
Convex Optimization

Adrien Taylor

Class notes by Antoine Groudiev



Last modified 22nd September 2024

Contents

1	Introduction	2
2	Convex sets	2
2.1	Definitions	2
2.2	Examples	2
2.2.1	Hyperplanes and halfspaces	2
2.2.2	Euclidian balls and ellipsoids	3
2.2.3	Cones	4
2.3	Convexity-preserving operations	5
2.3.1	Intersection and union	6
2.3.2	Affine functions	7
2.4	Geometric elements	7
2.4.1	Separating and supporting hyperplanes	7
2.4.2	Cone operators	9
3	Convex functions	9
3.1	Extended-valued functions	9
3.2	Definition and first properties	10
3.3	First-order conditions	11
3.4	Second-order conditions	12
3.5	Examples	12
3.5.1	One-dimensional examples	12
3.5.2	Examples on vectors	12
3.5.3	Examples on matrices	13
3.5.4	Log-determinant function	13
3.5.5	Softmax function	14
4	Convex problems	14

Abstract

This document is Antoine Groudiev's class notes while following the class *Deep Learning* at the Computer Science Department of ENS Ulm. It is freely inspired by the lectures of Adrien Taylor.

1 Introduction

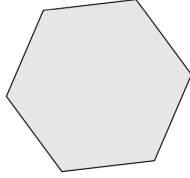
2 Convex sets

2.1 Definitions

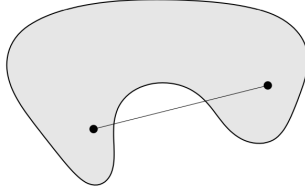
Definition (Convex set). A set C is a *convex set* if every segment that connects two points in C is in C . Formally:

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

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



Convex



Non-convex



Non-convex

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 $\text{Conv}(S)$, 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:

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

with $\theta_1, \dots, \theta_k \geq 0$ 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 :

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

2.2 Examples

2.2.1 Hyperplanes and halfspaces

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

$$H = \{ x \mid a^\top x = b \}$$

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.

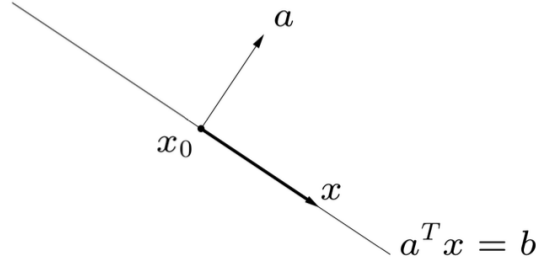


Figure 2.1: Hyperplane

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

$$H = \{ x \mid a^\top x \leq b \}$$

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.

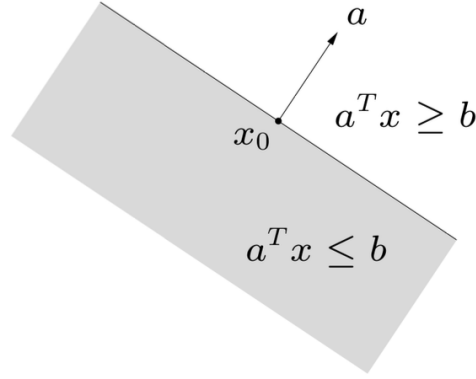


Figure 2.2: Halfspace

2.2.2 Euclidian balls and ellipsoids

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

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

Euclidian balls are convex.

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

$$E = \{ x \mid (x - x_c)^\top P^{-1} (x - x_c) \leq 1 \}$$

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

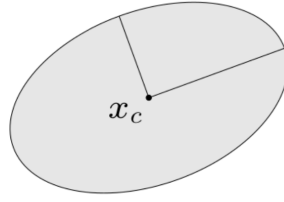


Figure 2.3: Ellipsoid

An alternative representation of an ellipsoid is:

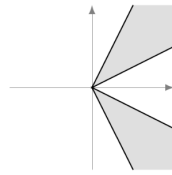
$$E = \{ x_c + Au \mid \|u\|_2 \leq 1 \}$$

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

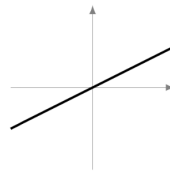
2.2.3 Cones

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

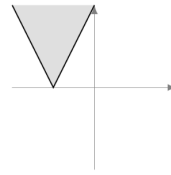
$$\forall x \in K, \forall \theta \in \mathbb{R}_+^*, \quad \theta x \in K$$



Cone



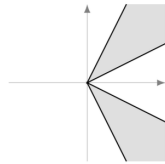
Cone



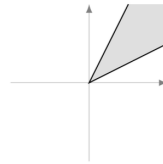
Not cone

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

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



Non-convex



Convex

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:

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

also denoted $X \succcurlyeq 0$.

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

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

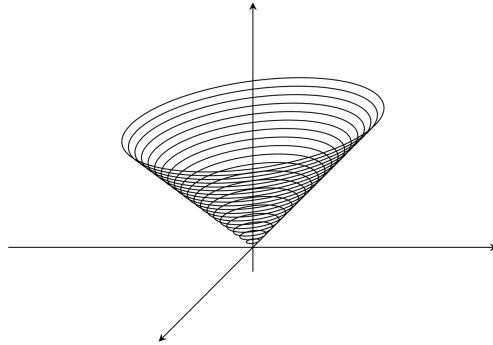
also denoted $X \succ 0$.

\mathbb{S}_+^n and \mathbb{S}_{++}^n are convex cones.

Special cases of cones include:

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

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



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

Positive semidefinite cone $\mathbb{S}_+^n = \{X \in \mathbb{S}^n \mid \forall z \in \mathbb{R}^n, z^\top X z \geq 0\}$

Co-positive cone $\mathbb{S}_+^n = \{X \in \mathbb{S}^n \mid \forall z \in \mathbb{R}_+^n, z^\top X z \geq 0\}$

Exponential cone $\{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+^* \times \mathbb{R} \mid z \geq y e^{x/y}\}$

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

$$K^* = \{y \mid \forall x \in K, y^\top x \geq 0\}$$

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:

$$K^\diamond = \{y \mid \forall x \in K, y^\top x \leq 0\}$$

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

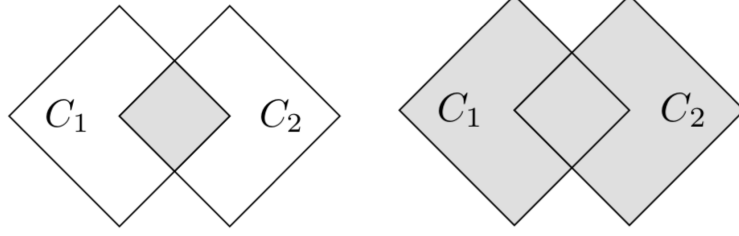
2.3 Convexity-preserving operations

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.

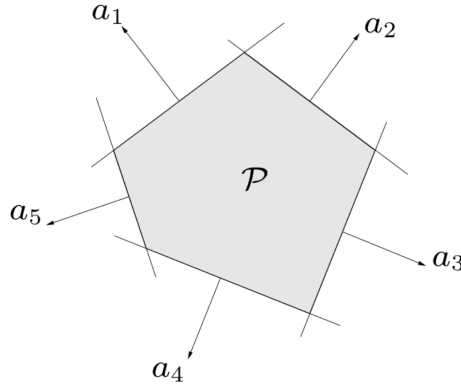


Remark. The union of convex sets is not necessarily convex. For instance in \mathbb{R} , both $[0, 1]$ and $[2, 3]$ are convex, but their union $[0, 1] \cup [2, 3]$ is not.

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

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

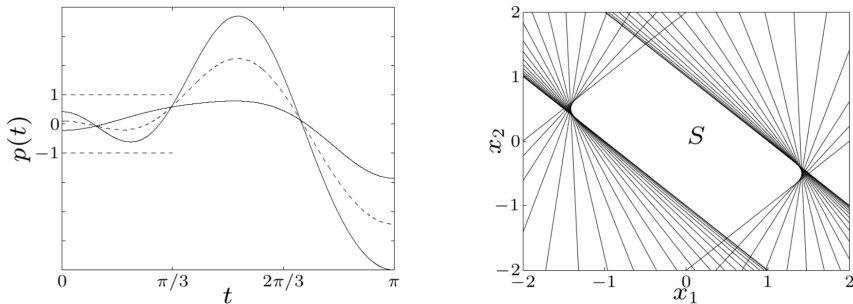
for $A \in \mathcal{M}_{m,n}(\mathbb{R})$ and $C \in \mathcal{M}_{p,n}(\mathbb{R})$. Polyhedra are convex, since they are the intersection of halfspaces and hyperplanes which are convex.



Example. Let:

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

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:

$$\mathbb{S}_+^n = \left\{ X \in \mathbb{S}^n \mid \forall z \in \mathbb{R}^n, z^\top X z \geq 0 \right\} = \mathbb{S}^n \cap \bigcap_{z \in \mathbb{R}^n} \left\{ X \in \mathcal{M}_n(\mathbb{R}) \mid z^\top X z \geq 0 \right\}$$

Each set $\left\{ X \in \mathcal{M}_n(\mathbb{R}) \mid z^\top X z \geq 0 \right\}$ 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 $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is an affine function, then if C is convex, $L(C)$ is convex.

More explicitly, let $A \in \mathcal{M}_{m,n}(\mathbb{R})$ 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 $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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 $A_1, \dots, A_m \in \mathbb{S}^n(\mathbb{R})$. The set:

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

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

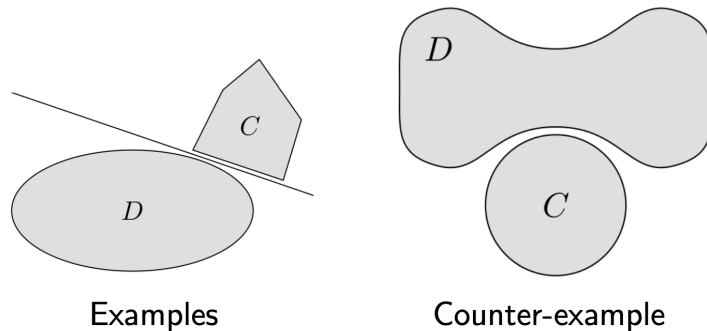
2.4 Geometric elements

2.4.1 Separating and supporting hyperplanes

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

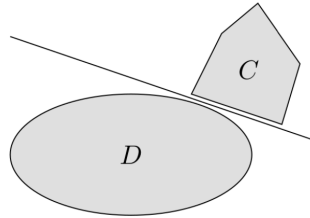
$$\exists s \in \mathbb{R}^n \setminus \{0\}, \exists r \in \mathbb{R}, \quad \forall x \in C, s^\top x \leq r \quad \text{and} \quad \forall x \in D, s^\top x \geq r$$

where $\{ x \in \mathbb{R}^n \mid s^\top x = t \}$ is called the *separating hyperplane*.

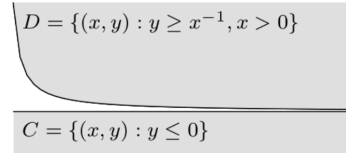


Property 2.6 (Strict separating hyperplanes). Suppose that $C, D \subseteq \mathbb{R}^n$ are two non-intersecting **closed** convex sets, and that one of them is compact (closed and bounded in finite dimension). Then there exists a hyperplane that strictly separates C and D , that is:

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

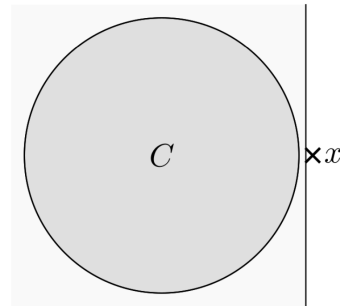
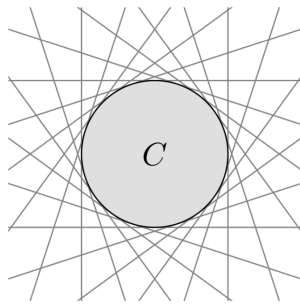


Examples



Counter-example

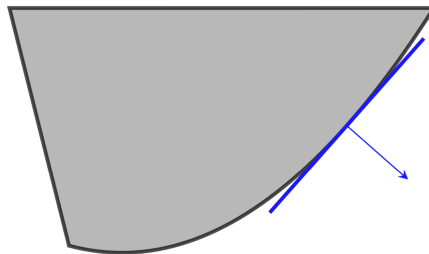
Note that a closed convex set C is the intersection of all halfspaces that contain it.



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

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

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:

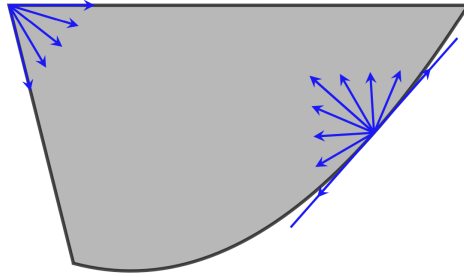
$$N_C(x) = \begin{cases} \left\{ g \in \mathbb{R}^n \mid \forall y \in C, \quad g^\top(y - x) \leq 0 \right\} & \text{if } x \in C \\ \emptyset & \text{if } x \notin C \end{cases}$$

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

For $x \in \overset{\circ}{C}$, 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:

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



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

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

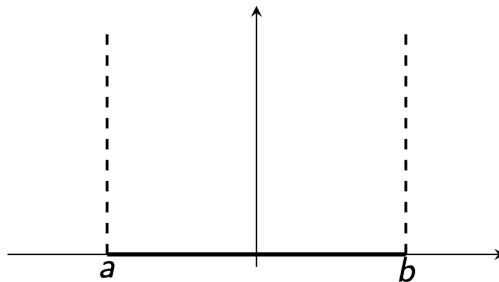
3 Convex functions

3.1 Extended-valued functions

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

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:

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

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

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:

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

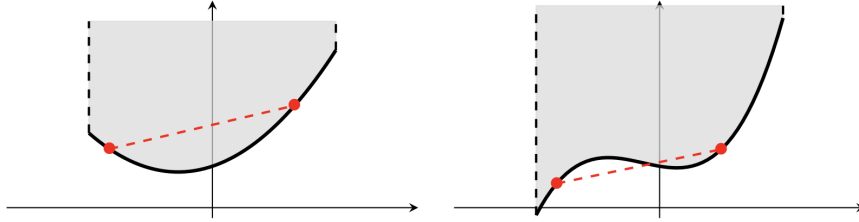
Definition (Concave function). A function $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ is *concave* if $-f$ is convex. That is:

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

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 :

$$\text{epi } f := \{ (x, t) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq t \} \quad (3.2.2)$$

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 $x \in \text{dom } f$. We define:

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

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

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

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 $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ be a differentiable function, that is that $\nabla f(x)$ exists for all $x \in \text{dom } f$. Then, f is convex if and only if $\text{dom } f$ is convex and:

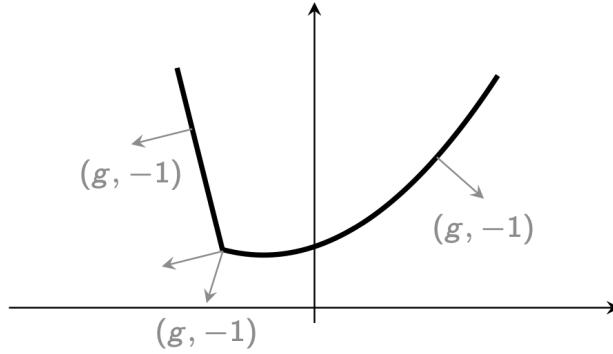
$$\forall x, y \in \text{dom } f, \quad f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$

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 $\text{epi } f$ at $(x, f(x))$ is a hyperplane such that:

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

This motivates the following definition.



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

$$\begin{aligned} \partial f(x) : \mathbb{R}^n &\rightarrow \mathcal{P}(\mathbb{R}^n) \\ x &\mapsto \left\{ g \in \mathbb{R}^n \mid \forall y \in \mathbb{R}^n, \quad f(y) \geq f(x) + g^\top (y - x) \right\} \end{aligned}$$

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

- If f is differentiable at x and $\partial f(x) \neq \emptyset$, 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 $x \in \text{int dom } f$, then:

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

$$\text{where } S(x) = \left\{ s \in \mathbb{R}^n \mid \nabla f(x_k) \xrightarrow{x_k \rightarrow x} s \right\}$$

- In general, for a convex function f :

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

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 extended-valued setting, let $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ be a convex function. Then:

1. Subgradients exist for all x in the relative interior of $\text{dom } f$.

2. Subgradients sometimes exist for x on the relative boundary of $\text{dom } f$.
3. No subgradient exists for x outside of $\text{dom } f$.

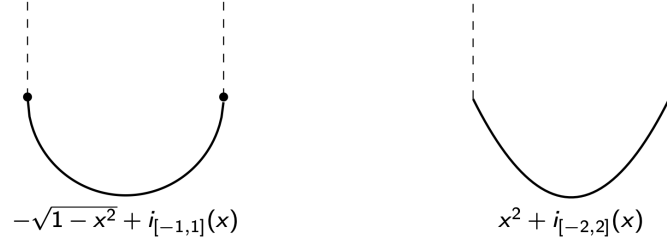


Figure 3.1: Examples for the second case, where boundary points exist on the relative boundary of $\text{dom } f$. No subgradient (affine minorizer) exists for the left function at $x = \pm 1$.

3.4 Second-order conditions

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

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

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

The following functions are concave:

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

3.5.2 Examples on vectors

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

- affine functions $x \mapsto a^\top x + b$, $a \in \mathbb{R}^n$, $b \in \mathbb{R}$
- norms: $x \mapsto \|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$, $p \geq 1$

- quadratic functions:

$$f : x \mapsto \frac{1}{2}x^\top Px + q^\top x + r$$

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:

$$f : x \mapsto \|Ax - b\|_2^2$$

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):

$$X \mapsto \text{Tr}(A^\top X) + b = \sum_{i=1}^m \sum_{j=1}^n A_{i,j} X_{i,j} + b$$

- spectral norm (maximum singular value):

$$X \mapsto \|X\|_2 = \sigma_{\max}(X) = \sqrt{\lambda_{\max}(X^\top X)}$$

- in general, all norms are convex

3.5.4 Log-determinant function

The log det function, defined on \mathbb{S}^n , is concave:

$$f : \mathbb{S}^n \longrightarrow \mathbb{R} \quad X \longmapsto \log \det X$$

with $\text{dom } f = \mathbb{S}_{++}^n$. 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 $g_{X,V}$:

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

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:

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

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

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

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

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

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

4 Convex problems