \\ x &\longmapsto \left\{\, g \in \mathbb{R}^n \;\middle|\; \forall y \in \mathbb{R}^n, \quad f(y) \geqslant f(x) + g^\top (y - x) \,\right\} \end{aligned}$$

Any $g \in \partial f(x)$ is called a *subgradient* of f at x.

- If f is differentiable at x and $\partial f(x) \neq \varnothing$, then $\partial f(x) = \{\nabla f(x)\}$.
- If f is convex, and $\partial f(x)$ is a singleton, then $\partial f(x) = \{\nabla f(x)\}$.
- If f is convex but not differentiable at $x \in \operatorname{int} \operatorname{dom} f$, then:

$$\partial f(x) = \overline{\operatorname{Conv} S(x)} \tag{3.3.1}$$

  where $S(x) = \left\{ s \in \mathbb{R}^n \;\middle|\; \nabla f(x_k) \xrightarrow[x_k \to x]{} s \right\}$
- In general, for a convex function f:

$$\partial f(x) = \overline{\operatorname{Conv} S(x)} + N_{\operatorname{dom} f}(x) \tag{3.3.2}$$

Property 3.5 (Existence of subgradient)**.** For finite-valued convex functions, a subgradient exists for every x.

Property 3.6 (Existence of subgradient for extended-valued functions)**.** In the extendevalued setting, let $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ be a convex function. Then:

1. Subgradients exist for all x in the relative interior of $\operatorname{dom} f$.
2. Subgradients sometimes exist for x on the relative boundary of $\operatorname{dom} f$.
3. No subgradient exists for x ourside of $\operatorname{dom} f$.

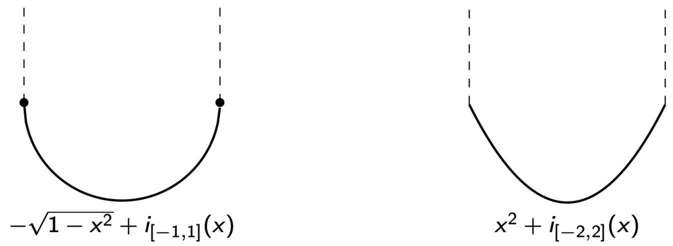


Figure 3.1: Examples for the second case, where boundary points exist on the relative boundary of $\operatorname{dom} f$. No subgradient (affine minorizer) exists for the left function at $x = \pm 1$.

3.4 Second-order conditions

Property 3.7 (Second-order condition for convexity)**.** Let $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ be a twice differentiable function (i.e. $\nabla^2 f(x)$ exists for all $x \in \operatorname{dom} f$ which is open). Then, f is convex if and only if $\operatorname{dom} f$ is convex and:

$$\forall x \in \operatorname{dom} f, \quad \nabla^2 f(x) \succcurlyeq 0 \tag{3.4.1}$$

3.5 Examples

In practice, we showed multiple practical ways to establish the convexity of a function:

- By definition, using the convexity criterion.
- By the existence of subgradients for all points of the domain.
- For twice differentiable functions, by checking the positive semidefiniteness of the Hessian.
- By decomposing the function into simpler functions through operations that preserve convexity.

3.5.1 One-dimensional examples

The following functions are convex:

- affine functions: $x \longmapsto ax + b$, $a, b \in \mathbb{R}$
- exponential functions: $x \longmapsto e^{ax}$, $a \in \mathbb{R}$
- power functions: $x : \mathbb{R}_+^* \longmapsto x^\alpha$, $a \geqslant 1$ or $\alpha \leqslant 0$
- powers of absolute value: $x \longmapsto |x|^p$, $p \geqslant 1$
- negative entropy: $x : \mathbb{R}_+^* \longmapsto x \log x$

The following functions are concave:

- affine functions: $x :\longmapsto ax + b$, $a, b \in \mathbb{R}$ (both convex and concave)
- power functions: $x : \mathbb{R}_+^* \longmapsto x^\alpha$, for $0 \leqslant \alpha \leqslant 1$
- logarithm: $x : \mathbb{R}_+^* \longmapsto \log x$

3.5.2 Examples on vectors

The following functions are convex on $\mathbb{R}^n$:

- affine functions $x \longmapsto a^\top x + b$, $a \in \mathbb{R}^n$, $b \in \mathbb{R}$
- norms: $x \longmapsto \|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$, $p \geqslant 1$
- quadratic functions:
$$f : x \longmapsto \frac{1}{2} x^\top P x + q^\top x + r$$
with $P \in \mathbb{S}^n$, $q \in \mathbb{R}^n$, $r \in \mathbb{R}$. Indeed, we have;
$$\nabla f(x) = Px + q \quad \text{and} \quad \nabla^2 f(x) = P \succcurlyeq 0$$
- least-squares objective:
$$f : x \longmapsto \|Ax - b\|_2^2$$
with $A \in \mathcal{M}_{m,n}(\mathbb{R})$, $b \in \mathbb{R}^m$. Indeed, we have:
$$\nabla f(x) = 2A^\top (Ax - b) \quad \text{and} \quad \nabla^2 f(x) = 2A^\top A \succcurlyeq 0$$

3.5.3 Examples on matrices

The following functions are convex on $\mathcal{M}_{m,n}(\mathbb{R})$:

- affine functions (convex and concave):
$$X \longmapsto \mathrm{Tr}(A^\top X) + b = \sum_{i=1}^m \sum_{j=1}^n A_{i,j} X_{i,j} + b$$
- spectral norm (maximum singular value):
$$X \longmapsto \|X\|_2 = \sigma_{\max}(X) = \sqrt{\lambda_{\max}(X^\top X)}$$
- in general, all norms are convex

3.5.4 Log-determinant function

The $\log\det$ function, defined on $\mathbb{S}^n$, is concave:
$$f : \mathbb{S}^n \longrightarrow \mathbb{R}X \quad \longmapsto \log\det X$$
with $\operatorname{dom} f = \mathbb{S}_{++}^n$. To show this, we will use Property 3.2; we define:
$$\begin{aligned} g_{X,V} : \mathbb{R} &\longrightarrow \mathbb{R} \\ t &\longmapsto \log\det(X + tV) \end{aligned}$$

Note that:

$$\begin{aligned} g_{X,V}(t) &= \log\det(X+tV) \\ &= \log\det X + \log\det(I + tX^{-1/2}VX^{-1/2}) \\ &= \log\det X + \sum_{i=1}^{n} \log(1+t\lambda_i) \end{aligned}$$

where λ_i are the eigenvalues of $X^{-1/2}VX^{-1/2}$.

We then apply the second-order condition to $g_{X,V}$:

$$g''_{X,V}(t) = -\sum_{i=1}^{n} \frac{\lambda_i}{(1+t\lambda_i)^2} \leqslant 0$$

Therefore, $g_{X,V}$ is concave for any X, V, hence f is concave.

3.5.5 Softmax function

The softmax function, defined on $\mathbb{R}^n$, is convex:

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \log \sum_{i=1}^{n} e^{x_i} \end{aligned}$$

If we denote by $z_i = e^{x_i} / \sum_j e^{x_j}$, then we get:

$$\nabla^2 f(x) = \operatorname{diag}(z) - zz^\top$$

with $z_i \geqslant 0$ and $\sum_i z_i = 1$. To show that $\nabla^2 f(x) \succcurlyeq 0$, we show that $\operatorname{diag}(z) - zz^\top$ is positive semidefinite. Let $v \in \mathbb{R}^n$, then:

$$\begin{aligned} v^\top \nabla^2 f(x) v &= v^\top (\operatorname{diag}(z) - zz^\top) v \\ &= \sum_{i=1}^{n} z_i v_i^2 - \left(\sum_{i=1}^{n} z_i v_i \right)^2 \end{aligned}$$

According to the Cauchy-Schwarz inequality applied to $\sqrt{z_i} \times \sqrt{z_i} v_i$, we have:

$$\left(\sum_{i=1}^{n} z_i v_i \right)^2 \leqslant \sum_{i=1}^{n} z_i \sum_{i=1}^{n} z_i v_i^2 = \sum_{i=1}^{n} z_i v_i^2$$

Therefore, $v^\top \nabla^2 f(x) v \geqslant 0$, and f is convex.

3.6 Convexity-preserving operations

3.6.1 Nonnegative weighted sum

Property 3.8 (Nonnegative scaling)**.** Let $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ be a convex function, and $\alpha > 0$. Then, αf is convex.

Property 3.9 (Sum)**.** Let $f_1, f_2 : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ be convex functions. Then, $f_1 + f_2$ is convex; this extends to infinite sums and integrals.

Property 3.10 (Nonnegative weighted sum)**.** Let $f_1, f_2 : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ be convex functions, and $\alpha_1, \alpha_2 > 0$. Then, $\alpha_1 f_1 + \alpha_2 f_2$ is convex; this extends to infinite sums and integrals.

3.6.2 Compositions by an affine function

Property 3.11 (Composition by an affine function)**.** Let $f : \mathbb{R}^n \longrightarrow \bar{R}$ be a convex function and let $A \in \mathcal{M}_{m,}(\mathbb{R})$, $b \in \mathbb{R}^m$. Then:

$$x \longmapsto f(Ax + b) \text{ is convex}$$

Example. The log barrier function for linear inequalities:

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^\top x)$$

with $\operatorname{dom} f = \left\{ x \in \mathbb{R}^n \,\middle|\, \forall i \in [\![1, m]\!], \quad a_i^\top x < b_i \right\}$, is convex.

Example. Any norm of an affine function:

$$f(x) = \|Ax + b\|$$

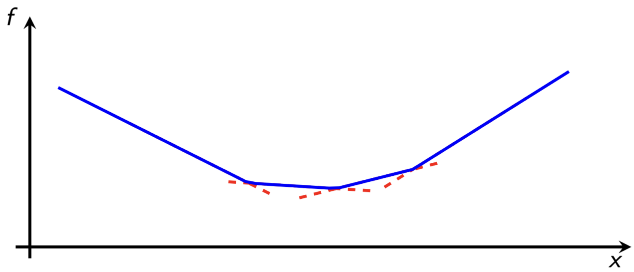
is convex.

3.6.3 Pointwise maximum

Property 3.12 (Pointwise maximum)**.** Let $f_1, f_2 : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ be convex functions. Then, $\max(f_1, f_2)$ is convex. This extends to the pointwise maximum of any finite number of convex functions.

Example. The following piecewise linear function is convex:

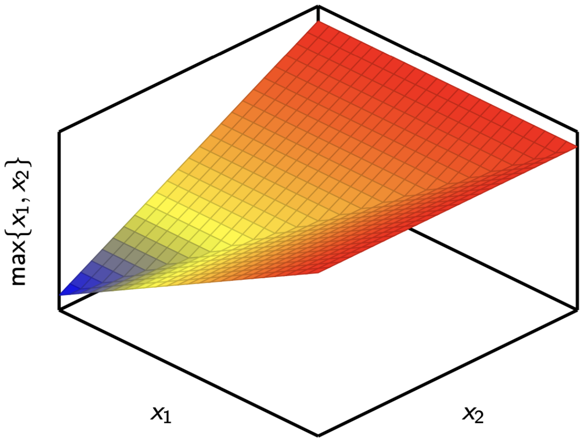
$$f(x) = \max_{i \in [\![1,m]\!]} a_i^\top x + b_i$$



Example (Sum of r largest components)**.** The sum of the r largest components of a vector $x \in \mathbb{R}^n$ is convex:

$$f(x) = x_{(1)} + \cdots + x_{(r)}$$

where $x_{(1)} \geqslant \ldots \geqslant x_{(n)}$ are the components of x sorted in decreasing order.



Indeed, we can write f as:

$$f(x) = \max \{ x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n \}$$

3.6.4 Pointwise supremum

Property 3.13 (Pointwise supremum)**.** If $\forall y \in \mathcal{A}, \quad x \longmapsto f(x, y)$ is convex, then:

$$g(x) = \sup_{y \in \mathcal{A}} f(x, y)$$

is convex.

Example (Support function)**.** The support function of a set C is convex:

$$S_C(x) = \sup_{y \in C} y^\top x$$

Example (Distance to farthest point)**.** The distance to the farthest point in a set C is convex:

$$f(x) = \sup_{y \in C} \|x - y\|$$

Example (Legendre-Fenchel conjugate)**.** Let $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ be a convex function. Then, its Legendre-Fenchel conjugate is convex:

$$f^*(x) = \sup_{y \in \mathbb{R}^n} x^\top y - f(y)$$

3.6.5 Eigenvalues

Property 3.14 (Maximum eigenvalue)**.** The function associating to a symmetric matrix $X \in \mathbb{S}_n$ its maximum eigenvalue is **convex** on $\mathbb{S}_n$:

$$\lambda_{\max}(X) = \sup_{\|y\|_2 = 1} y^\top X y$$

Property 3.15 (Minimum eigenvalue)**.** The function associating to a symmetric matrix $X \in \mathbb{S}_n$ its minimum eigenvalue is **concave** on $\mathbb{S}_n$:

$$\lambda_{\min}(X) = \inf_{\|y\|_2 = 1} y^\top X y$$

3.6.6 Composition with scalar functions

Property 3.16 (Composition with scalar functions)**.** Let $g : \mathbb{R}^n \to \bar{\mathbb{R}}$ and $h : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ be two functions. We define the composition:

$$f(x) = h(g(x))$$

If either:

- g is convex, h is convex and nondecreasing,
- g is concave, h is convex and nonincreasing,

then f is convex.

Proof. We will only prove the case where $n = 1$ and g, h are twice differentiable. We have:

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

If h is convex, then $h''(g(x)) \geqslant 0$ and $h''(g(x))g'(x)^2 \geqslant 0$. In the first case, g convex implies that $g''(x) \geqslant 0$, and h nondecreasing implies that $h'(g(x)) \geqslant 0$. Therefore, $f''(x) \geqslant 0$ and f is convex. In the second case, g concave implies that $g''(x) \leqslant$, and h nonincreasing implies that $h'(g(x)) \leqslant$. Therefore, $f''(x) \geqslant 0$ and f is also convex.

Note that the monotonicity must hold for h on the whole domain of g, including the extended values. □

Example. This allows us to deduce the following properties:

- If g is convex then $\exp g$ is convex.
- If g is concave and positive, then $-\log g$ is convex.
- If g is concave and positive, then $1/g$ is convex.
- If g is convex and nonnegative, then for $\alpha \geqslant 1$ we have that g^α is convex.
- For $\alpha \geqslant 1$, then $\|\cdot\|^\alpha$ is convex (with $h = [\cdot]_+^\alpha$, $g = \|\cdot\|$).

Counter-example. The following counter-example shows the importance of the monotonicity of h:

$$g(x) = x^2 \qquad \text{and} \qquad h = \mathbb{1}_{[1,2]}$$

Then, we have the following composition, which is not convex:

$$h(g(x)) = \mathbb{1}_{[-\sqrt{2},-1]\cup[1,\sqrt{2}]}(x)$$

3.6.7 Vector composition

We derive a property similar to Property 3.16 for vector functions.

Property 3.17 (Vector composition)**.** Let $g : \mathbb{R}^n \to \bar{\mathbb{R}}^k$ and $h : \mathbb{R}^k \to \bar{\mathbb{R}}$ be two functions. We define the composition:

$$f(x) = h(g(x)) = h(g_1(x), \dots, g_k(x))$$

If either:

- the g_i are convex, h is convex and nondecreasing in each argument,
- the g_i are concave, h is convex and nonincreasing in each argument,

then f is convex.

Proof. A proof similar to the one of Property 3.16 can be done, by considering the second derivative of f:

$$f''(x) = g'(x)^\top \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^\top g''(x)$$

This function is positive for similar reasons as in the scalar case. $\square$

Example. This allows us to deduce the following properties:

- If the g_i are concave and positive, then $-\log \sum_{i=1}^m \log g_i$ is convex.
- If the g_i are convex, then $\log \sum_{i=1}^m \exp g_i$ is convex.

3.6.8 Partial minimization

Property 3.18 (Partial minimization)**.** If $f(x, y)$ is convex in (x, y) and C is a non-empty convex set, then the minimization over one variable is convex:

$$g(x) = \inf_{y \in C} f(x, y)$$

Example (Distance to a convex set)**.** The distance to a convex set S is convex:

$$\operatorname{dist}(x, S) = \inf_{y \in S} \|x - y\|$$

4 Convex problems

4.1 Optimization problems in standard form

Definition (Optimization problem)**.** In its standard form, an optimization problem can be written as:

$$\text{minimize } f(x) \quad \text{subject to} \quad \begin{cases} \forall i \in [\![1, m]\!], & g_i(x) \leqslant 0 \\ \forall j \in [\![1, p]\!], & h_j(x) = 0 \end{cases}$$

where:

- $x \in \mathbb{R}^n$ is the optimization variable
- $f : \mathbb{R}^n \to \mathbb{R}$ is the *objectif* or *cost function*
- $g_i : \mathbb{R}^n \to \mathbb{R}$ are the inequality constraint functions
- $h_j : \mathbb{R}^n \to \mathbb{R}$ are the equality constraint functions

Remark. *This form can be generalized to support an infinity of constraints, and strict inequalities. Note that we can assume that the problem is subject only to inequations, without loss of generality: indeed, each equality $h_i(x) = 0$ can be expressed as two inequations $h_i(x) \leqslant 0$ and $-h_i(x) \leqslant 0$.*

Definition (Optimal value)**.** We define the optimal value associated to this optimization problem as:

$$p^* := \inf \{ f(x) \mid \forall i \in [\![1, m]\!],\ g_i(x) \leqslant 0 \quad \text{and} \quad \forall j \in [\![1, p]\!],\ h_j(x) = 0 \}$$

If $p^* = +\infty$, the problem is "infeasible": no x satisfies the constraints.
If $p^* = -\infty$, the problem is *unbounded below*.

Remark. *An optimization problem in standard form has an* implicit constraint *defined by the domain of the constraint functions:*

$$x \in \mathcal{D} := \bigcap_{i=0}^{m} \operatorname{dom} g_i \cap \bigcap_{j=0}^{p} \operatorname{dom} h_j$$

We call $\mathcal{D}$ the domain of the problem*. The constraints $g_i(x) \leqslant 0$ and $h_j(x) = 0$ are the explicit constraints, and the domain of the problem defines the implicit constraints. A problem is* unconstrained *if it has no explicit constraints ($m = p = 0$).*

Example. The following problem is unconstrained:

$$\text{minimize } -\sum_{i=1}^{k} \log(b_i - a_i^\top x)$$

The implit constraints are $a_i^\top x < b_i$ for all $i \in [\![1, k]\!]$.

Definition (Feasibility problem)**.** A feasibility problem is an optimization problem in which we seek a feasible point, i.e. a point that satisfies the constraints. It can be written as:

$$\text{find } x \quad \text{subject to} \quad \begin{cases} \forall i \in [\![1, m]\!], & g_i(x) \leqslant 0 \\ \forall j \in [\![1, p]\!], & h_j(x) = 0 \end{cases}$$

It can be considered a special case of the general problem with $f(x) = 0$:

$$\text{minimize } 0 \quad \text{subject to} \quad \begin{cases} \forall i \in [\![1, m]\!], & g_i(x) \leqslant 0 \\ \forall j \in [\![1, p]\!], & h_j(x) = 0 \end{cases}$$

If constraints are feasible, $p^* = 0$ and any feasible x is optimal.
If constraints are infeasible, $p^* = +\infty$.

4.2 Convex optimization problems

4.2.1 Definition

Definition (Convex Optimization problem)**.** In its standard form, a convex optimization problem can be written as:

$$\text{minimize } f(x) \quad \text{subject to} \quad \begin{cases} \forall i \in [\![1,m]\!], & g_i(x) \leqslant 0 \\ \forall j \in [\![1,p]\!], & a_j^\top x = b_j \end{cases}$$

where the g_i are convex, and the equality constraints are affine.

Such a problem is often written as:

$$\text{minimize } f(x) \quad \text{subject to} \quad \begin{cases} \forall i \in [\![1,m]\!], & g_i(x) \leqslant 0 \\ Ax = b \end{cases}$$

Remark. *The feasible set of a convex optimization problem is convex.*

Example. Consider the following optimization problem:

$$\text{minimize } x_1^2 + x_2^2 \quad \text{subject to} \quad \begin{cases} g_1(x) = x_1/(1+x_2^2) \leqslant 0 \\ h_1(x) = (x_1+x_2)^2 = 0 \end{cases}$$

The objective function $f(x) = x_1^2 + x_2^2$ is convex, and the feasible set

$$\{\, (x_1, x_2) \mid x_1 = -x_2 \leqslant 0 \,\}$$

is convex. Nevertheless, this is not a convex problem according to Definition 4.2.1 because the constraint $g_1(x)$ is not convex and h_1 is not affine. We can rewrite this problem in an equivalent but not identical form:

$$\text{minimize } x_1^2 + x_2^2 \quad \text{subject to} \quad \begin{cases} x_1 \leqslant 0 \\ x_1 + x_2 = 0 \end{cases}$$

This problem is now convex according to Definition 4.2.1.

Remark. *One could ask why we enforce this definition for a convex optimization problem, and why we do not open it to more general forms. In general, recognizing a convex optimization problem is a difficult task, and this allows to provide a simple definition that is easy to check. Note that software tools exist to recognize convex optimization problems via composition rules, such as* Disciplined Convex Programming (DCP)*.*

4.2.2 Optimal and locally optimal points

Definition (Feasible point)**.** A point x is *feasible* if $x \in \operatorname{dom} f$ and it satisfies the constraints:

$$\forall i \in [\![1,m]\!],\ g_i(x) \leqslant 0 \quad \text{and} \quad \forall j \in [\![1,p]\!],\ h_j(x) = 0$$

Definition (Optimal point)**.** A feasible point x is *optimal* if $f(x) = p^*$. We denote X_{opt} the set of optimal points.

Definition (Locally optimal point)**.** A point x is *locally optimal* if there is an $R > 0$ such that x is optimal for the problem restricted to the ball $B(x, R)$:

$$\text{minimize } f(z) \quad \text{subject to} \quad \begin{cases} \forall i \in [\![1, m]\!], \quad g_i(x) \leqslant 0 \\ \forall j \in [\![1, p]\!], \quad h_j(z) = 0 \\ \|z - x\|_2 \leqslant R \end{cases}$$

Example. With $n = 1, m = p = 0$:

- $f(x) = x \log x$, we have $\operatorname{dom} f = \mathbb{R}_+^*$, $p^* = -1/e$, and $x = 1/e$ is optimal
- $f(x) = 1/x$, we have $\operatorname{dom} f = \mathbb{R}_+^*$, $p^* = 0$, but no optimal point
- $f(x) = -\log x$, we have $\operatorname{dom} f = \mathbb{R}_+^*$, $p^* = -\infty$
- $f(x) = x^3 - 3x$, we have $p^* = -\infty$ but a local optimum at $x = 1$

Theorem (Global optimality for convex problems)**.** Any locally optimal point of a convex problem is globally optimal.

Proof. Suppose that x is locally optimal and y is optimal with $f(y) < f(x)$. Since x is locally optimal, there is an $R > 0$ such that:

$$\forall z \in B(x, R), \quad z \text{ feasible} \implies f(z) \geqslant f(x)$$

Now consider $z = \theta y + (1 - \theta)x$ with $\theta = R/(2\|y - x\|_2)$. Since $\|y - x\|_2 > R$, we must have $0 < \theta < 1/2$. z is a combination of two feasible points, hence it is feasible since the problem is convex. Finally, $\|z - x\|_2 = R/2$ hence $z \in B(x, R)$, and:

$$f(z) \leqslant \theta f(x) + (1 - \theta) f(y) < f(x)$$

which contradicts the assumption that x is locally optimal. □

4.2.3 Equivalent convex problems

4.3 Linear optimization

4.4 Quadratic optimization

4.5 Second-order cone optimization

4.6 Generalized inequality constraints

4.7 Semidefinite programming