Convex Optimization

Adrien Taylor

Class notes by Antoine Groudiev

Last modified 19th December 2024

Contents

1 Introduction **4**
- 1.1 Why Convex Optimization? . . . 4
- 1.2 Historical perspective . . . 4
- 1.3 Algorithmic persepective . . . 4
  - 1.3.1 Interior Point Methods . . . 4
  - 1.3.2 First-order methods . . . 5
  - 1.3.3 Time and error . . . 5
- 1.4 Exploiting the problem structure . . . 5

2 Convex sets **6**
- 2.1 Definitions . . . 6
- 2.2 Examples . . . 6
  - 2.2.1 Hyperplanes and halfspaces . . . 6
  - 2.2.2 Euclidean balls and ellipsoids . . . 7
  - 2.2.3 Cones . . . 8
- 2.3 Convexity-preserving operations . . . 9
  - 2.3.1 Intersection and union . . . 10
  - 2.3.2 Affine functions . . . 11
- 2.4 Geometric elements . . . 11
  - 2.4.1 Separating and supporting hyperplanes . . . 11
  - 2.4.2 Cone operators . . . 13

3 Convex functions **14**
- 3.1 Extended-valued functions . . . 14
- 3.2 Definition and first properties . . . 14
- 3.3 First-order conditions . . . 15
- 3.4 Second-order conditions . . . 16
- 3.5 Examples . . . 16
  - 3.5.1 One-dimensional examples . . . 16
  - 3.5.2 Examples on vectors . . . 17
  - 3.5.3 Examples on matrices . . . 17
  - 3.5.4 Log-determinant function . . . 17
  - 3.5.5 Softmax function . . . 18
- 3.6 Convexity-preserving operations . . . 18
  - 3.6.1 Nonnegative weighted sum . . . 18
  - 3.6.2 Compositions by an affine function . . . 19
  - 3.6.3 Pointwise maximum . . . 19
  - 3.6.4 Pointwise supremum . . . 20
  - 3.6.5 Eigenvalues . . . 20
  - 3.6.6 Composition with scalar functions . . . 21
  - 3.6.7 Vector composition . . . 21
  - 3.6.8 Partial minimization . . . 22

4 Convex problems **23**
- 4.1 Optimization problems in standard form . . . 23
- 4.2 Convex optimization problems . . . 24
  - 4.2.1 Definition . . . 24
  - 4.2.2 Optimal and locally optimal points . . . 24
  - 4.2.3 Equivalent convex problems . . . 25

4.3 Special classes of convex problems . . . 26
4.3.1 Linear programming (LP) . . . 26
4.3.2 Convex quadratic programming (QP) . . . 27
4.3.3 Quadratically constrained quadratic programming (QCQP) . . . 28
4.3.4 Second-order cone programming (SOCP) . . . 29
4.4 Robust linear programming . . . 29
4.4.1 Introduction . . . 29
4.4.2 Deterministic approach via SOCP . . . 29
4.4.3 Stochastic approach via SOCP . . . 30
4.5 Generalized inequalities . . . 30
4.5.1 Convex cone properties . . . 31
4.5.2 Semidefinite programming (SDP) . . . 32
4.5.3 LPs and SOCPs as SDPs . . . 33
4.6 Quasi-convex problems . . . 33
4.6.1 Quasi-convex functions . . . 33
4.6.2 Quasi-convex optimization . . . 34
4.6.3 Quasi-convex optimization via bisection . . . 35
4.7 Examples . . . 35
4.7.1 Regression . . . 35
4.7.2 Classification . . . 36

5 Duality I **37**
5.1 Recap on subdifferential calculus . . . 37
5.2 Fermat's rule . . . 38
5.2.1 Definition . . . 38
5.2.2 Constraint qualifications . . . 38
5.2.3 Optimality conditions . . . 39
5.3 Fenchel-Legendre conjugation . . . 39
5.3.1 Conjugate functions . . . 39
5.3.2 Biconjugate . . . 41
5.3.3 Fenchel-Young's equality . . . 42
5.4 A first approach to duality . . . 43
5.5 A second approach to duality: standard forms . . . 45
5.5.1 Formal definitions . . . 45
5.5.2 Examples . . . 46
5.5.3 The dual problem . . . 47
5.5.4 Karush-Kuhn-Tucker conditions . . . 48

6 Duality II **49**
6.1 Examples and interpretations . . . 49
6.1.1 Least-norm solution of linear equations . . . 49
6.1.2 Equality constrained norm minimization . . . 49
6.1.3 LP problem in standard form . . . 49
6.1.4 Two-way partitioning . . . 50
6.1.5 A non-convex problem with strong duality . . . 50
6.1.6 Water-filling . . . 51
6.2 Perturbations and sensitivity analysis . . . 51
6.3 Reformulations . . . 52
6.3.1 Introducing new variables and equality constraints . . . 52
6.3.2 Implicit constraints . . . 53

6.4 Generalized inequalities 53

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

Convex optimization is a subfield of optimization in which we are interested in minimizing a convex function over a convex set. Compared to the general optimization problem, convex optimization solves a simpler problem, allowing to develop efficient algorithms.

1.1 Why Convex Optimization?

Convex optimization is a very mature field, with a lot of theory and algorithms developed over the years. "Generic" working recipes are available for a wide range of problems, and the algorithms are often very efficient. Moreover, the theory of convex optimization is very well understood, and we can often prove that the algorithms converge to the optimal solution.

1.2 Historical perspective

There is a general belief that convex optimization problems are simple. While this is often true, this is not correct in general. Historically, non-linear problems were seen as hard, compared to the easier linear ones. The modern perspective differs:

> "In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity."[1]

Papers such as "¨roblem Complexity and Method Efficiency in Optimization." by Nemirovskii and Yudin in 1979, or "Interior-Point Polynomial Methods in Convex Programming." by Nesterov and Nemirovskii in 1994, progressively developed a robust theory and efficient algorithms for convex optimization.

The simplex algorithm, developed by Dantzig in 1949, solves linear programmin problems in exponential time for the worst case. However, it is very efficient in most cases. The ellipsoid method is used since Khachiyan in 1979 to show the polynomial complexity of LP. It is only since Karmakar in 1984 that an efficient polynomial time algorithm for LP was developed, using interior point methods.

The same interior point methods (IPM) were used by Nesterov and Nemirovskii to develop efficient algorithms for a larger class of structure convex problems. The self-concordance analysis that they introduce extends the polynomial time complexity proof for LPs; most operations that preserve convexity also preserve self-concordance.

1.3 Algorithmic persepective

1.3.1 Interior Point Methods

Interior point methods (IPM) essentially solved once and for all a broad range of medium-scale convex programs. For large-scale problems, computing a single Newton step is often too expensive.

[1] R. T. Rockafellar. "Lagrange multipliers and optimality." SIAM Review, 1993.

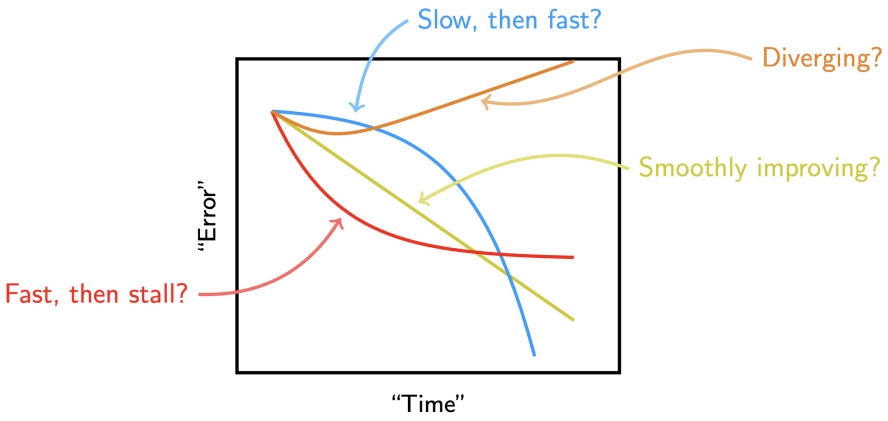
1.3.2 First-order methods

First-order methods are often used for large-scale problems. They are based on the gradient of the function, and are often very efficient. The dependence on precision becomes polynomial $O(1/\varepsilon^\alpha)$, and not logarithmic $O(\log(1/\varepsilon))$ as in IPM. This is acceptable in many applications (statistics, machine learning, etc.). It runs a much larger number of iterations, but each iteration is much cheaper. The lack of Hessian requires significantly less memory and CPU costs per iteration.

On the other hand, there is no unified analysis (as self-concordance for IPM) for first-order methods: there is therefore a huge jungle of disparate methods; the algorithmic choices are stricly constrained by the problem structure.

1.3.3 Time and error

In many optimization schemes, the usages depend on the application requirements (such as the target precision, the time budget, memory budget, etc.).



Diverse algorithms have different trade-offs between time and error.

1.4 Exploiting the problem structure

Formally, our target is to solve a problem of the form:

$$\underset{x \in C}{\text{minimize}} \; f(x)$$

To solve it efficiently, we must exploit the knowledge about the problem properties. What is known about f and C? What can I efficiently compute and use to minimize f over C? What am I aiming for?

For instance, the ability to compute gradients of f, the Hessian, or to solve Newton systems can be crucial. The structure of C can also be very important: can I project onto C? What is the target accuracy. This kind of questions drives the choice for optimization schemes.

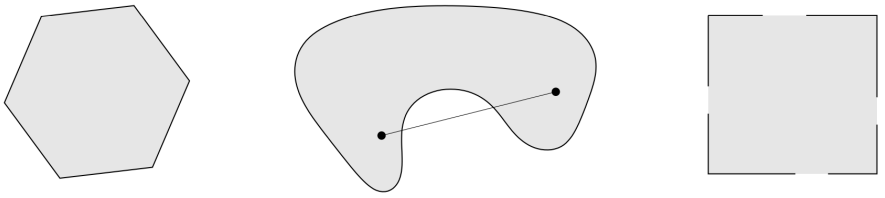
2 Convex sets

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

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

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:

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

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.

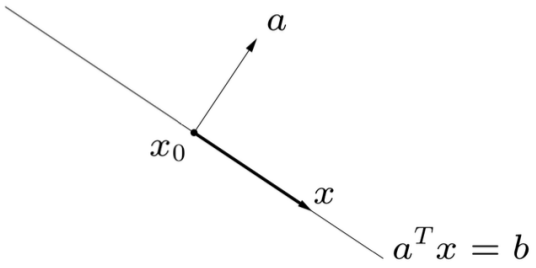


Figure 2.1: Hyperplane

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.

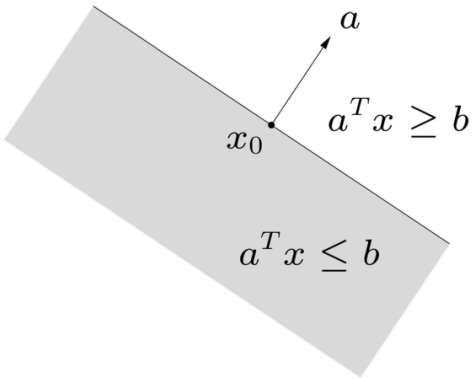


Figure 2.2: Halfspace

2.2.2 Euclidean balls and ellipsoids

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

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

Euclidean 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 $P \in \mathbb{S}^n_{++}$[2] and $x_c \in \mathbb{R}^n$. Ellipsoids are convex.

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

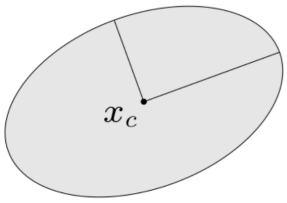


Figure 2.3: Ellipsoid

An alternative representation of an ellipsoid is:

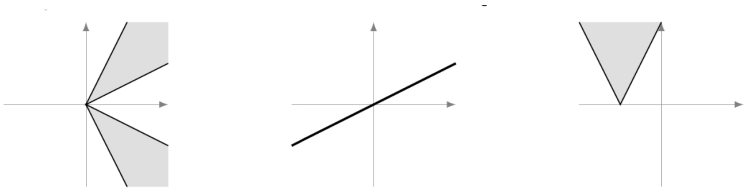
$$E = \{\, x_c + Au \mid \|u\|_2 \leqslant 1 \,\}$$

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

2.2.3 Cones

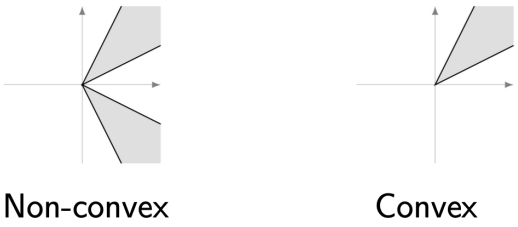
Definition (Cones)**.** A set K is a *cone*, or a *nonnegative homogeneous set*, if:

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



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:

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

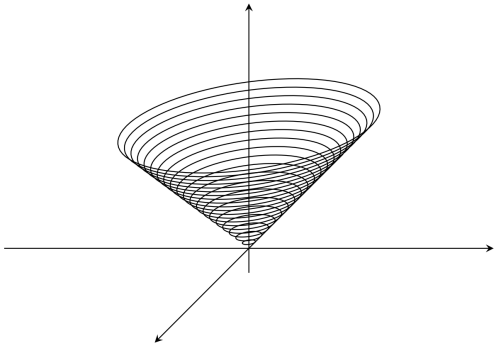
  also denoted $X \succ 0$.

$\mathbb{S}^n_+$ 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 $K^\diamond = -K^*$.

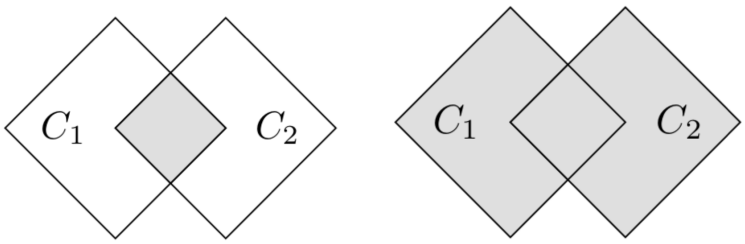
2.3 Convexity-preserving operations

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.

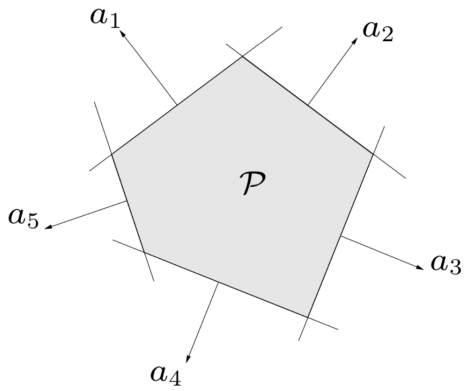


Remark. *The union of convex sets is not necessarily convex. For instance in* $\mathbb{R}$*, both* $[0, 1]$ *and* $[2, 3]$ *are convex, but their union* $[0, 1] \cup [2, 3]$ *is not.*

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

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

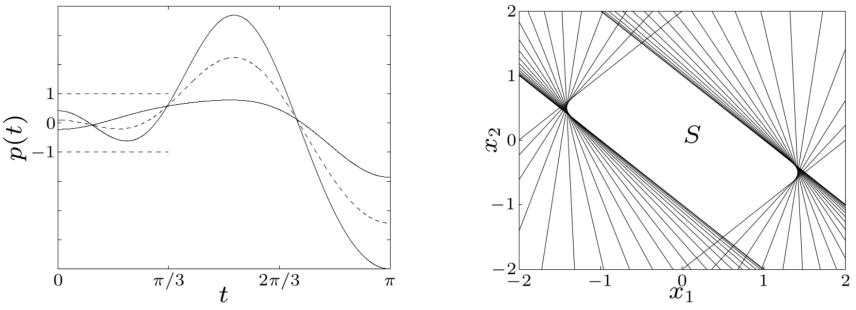
for $A \in \mathscr{M}_{m,n}(\mathbb{R})$ and $C \in \mathscr{M}_{p,n}(\mathbb{R})$. Polyhedra are convex, since they are the intersection of halfspaces and hyperplanes which are convex.



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*.

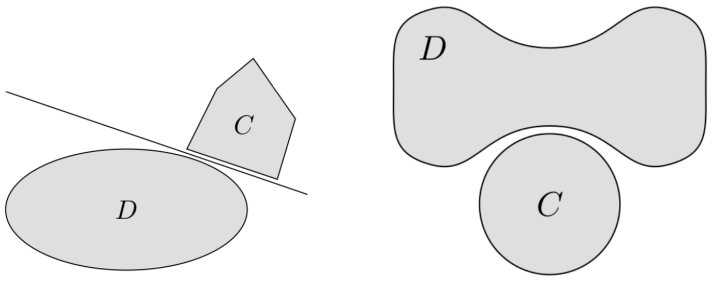
2.4 Geometric elements

2.4.1 Separating and supporting hyperplanes

Property 2.5 (Separating hyperplanes)**.** Suppose 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:

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

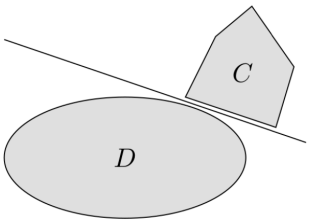
where $\left\{ x \in \mathbb{R}^n \,\middle|\, s^\top x = t \right\}$ is called the *separating hyperplane*.



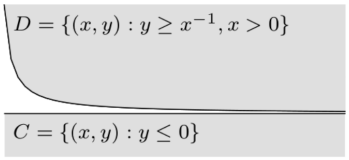
Examples Counter-example

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:

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

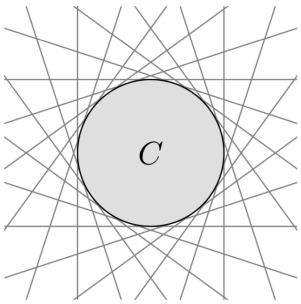


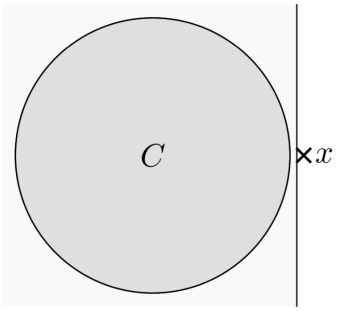
Examples



Counter-example

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

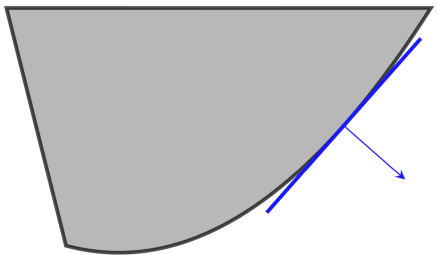




Definition (Supporting hyperplanes)**.** Supporting hyperplanes touch the boundary of a convex set, and have the entire set on one side. Formally, a hyperplane $H = \left\{ y \,\middle|\, s^\top y = r \right\}$ is a *supporting hyperplane* to a convex set C at a point $x \in \partial C$ if:

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

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:

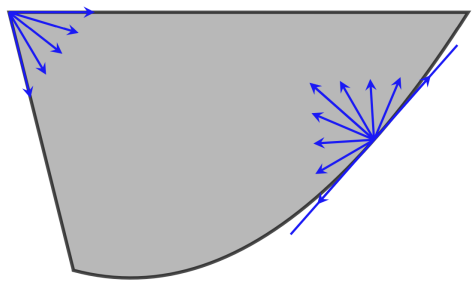
$$N_C(x) = \begin{cases} \left\{ g \in \mathbb{R}^n \,\middle|\, \forall y \in C, \quad g^\top (y - x) \leqslant 0 \right\} & \text{if } x \in C \\ \varnothing & \text{if } x \notin C \end{cases}$$

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:

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

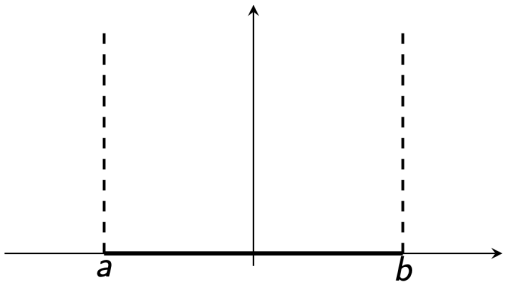
3 Convex functions

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:

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

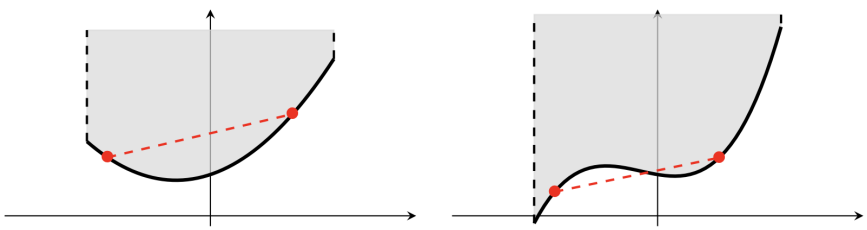
Definition (Concave function)**.** A function $f : \mathbb{R}^n \longrightarrow \bar{\mathbb{R}}$ is *concave* if $-f$ is convex. That is:

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

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:

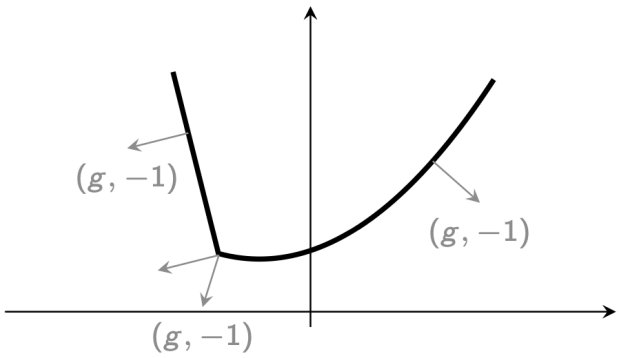
$$\forall x, y \in \operatorname{dom} f, \quad f(y) \geqslant f(x) + \nabla f(x)^\top (y - x)$$

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 $\operatorname{dom} f$.

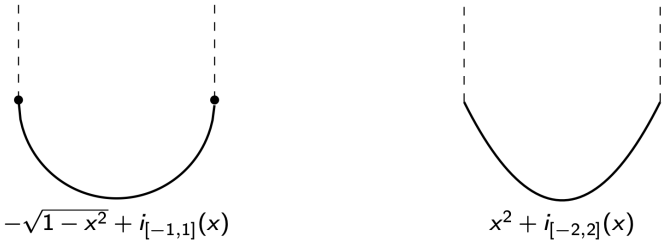


Figure 3.1: Examples for the second case, where boundary points exist on the relative boundary of $\operatorname{dom} f$. No subgradient (affine minorizer) exists for the left function at $x = \pm 1$.

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 $\mathrm{dom}\, f = \mathbb{S}_{++}^n$. 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:

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

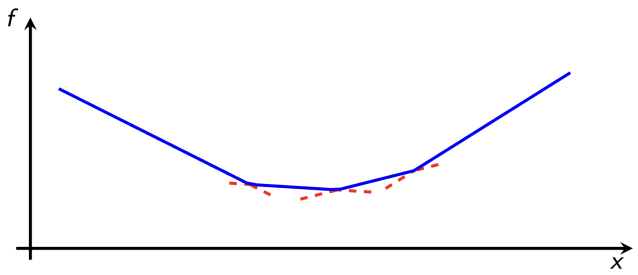
is convex.

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:

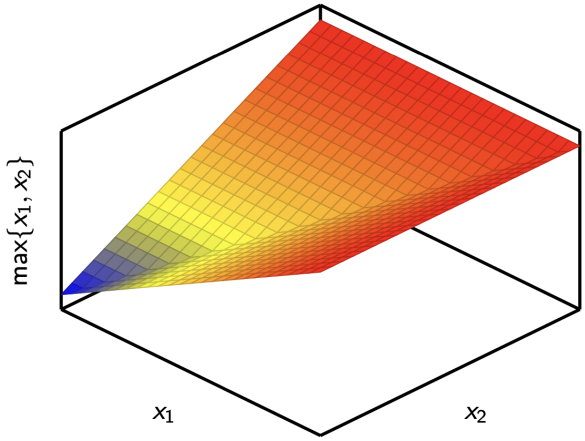
$$f(x) = \max_{i \in [\![1,m]\!]} a_i^\top x + b_i$$



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:

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

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:

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

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. □

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:

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

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

Two problems are informally equivalent if the solution of one is readily obtained from the solution of the other, and vice-versa. In the following, we will see multiple transformations that preserve both the solution and the convexity of an optimization problem.

Eliminating equality constraints Consider the problem:

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

If we can find F and x_0 such that:

$$Ax = b \iff \exists z,\ x = Fz + x_0$$

then we can rewrite the problem as:

$$\text{minimize } f(Fz + x_0) \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ g_i(Fz + x_0) \leqslant 0$$

For instance, one can choose F such that $\operatorname{Im}(F) = \operatorname{Ker}(A)$ and x_0 such that $Ax_0 = b$.

Introducing equality constraints Reciprocally, the problem:

$$\text{minimize } f(A_0 z + b_0) \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ g_i(A_i z + b_i) \leqslant 0$$

Can be rewritten as:

$$\text{minimize } f(y_0) \quad \text{subject to} \quad \begin{cases} \forall i \in [\![1, m]\!], & g_i(y_i) \leqslant 0 \\ \forall i \in [\![0, m]\!], & y_i = A_i x + b_i \end{cases}$$

Introducing slack variables for linear inequalities The idea is to replace linear inequalities by linear equalities and non-negativity constraints. Formally, the problem:

$$\text{minimize } f(x) \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ a_i^\top x \leqslant b_i$$

is equivalent to:

$$\text{minimize } f(y_0) \quad \text{subject to} \quad \begin{cases} \forall i \in [\![1, m]\!],\ a_i^\top x + s_i = b_i \\ \forall i \in [\![1, m]\!],\ s_i \geqslant 0 \end{cases}$$

Epigraph form We saw previously that the epigraph of a convex function is a convex set. We can use this property to rewrite a convex optimization problem in its standard form as:

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

Minimizing over some variables Consider the problem:

$$\text{minimize } f(x_1, x_2) \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ g_i(x_1) \leqslant 0$$

This can be rewritten as:

$$\text{minimize } \tilde{f}(x_1) \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ g_i(x_1) \leqslant 0$$

where $\tilde{f}(x_1) = \inf_{x_2} f(x_1, x_2)$. Said otherwise, we can start by minimizing over the unconstrained variables, and then minimize over the constrained variables.

4.3 Special classes of convex problems

If methods exist to solve general convex optimization problems, some classes of problems have specific structures that can be exploited to design more efficient algorithms. We will present some of these classes in the following.

4.3.1 Linear programming (LP)

Definition (Linear programming problem)**.** A *linear programming* (LP) problem is an optimization problem in which the objective function is affine and the constraints are linear:

$$\text{minimize } c^\top x + d \quad \text{subject to} \begin{cases} Gx \leqslant h \\ Ax = b \end{cases}$$

The feasible set of a linear programming problem is a polyhedron $\mathcal{P}$.

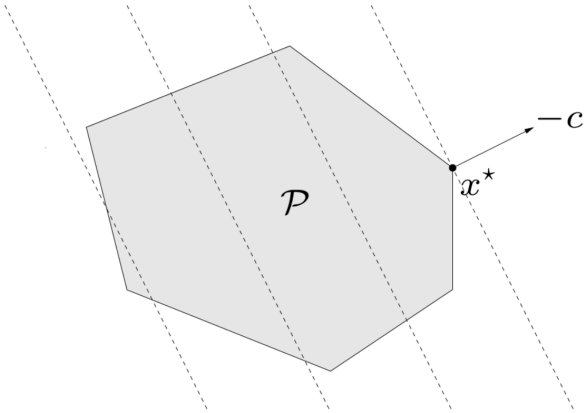


Figure 4.1: A polyhedron $\mathcal{P}$, the feasible set of a linear programming problem.

As an example, we consider the problem of finding the Chebyshev center of a polyhedron.

Definition (Chebyshev center)**.** Given a polyhedron $\mathcal{P}$ of the form:

$$\mathcal{P} = \left\{ x \;\middle|\; \forall i \in [\![1, m]\!],\ a_i^\top x \leqslant b_i,\ \right\}$$

its *Chebyshev center* is the center of the largest inscribed ball. Recall that the ball $B(x_c, r)$ of center x_c and radius r is defined as:

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

Then, the Chebyshev center $\hat{x}$ is the point x_c that maximizes r:

$$\hat{x} = \operatorname*{arg\,min}_{x_c, r} \{\, r \in \mathbb{R}_+ \mid B(x_c, r) \subseteq \mathcal{P} \,\} = \operatorname*{arg\,min}_{x_c} \max_{x \in \mathcal{P}} \|x - x_c\|_2$$

Property 4.1 (Chebyshev center as a linear programming problem)**.** The Chebyshev center of a polyhedron $\mathcal{P}$ can be computed as the solution of the following linear programming problem:

$$\text{maximize } r \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ a_i^\top x_c + r\|a_i\|_2 \leqslant b_i$$

4.3.2 Convex quadratic programming (QP)

Definition (Quadratic programming problem)**.** A *quadratic programming* (QP) problem is an optimization problem in which the objective function is quadratic and the constraints are linear:

$$\text{minimize } \frac{1}{2} x^\top P x + q^\top x + r \quad \text{subject to } \begin{cases} Gx \leqslant h \\ Ax = b \end{cases}$$

where $P \in \mathbb{S}_n^+(\mathbb{R})$ is positive semidefinite.

The feasible set of a quadratic programming problem is a still polyhedron $\mathcal{P}$ (since the constraints have the same form as an LP problem). Solving a QP problem corresponds to minimizing a quadratic function over a polyhedron.

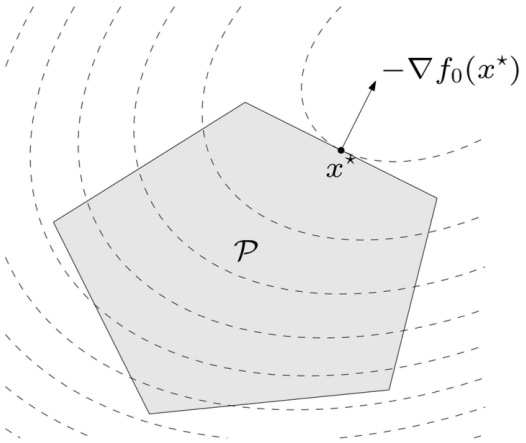


Figure 4.2: A quadratic function over a polyhedron.

Example (Least squares problem)**.** The least squares problem can be written as a QP problem:

$$\text{minimize } \|Ax - b\|_2^2$$

The analytical solution can be expressed using the Moore-Penrose pseudo-inverse of A, $A^\dagger$:

$$x^* = A^\dagger b = (A^\top A)^{-1} A^\top b$$

Linear constraints such as $l \leqslant x \leqslant u$ can be added.

Another common variant is the LASSO regularization:

$$\text{minimize } \|Ax - b\|_2^2 + \lambda \|x\|_1$$

Example (Linear program with random cost)**.** Consider a random vector $c \in \mathbb{R}^n$ with mean $\mathbb{E}[c] =: \bar{c}$ and covariance matrix $\mathbb{V}(c) =: \Sigma$. The following linear program is often used in economics and finance:

$$\text{minimize } \mathbb{E}[c^\top x] + \gamma \mathbb{V}(c^\top x) = \bar{c}^\top x + \gamma x^\top \Sigma x \quad \text{subject to} \quad \begin{cases} Gx \leqslant h \\ Ax = b \end{cases}$$

Where $\gamma > 0$ is a risk-aversion parameter: it controls how much we penalize the variance of the cost. Higher values of γ allow for solutions with higher expected cost but lower variance.

4.3.3 Quadratically constrained quadratic programming (QCQP)

Definition (Quadratically constrained quadratic programming problem)**.** A *quadratically constrained quadratic programming* (QCQP) problem is an optimization problem in which both the objective function and the constraints are quadratic:

$$\text{minimize } \frac{1}{2} x^\top P_0 x + q_0^\top x + r_0 \quad \text{subject to} \begin{cases} \frac{1}{2} x^\top P_i x + q_i^\top x + r_i \leqslant 0, \quad \forall i \in [\![1, m]\!] \\ Ax = b \end{cases}$$

where the objective and constraints matrices $P_i \in \mathbb{S}_n^+(\mathbb{R})$ are positive semidefinite. In the case where $P_1, \ldots, P_m \in \mathbb{S}_n^{++}(\mathbb{R})$ are positive definite, the feasible region is the intersection of m ellipsoids and an affine set.

4.3.4 Second-order cone programming (SOCP)

Definition (Second-order cone programming problem)**.** A *second-order cone programming* (SOCP) problem is an optimization problem in which the objective function is linear, and the constraints are second-order cones:

$$\text{minimize } f^\top x \quad \text{subject to } \begin{cases} \|A_i x + b_i\|_2 \leqslant c_i^\top x + d_i, & \forall i \in [\![1, m]\!] \\ Fx = g \end{cases}$$

where $f \in \mathbb{R}^n$, $A_i \in \mathscr{M}_{n_i,n}(\mathbb{R})$, $b_i \in \mathbb{R}^{n_i}$, $c_i \in \mathbb{R}^{n_i}$, $d_i \in \mathbb{R}$, $F \in \mathscr{M}_{p,n}(\mathbb{R})$, $g \in \mathbb{R}^p$.

Inequalities are second-order cone (SOC) constraints:

$$(A_i x + b_i, c_i^\top x + d_i) \in \text{second-order cone in } \mathbb{R}^{n_i+1}$$

For $n_i = 0$, this reduces to an LP problem. If $c_i = 0$, this reduces to a QCQP problem. In general, SOCP problems are more general than LP and QCQP problems.

4.4 Robust linear programming

4.4.1 Introduction

In many situations, the parameters of an optimization problems might be uncertain. For instance, in an LP problem such as:

$$\text{minimize } c^\top x \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ a_i^\top x \leqslant b_i$$

there can be uncertainty on the values of c, a_i, b_i, which can be modeled as taking any value in given intervals. There are two common approaches to handle this uncertainty: deterministic and stochastic models. Assume that the value of a_i can be any value in the set $\mathcal{E}_i$.

A **deterministic model** solves a harder problem, in which the constraints must hold for all $a_i \in \mathcal{E}_i$. This can be written as:

$$\text{minimize } c^\top x \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ \underbrace{\forall a_i \in \mathcal{E}_i}_{\text{uncertainty}},\ a_i^\top x \leqslant b_i$$

A **stochastic model** uses chance constraints: we consider that a_i is a random variable, and we require that the constraints hold with a given probability η. This can be written as:

$$\text{minimize } c^\top x \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ \underbrace{\mathbb{P}(a_i^\top x \leqslant b_i) \geqslant \eta}_{\text{probability constraint}}$$

In the following, we will develop both approaches using SOCP.

4.4.2 Deterministic approach via SOCP

We can model the uncertainty on a_i by choosing an ellipsoid as $\mathcal{E}_i$:

$$\mathcal{E}_i = \{\, \bar{a}_i + P_i u \mid \|u\|_2 \leqslant 1 \,\}$$

where $\bar{a}_i \in \mathbb{R}^n$ are the centers, and the semi-axes are determined by the singular values and vectors of $P_i \in \mathscr{M}_n(\mathbb{R})$. Therefore, the robust LP problem of the form:

$$\text{minimize } c^\top x \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ \forall a_i \in \mathcal{E}_i,\ a_i^\top x \leqslant b_i$$

can be written as the following SOCP problem:

$$\text{minimize } c^\top x \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ \bar{a}_i^\top x + \|P_i^\top x\|_2 \leqslant b_i$$

This comes from the fact that:

$$\sup_{\|u\|_2 \leqslant 1} (\bar{a}_i + P_i u)^\top x = \bar{a}_i^\top x + \|P_i^\top x\|_2$$

hence, if $\bar{a}_i^\top x + \|P_i^\top x\|_2 \leqslant b_i$ holds, then $a_i^\top x \leqslant b_i$ holds for all $a_i \in \mathcal{E}_i$.

4.4.3 Stochastic approach via SOCP

Assume that a_i is Gaussian with mean $\bar{a}_i$ and covariance matrix Σ_i:

$$a_i \sim \mathcal{N}(\bar{a}_i, \Sigma_i)$$

Therefore, $a_i^\top x$ is a Gaussian random vector with mean $\bar{a}_i^\top x$ and variance $x^\top \Sigma_i x$. We can model the probability constraint as:

$$\mathbb{P}(a_i^\top x \leqslant b_i) = \Phi\left(\frac{b_i - \bar{a}_i^\top x}{\|\Sigma_i^{1/2} x\|_2}\right)$$

where $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} e^{-t^2/2} dt$ is the *Cumulative Distribution Function* (CDF) of $\mathcal{N}(0, 1)$. The robust LP problem can be written as:

$$\text{minimize } c^\top x \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ \mathbb{P}(a_i^\top x \leqslant b_i) \geqslant \eta$$

This can be written as the following SOCP problem:

$$\text{minimize } c^\top x \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ \bar{a}_i^\top x + \Phi^{-1}(\eta) \|\Sigma_i^{1/2} x\|_2 \leqslant b_i$$

which is an SOCP problem when $\Phi^{-1}(\eta) \geqslant 0$, verified for $\eta \geqslant 1/2$. For values of $\eta < 1/2$, the problem is non-convex, as can be seen in the following visualization.

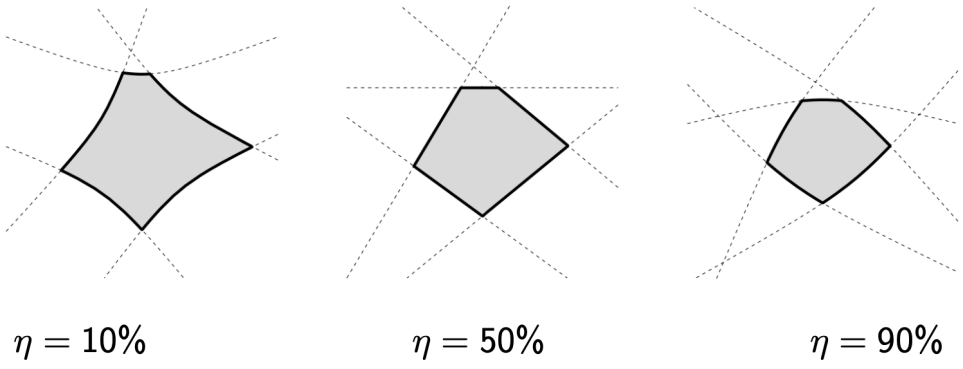


Figure 4.3: The set $\left\{ x \,\middle|\, \forall i \in [\![1, m]\!],\ \mathbb{P}(a_i^\top x \leqslant b_i) \geqslant \eta \right\}$ for multiple values of η. It is convex for $\eta \geqslant 1/2$ and in general non-convex for $\eta < 1/2$.

4.5 Generalized inequalities

For multiple reasons, it can be useful to consider more general inequalities than the standard ones, in the sense that the inequalities are not necessarily defined on the real line. This is mainly motivated by semi-definite programming, where the inequalities must hold on the cone of positive semidefinite matrices. We first define a partial ordering $\succcurlyeq$ on a cone K, and use it to introduce generalized inequalities.

4.5.1 Convex cone properties

Recall that a set K is a convex cone if:

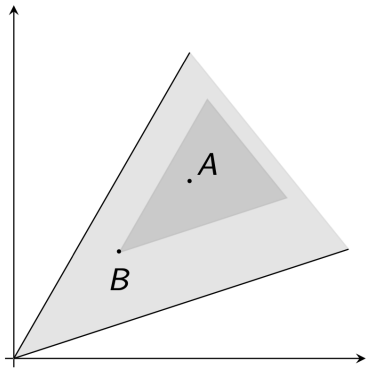
$$x_1, x_1 \in K \implies \forall \theta_1, \theta_2 > 0, \quad \theta_1 x_1 + \theta_2 x_2 \in K$$

Definition. A convex cone K can have the following properties:

- *pointed*: if it contains 0
- *salient*: if it contains no line, that is $x \in K \wedge -x \in K \implies x = 0$
- *solid*: if it has a non-empty interior ($\mathring{K} \neq \varnothing$)
- *closed*: if $K^{\subset}$ is an open set

Definition (Notation)**.** Let K be a pointed, salient and solid convex cone. Let A and B be two points of the ambient space.

- We note $A \succcurlyeq_K 0$ if and only if $A \in K$.
- We note $A \succcurlyeq_K B$ if and only if $A - B \succcurlyeq_K 0$.
- We note $A \succ_K 0$ if and only if $A \in \mathring{K}$ (which makes sense if K is solid).



Property 4.2. $\succcurlyeq_K$ defines a partial ordering.

Proof.

- K is pointed, hence $0 \in K$, hence for any A we have that $A - A \succcurlyeq_K 0$ and $A \succcurlyeq_K A$.
- $A \succcurlyeq_K B$ and $B \succcurlyeq_K A$ implies that $A - B \in K$ and $B - A \in K$. Since K is salient, $A - B = 0$ and $A = B$.
- $A \succcurlyeq_K B$ and $B \succcurlyeq_K C$ implies that $A - B \in K$ and $B - C \in K$. Therefore, $A - C = (A - B) + (B - C) \in K$ since K is a cone, and therefore $A \succcurlyeq_K C$.

□

Definition (Convex problem with generalized inequality constraints)**.** A *convex optimization problem with generalized inequality* constraints is a problem of the form:

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex, and the $g_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$ are K_i-convex with respect to the proper cones K_i:

$$\forall i \in [\![1, m]\!],\ \forall x, y \in \mathbb{R}^n,\ \forall \theta \in [0, 1], \quad g_i(\theta x + (1 - \theta)y) \preccurlyeq_{K_i} \theta g_i(x) + (1 - \theta) g_i(y)$$

A convex optimization problem with generalized inequality constraints as the same properties as a standard convex problem: its feasible set is convex, any local optimum is a global optimum, etc. When the K_i are clear from the context, we can simply use $\preccurlyeq$ and omit the K_i.

Definition (Conic convex optimization problem)**.** A *conic convex optimization problem* is a special case of the generalized convex problem, where the objective and the constraints are affine:

$$\text{minimize } c^\top x \quad \text{subject to} \quad \begin{cases} Fx + g \preccurlyeq_K 0 \\ Ax = b \end{cases}$$

where $c \in \mathbb{R}^n$, $F \in \mathscr{M}_{p,n}(\mathbb{R})$, $g \in \mathbb{R}^p$, $A \in \mathscr{M}_{m,n}(\mathbb{R})$. This extends linear programming to non-polyedral cones. The linear programming case can be obtained by choosing $K = \mathbb{R}^m_+$.

4.5.2 Semidefinite programming (SDP)

Definition. A *semidefinite programming* (SDP) problem is a problem of the form:

$$\text{minimize } c^\top x \quad \text{subject to} \quad \begin{cases} x_1 G_1 + \cdots + x_n G_n + H \preccurlyeq 0 \\ Ax = b \end{cases}$$

where $G_i, H \in \mathbb{S}_k(\mathbb{R})$.

Remark. *The inequality constraint is called a* linear matrix inequality *(LMI). Note that problems with multiple LMIT constraints can be transformed into a single LMI constraint of higher dimension. Suppose that you have the constraints:*

$$x_1 \hat{G}_1 + \cdots + x_n \hat{G}_n + \hat{H} \preccurlyeq 0 \quad \textit{and} \quad x_1 \tilde{G}_1 + \cdots + x_n \tilde{G}_n + \tilde{H} \preccurlyeq 0$$

They can be rewritten as a single LMI:

$$x_1 \begin{bmatrix} \hat{G}_1 & 0 \\ 0 & \tilde{G}_1 \end{bmatrix} + \cdots + x_n \begin{bmatrix} \hat{G}_n & 0 \\ 0 & \tilde{G}_n \end{bmatrix} + \begin{bmatrix} \hat{H} & 0 \\ 0 & \tilde{H} \end{bmatrix} \preccurlyeq 0$$

Example (Largest eigenvalue minimization)**.** Consider the following problem:

$$\text{minimize } \lambda_{\max}(A(x))$$

where $A(x) = A_0 + x_1 A_x + \cdots + x_n A_n$ and $A_0, A_1, \ldots, A_n \in \mathbb{S}_k(\mathbb{R})$. This problem can be written as an equivalent SDP problem:

$$\text{minimize } t \quad \text{subject to} \quad A(x) \preccurlyeq tI$$

over variables $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$. This follows from:

$$\lambda_{\max}(A(x)) \leqslant t \iff A(x) \preccurlyeq tI$$

Example (Matrix norm minimization)**.** Consider the following problem:

$$\text{minimize } \|A(x)\|_2 = \sqrt{\lambda_{\max}(A(x)^\top A(x))}$$

where $A(x) = A_0 + x_1 A_x + \cdots + x_n A_n$ and $A_i \in \mathbb{S}_k(\mathbb{R})$. This problem can be written as an equivalent SDP problem:

$$\text{minimize } t \quad \text{subject to} \quad \begin{bmatrix} tI & A(x) \\ A(x)^\top & tI \end{bmatrix} \succeq 0$$

over variables $x \in \mathbb{R}^n$ and $t \in \mathbb{R}_+$. This follows from:

$$\|A(x)\|_2 \leqslant t \iff A^\top A \preccurlyeq t^2 I \iff \begin{bmatrix} tI & A(x) \\ A(x)^\top & tI \end{bmatrix} \succeq 0$$

where the last equivalence comes from the Schur complement.

4.5.3 LPs and SOCPs as SDPs

Property 4.3 (Any LP problem is equivalent to an SDP problem)**.** Consider the LP problem:

$$\text{minimize } c^\top x \quad \text{subject to} \quad Ax \leqslant b$$

This can be written as the following SDP problem:

$$\text{minimize } c^\top x \quad \text{subject to} \quad \operatorname{diag}(Ax - b) \preccurlyeq 0$$

Property 4.4 (Any SOCP problem is equivalent to an SDP problem)**.** Consider the SOCP problem:

$$\text{minimize } f^\top x \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ \|A_i x + b_i\|_2 \leqslant c_i^\top x + d_i$$

This can be written as the following SDP problem:

$$\text{minimize } f^\top x \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ \begin{bmatrix} c_i^\top x + d_i & A_i x + b_i \\ (A_i x + b_i)^\top & c_i^\top x + d_i \end{bmatrix} \succcurlyeq 0$$

4.6 Quasi-convex problems

4.6.1 Quasi-convex functions

Definition (Quasi-convex function)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is *quasi-convex* if $\operatorname{dom} f$ is convex at the sublevel sets; that is, for all $\alpha \in \mathbb{R}$, the set S_α is convex:

$$S_\alpha := \{\, x \in \operatorname{dom} f \mid f(x) \leqslant \alpha \,\}$$

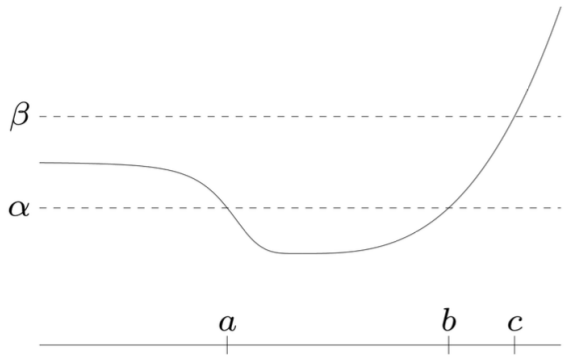


Figure 4.4: A quasi-convex function.

Definition (Quasi-concave function)**.** A function f is *quasi-concave* if $-f$ is quasi-convex.

Definition (Quasi-linear function)**.** A function f is *quasi-linear* if it is both quasi-convex and quasi-concave.

Example.

- $x \mapsto \sqrt{|x|}$ is quasi-convex on $\mathbb{R}$
- $x \mapsto \lfloor x \rfloor$ is quasi-linear
- $\log x$ is quasi-linear on $\mathbb{R}_+^*$
- $f(x_1, x_2) = x_1 x_2$ is quasi-concave on $(\mathbb{R}_+^*)^2$

- Any linear-fractional function is quasi-linear:

$$f(x) = \frac{a^\top x + b}{c^\top x + d}$$

  where $\operatorname{dom} f = \left\{ x \;\middle|\; c^\top x + d > 0 \right\}$
- The distance ratio is quasi-convex:

$$f(x) = \frac{\|x - a\|_2}{\|x - b\|_2}$$

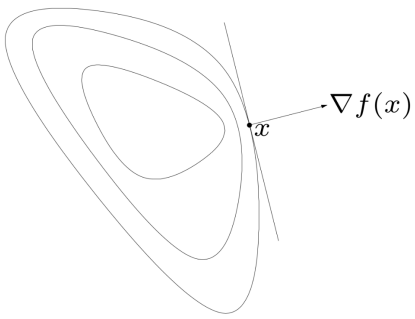
  where $\operatorname{dom} f = \{ x \mid \|x - a\|_2 \leqslant \|x - b\|_2 \}$.

Property 4.5 (Modified Jensen inequality)**.** Let f be a quasi-convex function, and $x, y \in \operatorname{dom} f$. Then, for all $\theta \in [0, 1]$:

$$f(\theta x + (1 - \theta)y) \leqslant \max(f(x), f(y))$$

Property 4.6 (First-order condition)**.** A differentiable function f with convex domain is quasi-convex if and only if:

$$\forall x \in \operatorname{dom} f,\ \forall y \in \operatorname{dom} f,\ f(y) \leqslant f(x) \implies \nabla f(x)^\top (y - x) \leqslant 0$$



Remark. *The sum of two quasi-convex functions is not necessarily quasi-convex.*

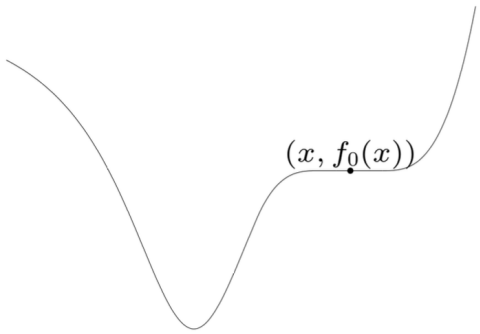
4.6.2 Quasi-convex optimization

Definition (Quasi-convex optimization problem)**.** A *quasi-convex optimization problem in standard form* is an optimization problem in which the objective function is quasi-convex and the constraints are convex:

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

where f is **quasi-convex**, and the g_i are convex.

Remark. *Such a problem can have locally optimal points that are not globally optimal.*



4.6.3 Quasi-convex optimization via bisection

Consider for a fixed t consider the following feasibility problem in x:

$$\text{find } x \quad \text{subject to} \quad \begin{cases} f(x) \leqslant t \\ \forall i \in [\![1, m]\!],\ g_i(x) \leqslant 0 \\ Ax = b \end{cases} \tag{4.6.1}$$

where f is quasi-convex and the g_i are convex.

If this problem is feasible, we can conclude that $t \geqslant p^*$; if it is infeasible, $t \leqslant p^*$. This leads to the following bisection algorithm, acting as a binary search. Consider two initial values $I_0 < u_0$. We simply repeat the following steps:

1. $t := (I + u)/2$
2. Solve the convex feasibility problem (4.6.1)
3. If (4.6.1) is feasible, $u := t$; otherwise, $I = t$.

This requires $\lceil \log_2((u_0 - I_0)/\varepsilon) \rceil$ iterations to obtain an optimality gap of at most ε.

4.7 Examples

We present two common applications in statistical learning, taking the form of optimization problems: regression and classification.

4.7.1 Regression

Consider a set of data formed of vectors $x_i \in \mathbb{R}^d$ and labels $y_i \in \mathbb{R}$:

$$\{\, (x_i, y_i) \mid i \in [\![1, n]\!] \,\} \subset \mathbb{R}^d \times \mathbb{R}$$

We can form the following data matrix X and labels vector Y:

$$X := \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \quad \text{and} \quad Y := \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

The goal of linear regression is to find a $\theta \in \mathbb{R}^d$ such that:

$$X^\top \theta \simeq Y$$

This can be done by solving the *Least Absolute Shrinkage and Selection Operator* (LASSO) of parameter $\lambda > 0$:

$$\min_\theta \|X^\top \theta - Y\|_2^2 + \lambda \|\theta\|_1$$

4.7.2 Classification

Consider a set of data formed of vectors $x_i \in \mathbb{R}^d$ and labels $y_i \in \{-1, 1\}$:

$$\{ (x_i, y_i) \mid i \in [\![1, n]\!] \} \subset \mathbb{R}^d \times \{-1, 1\}$$

The hard-margin *Support Vector Machine* (SVM) is defined as:

$$\text{find } w \in \mathbb{R}^d \quad \text{such that} \quad \forall i \in [\![1, n]\!],\ y_i(w^\top x_i + b) \geqslant 1$$

This problem can be solved via optimization by maximizing the margin $2/\|w\|_2$:

$$\underset{w}{\text{minimize}}\ \|w\|_2^2 \quad \text{such that} \quad \forall i \in [\![1, n]\!],\ y_i(w^\top x_i + b) \geqslant 1$$

The soft-margin SVM is defined as:

$$\underset{s \geqslant 0, w}{\text{minimize}}\ \lambda \|w\|_2^2 + \sum_{i=1}^{n} s_i \quad \text{such that} \quad \forall i \in [\![1, n]\!],\ y_i(w^\top x_i + b) \geqslant 1$$

5 Duality I

An optimization problem can always be seen from two perspectives: the primal problem, which is the problem we want to solve, and the dual problem, which is a reformulation of the primal problem. While their solutions are not always equal, the dual problem is often easier to solve than the primal problem, and it can provide useful information about it.

5.1 Recap on subdifferential calculus

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. The subdifferential of f at a point x is defined as:

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

and any $g \in \partial f(x)$ is called a subgradient of f at x. The subdifferential is a set-valued function, and it is always non-empty for convex functions. If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$. The subdifferential is a generalization of the gradient for non-differentiable functions.

In the following, we derive formulae for the subdifferential of some common operations.

Property 5.1 (Subdifferential of a scaled function)**.** Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ and $\alpha > 0$. We define $h(x) = \alpha f(x)$. Then, for all $x \in \mathbb{R}^n$:

$$\partial h(x) = \alpha \partial f(x)$$

Property 5.2 (Subdifferential of a composition with linear mapping)**.** Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ and $A \in \mathscr{M}_{n,d}(\mathbb{R})$. We define $h(x) = f(Ax)$. Then, for all $x \in \mathbb{R}^n$:

$$\partial h(x) = A^\top \partial f(Ax)$$

when $\operatorname{relint} \operatorname{dom} h \neq \varnothing$.

Property 5.3 (Subdifferential of a sum of functions)**.**
Let $f_1, f_2 : \mathbb{R}^n \to \bar{\mathbb{R}}$ and $(\operatorname{relint} \operatorname{dom} f_1) \cap (\operatorname{relint} \operatorname{dom} f_2) \neq \varnothing$. Then, for all $x \in \mathbb{R}^n$:

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x)$$

Remark. *The assumption* $(\operatorname{relint} \operatorname{dom} f_1) \cap (\operatorname{relint} \operatorname{dom} f_2) \neq \varnothing$ *is often referred to as a* constraint qualification. *Such a constraint is necessary. Consider the following example: the function* $\mathbb{1}_{(0,0)}$ *is the indicator function of the origin in* $\mathbb{R}^2$*. Its subdifferential at* $(0,0)$ *is* $\partial \mathbb{1}_{(0,0)}((0,0)) = \mathbb{R}^2$*. Now, consider the following decomposition:*

$$\mathbb{1}_{(0,0)} = \frac{\mathbb{1}_{B_1} + \mathbb{1}_{B_1}}{2}$$

where $B_1 = B((-1,0),1)$ *and* $B_2 = B((1,0),1)$*. The subdifferential of the sum is:*

$$\partial(\mathbb{1}_{B_1} + \mathbb{1}_{B_2})((0,0)) = \{\, (\alpha, 0) \mid \alpha \in \mathbb{R} \,\}$$

which is not equal to $\mathbb{R}^2$*.*

Remark (About the composition rule)**.** *Let* $S = \{\, x \in \mathbb{R}^2 \mid x_1 = x_2 \,\}$*, and* $f = \mathbb{1}_S$*. We define:*

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

We do have $f(Ax) = \mathbb{1}_{(0,0)}$*, but the sets* $\partial h(0)$ *and* $A^\top \partial f(0)$ *are not equal. This does not contradict the rule, as the constraint qualification is not satisfied (*$\operatorname{relint} \operatorname{dom} f(A\cdot) = \varnothing$*).*

5.2 Fermat's rule

5.2.1 Definition

Theorem (Fermat's rule)**.** Let $f : \mathbb{R}^n \to \bar{R}$ be any function, and $x \in \mathbb{R}^n$. Then, x is a minimizer of f if and only if $0 \in \partial f(x)$.

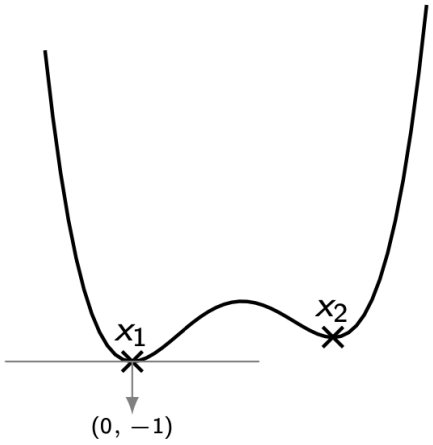
Proof. A point x minimizes f if and only if:

$$\forall y \in \mathbb{R}^n,\ f(y) \geqslant f(x) = f(x) + 0^\top (y - x)$$

which by definition of the subdifferential is equivalent to $0 \in \partial f(x)$. □

Note that Fermat's rule also holds for non-convex functions, even with local minimizers.

Example. Consider the following function:



Then we have a global minimum at x_1:

$$\partial f(x_1) = \{0\} \quad \text{and} \quad \nabla f(x_1) = 0$$

and a local minimum at x_2:

$$\partial f(x_2) = \varnothing \quad \text{and} \quad \nabla f(x_2) = 0$$

In general, it is difficult to check Fermat's rule directly; we need to resort on the structure of the problem. For instance, if the optimization problem involves several functions, such as:

$$\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

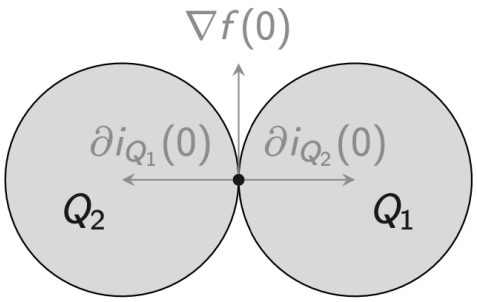
then we can check the optimality of x by verifying that $0 \in \partial f(x) + \partial g(x)$. This also works under constraint qualifications (remember Remark 5.1).

5.2.2 Constraint qualifications

Suppose that we want to solve the following optimization problem:

$$\min_x f(x) + \mathbb{1}_{Q_1}(x) + \mathbb{1}_{Q_2}(x)$$

For instance in $\mathbb{R}^2$, we might have $f((x_1, x_2)) = x_2$, $Q_1 = \{ x \in \mathbb{R}^2 \mid \|x - (1,0)\| \leqslant 1 \}$, and $Q_2 = \{ x \in \mathbb{R}^2 \mid \|x + (1,0)\| \leqslant 1 \}$:



There is no way to cancel the gradient using the sum of the subdifferentials, though $x = 0$. This is because the constraint qualification is not satisfied. We need to directly check the subdifferential $\partial(f(x) + \mathbb{1}_{Q_1} + \mathbb{1}_{Q_2})$.

5.2.3 Optimality conditions

Let $f : \mathbb{R}^m \to \bar{\mathbb{R}}$, $g : \mathbb{R}^n \to \bar{\mathbb{R}}$ be closed, convex functions, and $A \in \mathscr{M}_{m,n}(\mathbb{R})$. We study the following optimization problem:

$$\min_x f(Ax) + g(x) \tag{5.2.1}$$

Theorem (Slater's Constraint Qualification)**.**

$$\exists \tilde{x}, \quad \tilde{x} \in (\operatorname{relint} \operatorname{dom}(f \circ A)) \cap (\operatorname{relint} \operatorname{dom} g) \tag{Slater-CQ}$$

Assuming Slater's constraint qualification, then $x_* \in \mathbb{R}^n$ is a solution to Equation 5.2.1 if and only if x_* satisfies:

$$0 \in A^\top \partial f(Ax_*) + \partial g(x_*)$$

This optimality condition implies that for such x_* there is some $\lambda \in \mathbb{R}^n$ such that:

$$\lambda \in A^\top \partial f(Ax_*) \quad \text{and} \quad -\lambda \in \partial g(x_*)$$

5.3 Fenchel-Legendre conjugation

5.3.1 Conjugate functions

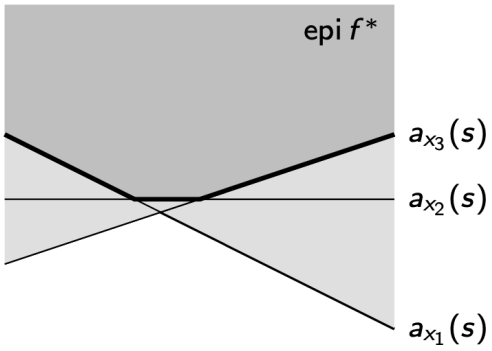
Definition (Conjugate function)**.** The *conjugate function* of $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is the function f^* defined as:

$$f^*(s) := \sup_x \left(s^\top x - f(x) \right) \tag{5.3.1}$$

This definition is implicit via the optimization problem above.

The conjugate f^* is the supremum of a family of affine functions. If we define $a_x(s) = s^\top x - f(x)$ an affine function parametrized by x, then $f^*(s) = \sup_x a_x(s)$.

The epigraph of f^* is the intersection of the epigraphs of the a_x, which results in interesting properties for f^*.

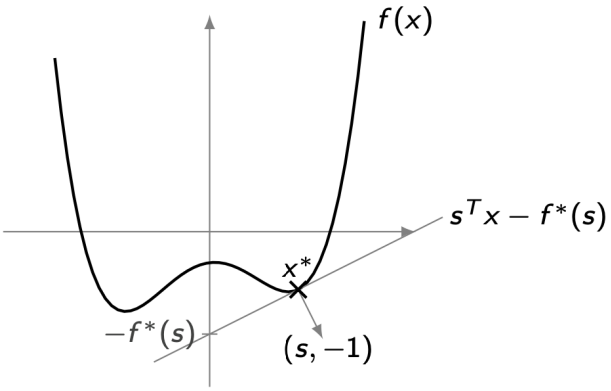


Property 5.4.

- f^* is convex, since its epigraph is the intersection of the convex halfspaces epi a_x
- f^* is closed, since its epigraph is the intersection of the closed halfspaces epi a_x
- f^* is proper if $\partial f(x) \neq \varnothing$ for some $x \in \mathbb{R}^n$

In the following, we will always assume this last hypothesis.

An interpretation of the conjugate $f^*(s)$ is that is defines an affine minorant of f with slope s, where $-f^*(s)$ decides the constant offset to get support.



Formally:

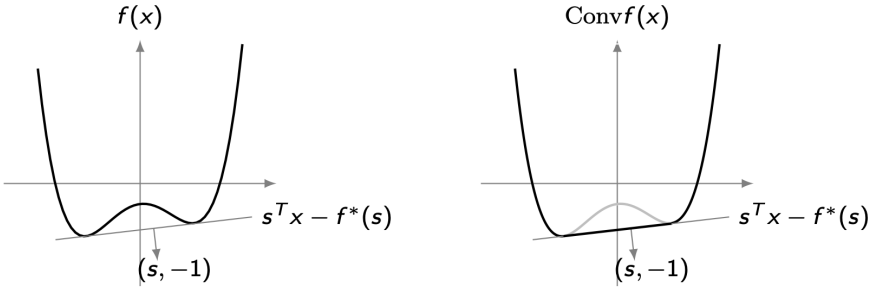
$$\begin{aligned} f^*(s) = \sup_x \left(s^\top x - f(x) \right) &\iff \forall x,\ f^s(s) \geqslant s^\top x - f(x) \\ &\iff \forall x,\ f(x) \geqslant s^\top x - f^*(s) \end{aligned}$$

The maximizing argument x_* gives support: $f(x_*) = s^\top x_* - f^*(s)$. Furthermore, we have $f(x_*) = s^\top x_* - f^*(s)$ if and only if $s \in \partial f(x_*)$.

Property 5.5. The conjugate of f and Conv f[3] are the same, that is:

$$f^* = (\operatorname{Conv} f)^*$$

[3]The *convex closure* of f, noted Conv f, is the function such that epi Conv f = Conv epi f.



Both functions have the same supporting affine functions, and both epigraphs have the same supporting hyperplanes.

Example (Conjugate of the absolute value)**.** We aim at computing the conjugate of $f(x) = |x|$. By definition (Equation 5.3.1):

$$f^*(s) = \sup_{x \in \mathbb{R}} \left(s^\top x - f(x) \right)$$

If we choose $s < -1$, then $f^*(s) = +\infty$. For $-1 \leqslant s \leqslant 1$, $f^*(s) = 0$. For $s > 1$, then $f^*(s) = +\infty$.

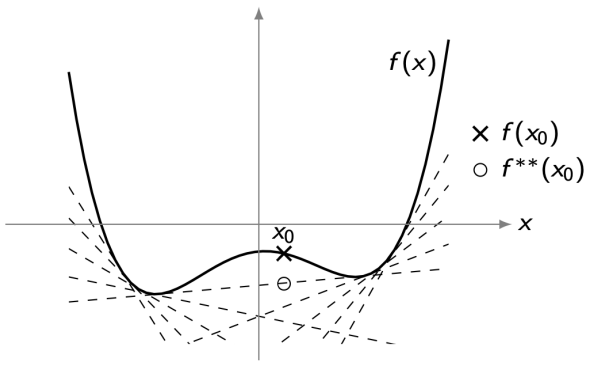
Therefore, $f^*(s) = \mathbb{1}_{[-1,1]}(s)$.

5.3.2 Biconjugate

Definition (Biconjugate)**.** The *biconjugate* of $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is the function f^{**} defined as:

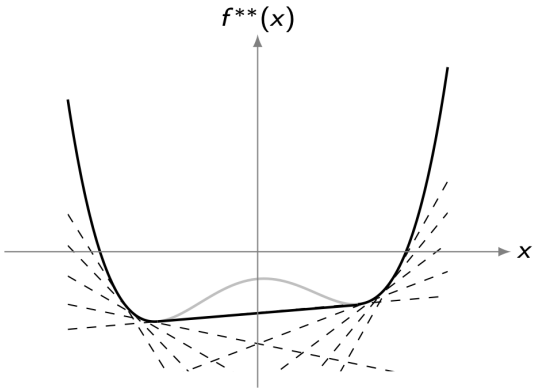
$$f^{**}(x) := \sup_s \left(s^\top x - f^*(s) \right) \tag{5.3.2}$$

For every x, it is the largest value of all affine minorants of f.



Indeed, $x^\top s - f^*(s)$ supports the affine minorant of f with slope s. The biconjugate $f^{**}(x)$ picks the largest value over all these affine minorants evaluated at x.

Property 5.6. The biconjugate f^{**} is the closed convex envelope of f:



Property 5.7. The following inequality always holds:

$$f^{**} \leqslant f$$

with equality if and only if f is closed and convex.

5.3.3 Fenchel-Young's equality

Theorem (Fenchel-Young's inequality)**.** Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function. Then, for all $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$:

$$f(x) \geqslant s^\top x - f^*(s) \tag{5.3.3}$$

with equality if and only if $s \in \partial f(x)$.

Theorem (Fenchel-Young's equality)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a closed, proper, convex function. Then, given $x, s \in \mathbb{R}^n$, the following statements are equivalent:

- $f(x) + f^*(s) = s^\top x$
- $s \in \partial f(x)$
- $x \in \partial f^*(s)$

An equivalent way to write this result is that for a closed, convex function f:

$$\partial f^* = (\partial f)^{-1} \tag{5.3.4}$$

Example (Stricly convex quadratic function)**.** For $Q \succ 0$, we define:

$$f(x) = \frac{1}{2} x^\top Q x$$

Its conjugate is:

$$f^*(s) = \frac{1}{2} s^\top Q^{-1} s$$

and $\nabla f^*(y) = Q^{-1} y$.

Example (Indicator of a point)**.** Let $x_0 \in \mathbb{R}^n$. We consider $\mathbb{1}_{x_0}$, the indicator function of x_0. Its conjugate is:

$$\mathbb{1}^*_{x_0}(s) = a^\top s$$

Their subdifferentials are:

$$\partial \mathbb{1}_{\{x_0\}}(x) = \begin{cases} \{0\} & \text{if } x = x_0 \\ \varnothing & \text{otherwise} \end{cases} \quad \text{and} \quad \partial \mathbb{1}^*_{\{x_0\}}(s) = a$$

Example (Linear function)**.** The conjugate of a linear function is the indicator of a point, the slope of the linear function:

$$f(x) = a^\top x \implies f^*(s) = \mathbb{1}_a(s)$$

5.4 A first approach to duality

In this section, we consider an optimization problem, and we derive its *dual problem* to understand what duality is, and how it can be helpful.

Let $f : \mathbb{R}^m \to \bar{\mathbb{R}}$, $g : \mathbb{R}^n \to \bar{\mathbb{R}}$ be closed, convex functions, and $A \in \mathscr{M}_{m,n}(\mathbb{R})$. We consider the following *primal* optimization problem:

$$p_* = \min_{x \in \mathbb{R}^n} f(Ax) + g(x) \tag{Primal}$$

We can reformulate this problem by decoupling the objective functions:

$$p_* = \min_{x,y} f(y) + g(x) \quad \text{s.t.} \quad y = Ax$$

We introduce the *Lagrangian* $\mathcal{L}$ of this problem:

$$\mathcal{L}(x, y, \lambda) := f(y) + g(x) + \lambda^\top (Ax - y) \tag{5.4.1}$$

for $\lambda \in \mathbb{R}^m$. An equivalent way to write our original problem using the Lagrangian is:

$$p_* = \min_{x,y} \max_{\lambda} \mathcal{L}(x, y, \lambda)$$

Indeed, $\lambda \mapsto \mathcal{L}(x, y, \lambda)$ is an affine function, whose maximum is different from $+\infty$ if and only if the slope of the affine function is null, that is $Ax - y = 0$ (when the constraint is satisfied). Therefore, if the constraint can be satisfied, then the maximum of the Lagrangian is $f(y) + g(x)$, and we obtain the original problem back. If the constraint cannot be satisfied, then the maximum is $+\infty$, and the minimum is $+\infty$ as well.

Notice that by definition of the maximum, we have:

$$p_* \geqslant \min_{x,y} \mathcal{L}(x, y, \lambda)$$

which is called *Lagrangian relaxation*; this provides a lower bound on p_*. Finally, the *dual problem* consists in maximizing this lower bound:

$$p_* \geqslant \max_{\lambda} \min_{x,y} \mathcal{L}(x, y, \lambda) \tag{5.4.2}$$

Writing it out explicitly:

$$\begin{aligned} \max_{\lambda} \min_{x,y} f(y) + g(x) + \lambda^\top (Ax - y) &= \max_{\lambda} \left[\min_x (g(x) + \lambda^\top Ax) + \min_y (f(y) - \lambda^\top y) \right] \\ &= \max_{\lambda} \left[- \max_x (-\lambda^\top Ax - g(x)) - \max_y (\lambda^\top y - f(y)) \right] \end{aligned}$$

Hence, we reach the explicit *Fenchel dual formulation* of the problem:

$$d_* = \max_{\lambda} -g_*(-A^\top \lambda) - f^*(\lambda) \tag{Dual}$$

Property 5.8 (Weak duality)**.** Given our primal problem:

$$p_* = \min_{x \in \mathbb{R}^n} f(Ax) + g(x)$$

and its dual problem:

$$d_* = \max_{\lambda} -g_*(-A^\top \lambda) - f^*(\lambda)$$

we have the following inequality, called *weak duality*:

$$d_* \leqslant p_* \tag{5.4.3}$$

Property 5.9 (Strong duality)**.** If Slater's constraint qualification ((Slater-CQ)) is satisfied, and if there exists x_* such that:

$$0 \in A^\top \partial f(Ax_*) + \partial g(x_*)$$

then we have *strong duality*:

$$d_* = p_* \tag{5.4.4}$$

Proof. Since:

$$0 \in A^\top \partial f(Ax_*) + \partial g(x_*)$$

we can choose λ_* such that:

$$\lambda_* \in \partial f(Ax_*) \quad \text{and} \quad -A^\top \lambda_* \in \partial g(x_*)$$

By Fenchel-Young's equality, this is equivalent to:

$$f^*(\lambda_*) + f(Ax_*) = \lambda_*^\top Ax_* \quad \text{and} \quad g_*(-A^\top \lambda_*) + g(x_*) = -\lambda_*^\top$$

The *primal-dual gap* is then:

$$p_* - d_* = f(Ax_*) + g(x_*) + f^*(\lambda_*) + g_*(-A^\top \lambda_*) = 0$$

□

Why should we care about duality? The dual problem is often easier to solve than the primal problem. Furthermore, the dual problem can provide useful information about the primal problem, such as the optimal value, the optimal solution, or the structure of the optimal solution. If strong duality holds, then we can solve:

$$\min_{\lambda} f^*(\lambda) + g_*(-A^\top \lambda)$$

Example (Quadratic minimization)**.** We consider the primal quadratic minimization problem:

$$\boxed{\min_x \frac{1}{2} x^\top Q x \quad \text{s.t.} \quad Ax \leqslant b}$$

We can reformulate this problem as:

$$\min_{x,y} \frac{1}{2} x^\top Q x + \mathbb{1}_{\mathbb{R}_-}(Ax - b)$$

Its Fenchel dual is:

$$\max_{\lambda} -\frac{1}{2} \lambda^\top A Q^{-1} A^\top \lambda - \mathbb{1}^*_{\mathbb{R}_-}(\lambda) - \lambda^\top b$$

If the set $\{ x \mid Ax \leqslant b \}$ is feasible, then the duality gap is null. We can compute $\mathbb{1}^*_{\mathbb{R}_-}$, the support function of the negative orthant:

$$\partial \mathbb{1}^*_{\mathbb{R}_-}(y) = \begin{cases} \mathbb{R}_+ & \text{if } y = 0 \\ \{0\} & \text{if } y < 0 \\ \varnothing & \text{otherwise} \end{cases}$$

Note that $\partial \mathbb{1}^*_{\mathbb{R}_-} = \partial \mathbb{1}_{\mathbb{R}_+}$, and that $\mathbb{1}^*_{\mathbb{R}_-} = \mathbb{1}_{\mathbb{R}_+}$. Therefore, the dual problem to quadratic minimization is:

$$\boxed{\max_{\lambda \geqslant 0} -\frac{1}{2}\lambda^\top A Q^{-1} A^\top \lambda - \lambda^\top b} \tag{5.4.5}$$

Example (LASSO)**.** Consider the LASSO primal problem:

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 + \alpha \|x\|_1$$

Its dual problem is:

$$\max_\lambda -\frac{1}{2}\|\lambda - b\|_2^2 + \|b\|_2^2 - \mathbb{1}_{\{\|\cdot\| \leqslant 1\}}\left(\frac{A^\top \lambda}{\alpha}\right)$$

It has a null duality gap, since the primal has full domain.

5.5 A second approach to duality: standard forms

5.5.1 Formal definitions

We present another point of view on duality, which is essentially equivalent in spirit, but uses a different parametrization that exploits structure.

Consider the following optimization problem in standard form (which is not necessarily convex):

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} \forall i \in [\![1, m]\!],\ g_i(x) \leqslant 0 \\ \forall j \in [\![1, p]\!],\ h_j(x) = 0 \end{cases} \tag{5.5.1}$$

of variable $x \in \mathbb{R}^n$, over a domain $\mathcal{D}$, with optimal value p^*.

Definition (Lagrangian)**.** The *Lagrangian* of a problem on standard form (5.5.1) is the function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ such that, with $\operatorname{dom} \mathcal{L} = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$, defined by:

$$\mathcal{L}(x, \lambda, \nu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^p \nu_j h_j(x) \tag{5.5.2}$$

It is a weighted sum of objective and constraint functions, where λ_i is the *Lagrange multiplier* associated with $g_i(x) \leqslant 0$, and ν_j is the *Lagrange multiplier* associated with $h_j(x) = 0$.

Definition (Lagrange dual function)**.** The *Lagrange dual function* is the function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \cup \{-\infty\}$ defined by:

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \tag{5.5.3}$$

g is always concave, and can take the value $-\infty$ for some values of λ, ν.

Property 5.10 (Lower bound property)**.** If $\lambda \geqslant 0$, then $g(\lambda, \nu) \leqslant p^*$.

Proof. Consider $\tilde{x}$ feasible and $\lambda \geqslant 0$. Then:

$$f(\tilde{x}) \geqslant \mathcal{L}(\tilde{x}, \lambda, \nu) \geqslant \inf_{x \in \mathcal{D}} \lambda(x, \lambda, \nu) = g(\lambda, \nu)$$

Minimizing this lower bound over all feasible $\tilde{x}$ gives $p^* \geqslant g(\lambda, \nu)$. □

5.5.2 Examples

Least-norm solution of linear equations We consider the following problem:

$$\underset{x}{\text{minimize}}\ x^\top x \quad \text{subject to} \quad Ax = b$$

The Lagrangian is:

$$\mathcal{L}(x, \nu) = x^\top x + \nu^\top (Ax - b)$$

To minimize $\mathcal{L}$ over x, we set its gradient to zero:

$$\nabla_x \mathcal{L}(x, \nu) = 2x + A^\top \nu = 0 \implies x = -\frac{1}{2} A^\top \nu$$

Plugging it into $\mathcal{L}$ to obtain the dual function g:

$$g(\nu) = \mathcal{L}\left(-\frac{1}{2} A^\top \nu, \nu\right) = -\frac{1}{4} \nu^\top A A^\top \nu - b^\top \nu$$

which is a concave function of ν. The lower bound property gives us:

$$\forall \nu, \quad p^* \geqslant -\frac{1}{4} \nu^\top A A^\top \nu - b^\top \nu$$

Standard form LP Recall that a linear programming (LP) problem has the following standard form:

$$\underset{x}{\text{minimize}}\ c^\top x \quad \text{subject to} \quad \begin{cases} Ax = b \\ x \geqslant 0 \end{cases}$$

Its Lagrangian is therefore:

$$\mathcal{L}(x, \lambda, \nu) = c^\top x + \nu^\top (Ax - b) - \lambda^\top x = -b\top \nu + (c + A^\top \nu - \lambda)^\top x$$

which is linear in x. Hence, it has a finite minimum if and only if its slope is null:

$$g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu) = \begin{cases} -b^\top \nu & \text{if } A^\top \nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Note that g is linear on the affine domain $\left\{ (\lambda, \nu) \,\middle|\, A^\top \nu - \lambda + c = 0 \right\}$, and is therefore concave as expected. The lower bound property gives us:

$$p^* \geqslant -b^\top \nu \quad \text{subject to} \quad A^\top \nu + c \geqslant 0$$

Equality constrained norm minimization We consider the following problem in standard form, called *equality constrained norm minimization*:

$$\underset{x}{\text{minimize}}\ \|x\| \quad \text{subject to} \quad Ax = b$$

The dual function of this problem is:

$$g(\nu) = \inf_x \left(\|x\| - \nu^\top A x + b^\top \nu \right) = \begin{cases} b^\top \nu & \text{if } \|A^\top \nu\|_* \leqslant 1 \\ -\infty & \text{otherwise} \end{cases}$$

where $\|v\|_* = \sup_{\|u\| \leqslant 1} u^\top v$ is the dual norm of $\|\cdot\|$. The lower bound property gives us that:

$$p^* \geqslant b^\top \nu \quad \text{subject to} \quad \|A^\top \nu\| \leqslant 1$$

Proof. Let's show that $\inf_x \left(\|x\| - y^\top x \right) = 0$ if $\|y\|_* = 1$, and $-\infty$ otherwise:

- If $\|y\|_* \leqslant 1$, then $\|x\| - y^\top x \geqslant 0$ for any x, with equality if $x = 0$
- If $\|y\|_* > 1$, we choose $x = tu$ where $\|u\| \leqslant 1$ and $u^\top y = \|y\|_* > 1$. Then:

$$\|x\| - y^\top x = t \left(\|u\| - \|y\|_* \right) \xrightarrow[t \to +\infty]{} -\infty$$

□

5.5.3 The dual problem

Definition (Lagrange dual problem)**.** Using the definitions of 5.5.1, the *Lagrange dual problem* is the following optimization problem:

$$\max_{\lambda,\nu} g(\lambda,\nu) \quad \text{such that} \quad \lambda \geqslant 0 \tag{5.5.4}$$

This corresponds to finding the best lower bound on p^* obtained from the Lagrange dual functions.

In general, it is a convex optimization problem; we will denote its optimal value d^*. A couple (λ,ν) is dual feasible if $\lambda \geqslant 0$ and $(\lambda,\nu) \in \operatorname{dom} g$. It is often simplified by making the implicit constraint $(\lambda,\nu) \in \operatorname{dom} g$ explicit.

Example (Standard form LP)**.** The standard form LP problem is:

$$\min c^\top x \quad \text{subject to} \quad Ax = b, x \geqslant 0$$

Its dual problem is:

$$\max -b^\top \nu \quad \text{such that} \quad A^\top \nu + c \geqslant 0$$

Weak duality The following inequality, called *weak duality*, always holds for both convex and non-convex problems:

$$d^* \leqslant p^*$$

It can be used to find nontrivial lower bounds for difficult problems.

Strong duality The following equality, called *strong duality*, holds for convex problems but does not hold in general:

$$d^* = p^*$$

Conditions that guarantee strong duality in convex problems are called *constraint qualifications* (for instance, Slater's constraint qualification).

Theorem (Slater's constraint qualification)**.** For a convex problem of the form:

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

if there exists $\tilde{x}$ that is strictly feasible, that is:

$$\forall i \in [\![1,m]\!],\ g_i(\tilde{x}) < 0 \quad \text{and} \quad A\tilde{x} = b$$

then strong duality holds. Furthermore, the dual optimal value is attained: $p^* > -\infty$.

Example (Inequality form of the LP problem)**.** Consider the following LP problem:

$$\min c^\top x \quad \text{subject to} \quad Ax \leqslant b$$

Its dual function is:

$$g(\lambda) = \inf_x \left((c + A^\top \lambda)^\top x - b^\top \lambda \right) = \begin{cases} -b^\top \lambda & \text{if } A^\top \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Therefore, the dual problem is:

$$\max -b^\top \nu \quad \text{subject to} \quad A^\top \nu + c = 0, \lambda \geqslant 0$$

According to Slater's condition, if there exists $\tilde{x}$ such that $A\tilde{x} < b$, then $p^* = d^*$; in fact, $p^* = d^*$ except when the primal and dual are infeasible.

Example (Quadratic program)**.** Consider the following quadratic program for some $P \in \mathbb{S}_n^{++}(\mathbb{R})$:

$$\min_x x^\top P x \quad \text{subject to} \quad Ax \leqslant b$$

Its dual function is:

$$g(\lambda) = \inf_x \left(x^\top P x + \lambda^\top (Ax - b) \right) = -\frac{1}{4} \lambda^\top A P^{-1} A^\top \lambda - b^\top \lambda$$

Therefore, the dual problem is:

$$\max -\frac{1}{4} \lambda^\top A P^{-1} A^\top \lambda - b^\top \lambda \quad \text{subject to} \quad \lambda \geqslant 0$$

According to Slater's condition, if there exists $\tilde{x}$ such that $A\tilde{x} < b$, then $p^* = d^*$; in fact, we always have $p^* = d^*$.

Property 5.11 (Complementary slackness)**.** Assume that strong duality holds, and that x^* is a primal optimal, and (λ^*, ν^*) is a dual optimal. Then, x^* minimizes $x \to \mathcal{L}(x, \lambda^*, \nu^*)$, and the following *complementary slackness* conditions hold:

$$\forall i \in [\![1, m]\!],\ \lambda_i^* g_i(x^*) = 0 \tag{5.5.5}$$

Said otherwise:

$$\lambda_i^* > 0 \implies g_i(x^*) = 0 \quad \text{and} \quad g_i(x^*) < 0 \implies \lambda_i^* = 0$$

5.5.4 Karush-Kuhn-Tucker conditions

Theorem (Karush-Kuhn-Tucker conditions)**.** Consider an optimization problem in standard form where the g_i and h_i are differentiable. Let x^* be a primal optimal, and (λ^*, ν^*) be a dual optimal. Then, the following necessary conditions hold:

1. **Primal feasibility**: $g_i(x^*) \leqslant 0$ and $h_j(x^*) = 0$
2. **Dual feasibility**: $\lambda^* \geqslant 0$
3. **Complementary slackness**: $\forall i \in [\![1, m]\!],\ \lambda_i^* g_i(x^*) = 0$
4. **Stationarity**:
$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(x^*) = 0$$

If the problem is **convex**, they are also sufficient.

Reciprocally, if $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy the KKT conditions for a **convex problem**, then they are optimal:

- from complementary slackness: $f(\tilde{x}) = \mathcal{L}(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$
- from the 4th condition (and convexity): $g(\tilde{\lambda}, \tilde{\nu}) = \mathcal{L}(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$, hence $f(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$

If Slater's condition is satisfied, then strong duality holds, and the KKT conditions are necessary and sufficient for optimality. This generalizes the optimality conditions $\nabla f(x) = 0$ for unconstrained problems.

6 Duality II

This second part about duality is focused on optimality conditions related to duality: we will introduce conditions for strong duality (with proofs), and multiple examples.

6.1 Examples and interpretations

This part introduces classical optimization problems and provide a step-by-step resolution using duality.

6.1.1 Least-norm solution of linear equations

We consider the following problem, which aims at finding the solution of the equation $Ax = b$ having the smallest possible norm:

$$\underset{x}{\text{minimize}}\ x^\top x \quad \text{subject to} \quad Ax = b$$

Its Lagrangian is $\mathcal{L}(x, \nu) = x^\top x + \nu^\top (Ax - b)$. To obtain g, we need to minimize $\mathcal{L}$ over x for any fixed ν; therefore we set its gradient to zero:

$$\nabla_x \mathcal{L}(x, \nu) = 2x + A^\top \nu = 0 \implies x = -\frac{1}{2} A^\top \nu$$

Which gives us:

$$g(\nu) = \mathcal{L}\left(-\frac{1}{2} A^\top \nu, \nu\right) = -\frac{1}{4} \nu^\top A A^\top \nu - b^\top \nu$$

The lower-bound property is therefore:

$$\forall \nu, \quad p^* \geqslant -\frac{1}{4} \nu^\top A A^\top \nu - b^\top \nu$$

Strong duality holds if the set of solutions $\{\, x \mid Ax = b \,\}$ is non-empty.

6.1.2 Equality constrained norm minimization

We consider a variation of the previous problem, but consider the norm instead of the squared-norm:

$$\underset{x}{\text{minimize}}\ \|x\| \quad \text{subject to} \quad Ax = b$$

We already showed that:

$$g(\nu) = \inf_x \left(\|x\| - \nu^\top A x + b^\top \nu \right) = \begin{cases} b^\top \nu & \text{if } \|A^\top \nu\| \leqslant 1 \\ -\infty & \text{otherwise} \end{cases}$$

where $\|v\|_* = \sup_{\|u\| \leqslant 1} u^\top v$ is the dual norm of $\|\cdot\|$. Like previously, strong duality holds if the set of solutions $\{\, x \mid Ax = b \,\}$ is non-empty.

6.1.3 LP problem in standard form

We already considered the LP problem, whose standard form is:

$$\underset{x}{\text{minimize}}\ c^\top x \quad \text{subject to} \quad \begin{cases} Ax = b \\ x \geqslant 0 \end{cases}$$

Strong duality holds once again if the constrained set is not empty, that is $\{\, x \mid Ax = b, x \geqslant 0 \,\} \neq \varnothing$.

6.1.4 Two-way partitioning

We introduce another problem, *two-way partitioning*:

$$\underset{x}{\text{minimize}}\ x^\top W x \quad \text{subject to} \quad \forall i \in [\![1,n]\!],\ x_i^2 = 1$$

It is a non-convex problem whose feasible set contains 2^n discrete points. An interpretation is that we partition $[\![1,n]\!]$ in two sets, respectively $\{\, i \in [\![1,n]\!] \mid x_i = 1 \,\}$ and $\{\, i \in [\![1,n]\!] \mid x_i = -1 \,\}$. The weights W_{ij} corresponds to the cost of assigning i and j to the same set; reciprocally, $-W_{ij}$ is the cost of assigning i and j to different sets.

The dual function of this problem is given by:

$$g(\nu) = \inf\left(x^\top W x + \sum_i \nu_i (x_i^2 - 1)\right) = \inf_x x^\top (W + \operatorname{diag}(\nu)) x - \mathbf{1}^\top \nu = \begin{cases} -\mathbf{1}^\top \nu & \text{if } W + \operatorname{diag}(\nu) \succcurlyeq 0_n \\ -\infty & \text{otherwise} \end{cases}$$

We can derive the following lower-bound property:

$$W + \operatorname{diag}(\nu) \succcurlyeq 0_n \implies p^* \geqslant -\mathbf{1}^\top \nu$$

For instance, choosing $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives us:

$$p^* \geqslant n\lambda_{\min}(W)$$

Strong duality does not hold a priori.

6.1.5 A non-convex problem with strong duality

Strong duality does not hold only in the case where the problem is convex. An example of such a non-convex problem with strong duality is given by:

$$\underset{x}{\text{minimize}}\ x^\top A x + 2b^\top x \quad \text{subject to} \quad x^\top x \leqslant 1$$

which is non-convex if we do not have $A \succcurlyeq$.

Its dual function is:

$$g(\lambda) = \inf_x \left(x^\top (A + \lambda I) x + 2b^\top x - \lambda\right)$$

The quantity in the $\inf_x$ is unbounded below if we do not have $A + \lambda I \succcurlyeq 0$ or if $A + \lambda I \succcurlyeq 0$ and $b \notin \operatorname{Im}(A + \lambda I)$. Otherwise, it is minimized by $x = -(A + \lambda I)^\dagger b$:

$$g(\lambda) = -b^\top (A + \lambda I)^\dagger b - \lambda$$

Therefore, its dual problem is:

$$\underset{x}{\text{maximize}}\ -b^\top (A + \lambda I)^\dagger b - \lambda \quad \text{subject to} \quad \begin{cases} A + \lambda I \succcurlyeq 0 \\ b \in \operatorname{Im}(A + \lambda I) \end{cases}$$

It can be rewritten as an equivalent SDP of higher dimension:

$$\underset{x}{\text{maximize}}\ -t - \lambda \quad \text{subject to} \quad \begin{bmatrix} A + \lambda I & b \\ b^\top & t \end{bmatrix} \succcurlyeq 0$$

6.1.6 Water-filling

Given some $\alpha_i > 0$ for $i \in [\![1, n]\!]$, consider the following problem:

$$\underset{x}{\text{minimize}} \; -\sum_{i=1}^{n} \log(x_i + \alpha_i) \quad \text{subject to} \quad \begin{cases} x \geqslant 0 \\ \mathbf{1}^\top x = 1 \end{cases}$$

Using the KKT conditions, notice that x is optimal if and only if $x \geqslant 0$, $\mathbf{1}^\top x = 1$, and there exists $\lambda \in \mathbb{R}^n$ and $\nu \in \mathbb{R}$ such that:

$$\lambda \geqslant 0, \quad \lambda_i x_i = 0, \quad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

If $\nu < 1/\alpha_i$, then $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$. If $\nu \geqslant 1$, then $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$. We can therefore determine ν from $\mathbf{1}^\top x = \sum_{i=1}^n \max(0, 1/\nu - \alpha_i) = 1$.

The optimization problem can be interpreted as a modeling of "water-filling": considering n patches, the level of patch i being of height α_i. We flood the area with one unit of water, such that the cumulative height $(x_i + \alpha_i)$ is the same everywhere. Hence, the resulting level is $1/\nu^*$.

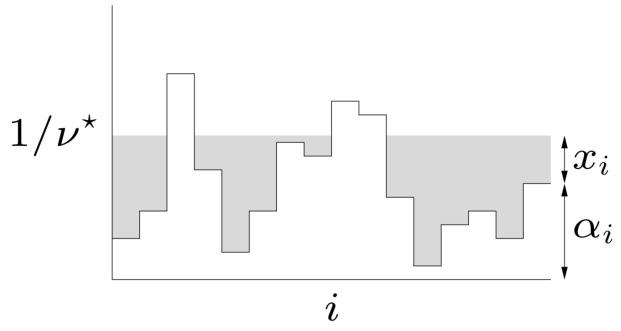


Figure 6.1: Interpretation of the problem as water-filling.

6.2 Perturbations and sensitivity analysis

Consider an (unperturbed) optimization problem:

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

and its dual:

$$\text{maximize} \; g(\lambda, \nu) \quad \text{subject to} \quad \lambda \geqslant 0$$

For parameters u and v, we introduce the *perturbed problem*:

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

and its dual:

$$\text{maximize} \; g(\lambda, \nu) - u^\top \lambda - v^\top \nu \quad \text{subject to} \quad \lambda \geqslant 0$$

The optimal value p^* becomes a function of u and v, such that $p^*(u, v)$ is the optimal value of the problem above. We are interested in information about $p^*(u, v)$ that we can obtain from the solution of the unperturbed problem and its dual.

Assume that strond duality holds for the unperturbed problem, and let (λ^*, ν^*) be the dual optimal of the unperturbed problem. Applying weak duality to the perturbed problem results in:

$$p^*(u, v) \geqslant g(\lambda^*, \nu^*) - u^\top \lambda^* - v^\top \nu^* = p^*(0, 0) - u^\top \lambda^* - v^\top \nu^*$$

This result is called *global sensitivity*. Assuming that $p^*(u, v)$ is differentiable at $(0, 0)$, we can also use *local sensitivity*:

$$\lambda_i^* = -\frac{\partial p^*}{\partial u_i}(0, 0) \quad \text{and} \quad \nu_i^* = -\frac{\partial p^*}{\partial v_i}(0, 0)$$

6.3 Reformulations

Equivalent formulations of a problem can lead to very different duals; reformulating the primal problem can therefore be useful when the dual is difficult to derive, or do not bring any interesting information about the initial primal problem.

There is multiple common ways to reformulate a problem, that often lead to improved duals. Introducing new variables and equality constraints can make the dual more interesting. One can also make explicit some constraints that were implicit, or vice-versa. Finally, it might be useful to transform the objective or cosntraint functions. For instance, replacing $f(x)$ by $\phi \circ f(x)$ with ϕ a convex and increasing function is a trick often used in practice. We will give a few examples to illustrate those methods.

6.3.1 Introducing new variables and equality constraints

Consider a problem of the form:

$$\underset{x}{\text{minimize}} \; f(Ax + b)$$

The dual function is constant, $g = \inf_x \mathcal{L}(x) = \inf_x f(Ax+b) = p^*$, since there are no constraints. In this case, we have strong duality, but the dual problem is quite useless.

We can reformulate it by introducing a new variable, and a matching new equality constraint:

$$\underset{x}{\text{minimize}} \; f(y) \quad \text{subject to} \quad Ax + b - y = 0$$

Since we added a new constraint, the dual problem becomes:

$$\underset{\nu}{\text{maximize}} \; b^\top \nu - f^*(\nu) \quad \text{subject to} \quad A^\top \nu = 0$$

Indeed, the dual function follows from:

$$g(\nu) = \inf_{x,y} \left(f(y) - \nu^\top y + \nu^\top A x + b^\top \nu \right) = \begin{cases} -f^*(\nu) + b^\top \nu & \text{if } A^\top \nu = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Example (Norm approximation problem)**.** The norm approximation problem:

$$\underset{x}{\text{minimize}} \; \|Ax - b\|$$

can be reformulated as:

$$\underset{x}{\text{minimize}} \; \|y\| \quad \text{subject to} \quad y = Ax - b$$

Hence, we can obtain a much more interesting dual problem:

$$\begin{aligned} g(\nu) &= \inf_{x,y} \left(\|y\| - \nu^\top y - \nu^\top A x + b^\top \nu \right) \\ &= \begin{cases} b^\top \nu + \inf_y \left(\|y\| + \nu^\top y \right) & \text{if } A^\top \nu = 0 \\ -\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} b^\top \nu & \text{if } A^\top \nu = 0 \text{ and } \|\nu\|_* \leqslant 1 \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

This finally gives the following dual approximation problem:

$$\text{maximize } b^\top \nu \quad \text{subject to} \quad \begin{cases} A^\top \nu = 0 \\ \|\nu\|_* \leqslant 1 \end{cases}$$

6.3.2 Implicit constraints

In some cases, it is useful to transform explicit constraints into implicit one, by changing the cost function. This allows to keep the dual problem simple, which makes it easier to solve.

Example (LP with box constraints)**.** Consider the following LP problem with *box constraints*:

$$\underset{x}{\text{minimize}}\ c^\top x \quad \text{subject to} \quad \begin{cases} Ax = b \\ -\mathbf{1} \leqslant x \leqslant \mathbf{1} \end{cases}$$

Because of all the constraints, the dual problem is quite complicated:

$$\underset{\nu}{\text{maximize}}\ -b^\top \nu - \mathbf{1}^\top \lambda_1 - \mathbf{1}^\top \lambda^2 \quad \text{subject to} \quad \begin{cases} c + A^\top \nu + \lambda_1 - \lambda_2 = 0 \\ \lambda_1 \geqslant 0 \\ \lambda_2 \geqslant 0 \end{cases}$$

We can reformulate it by introducing the box constraints in the cost function, making them implicit:

$$\underset{x}{\text{minimize}}\ \tilde{f}(x) := \begin{cases} c^\top x & \text{if } -\mathbf{1} \leqslant x \leqslant \mathbf{1} \\ +\infty & \text{otherwise} \end{cases} \quad \text{subject to} \quad Ax = b$$

The dual function becomes:

$$\tilde{g}(\nu) = \inf_{-\mathbf{1} \leqslant x \leqslant \mathbf{1}} \left(c^\top x + \nu^\top (Ax - b) \right) = -b^\top \nu - \|A^\top \nu + c\|_1$$

Hence, the dual problem simply becomes:

$$\underset{\nu}{\text{maximize}}\ -b^\top \nu - \|A^\top \nu + c\|_1$$

6.4 Generalized inequalities

Consider K_i to be pointed, salient and solid convex cones. We previously introduced the generalized inequality $\preccurlyeq_{K_i}$ on $\mathbb{R}^{k_i}$; we would like to generalize duality to such inequalities. In the following, we will consider the normal form of a problem with generalized inequalities to be:

$$\underset{x}{\text{minimize}}\ f(x) \quad \text{subject to} \quad \begin{cases} \forall i \in [\![1, m]\!]\ g(x) \preccurlyeq_{K_i} 0 \\ \forall j \in [\![1, p]\!]\ h(x) = 0 \end{cases}$$

The definitions are similar to the scalar case. The Lagrange multiplier for $g_i(x) \preccurlyeq_{K_i} 0$ becomes a vector $\lambda_i \in \mathbb{R}^{k_i}$.

Definition (Generalize Inequalities Lagrangian)**.** The Lagrangian in the generalized inequalities case is the function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R}$ defined by:

$$\mathcal{L}(x, \lambda_1, \ldots, \lambda_m, \nu) := f(x) + \sum_{i=1}^{m} \lambda_i^\top g_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

Definition (Dual function)**.** The dual function in the generalized inequalities case is the function $g : \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R}$ defined as:

$$g(\lambda_1, \ldots, \lambda_m, \nu) := \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda_1, \ldots, \lambda_m, \nu)$$

Property 6.1 (Lower bound property)**.** If $\lambda_i \preccurlyeq_{K_i^*} 0$, then $g(\lambda_1, \ldots, \lambda_m, \nu) \leqslant p^*$.

Proof. If $\tilde{x}$ is feasible and $\lambda_i \preccurlyeq_{K_i^*} 0$, then:

$$\begin{aligned} f(\tilde{x}) &\geqslant f(\tilde{x}) + \sum_{i=1}^{m} \lambda_i^\top g_i(\tilde{x}) + \sum_{j=1}^{p} \nu_j h_j(\tilde{x}) \\ &\geqslant \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda_1, \ldots, \lambda_m, \nu) \\ &= g(\lambda_1, \ldots, \lambda_m, \nu) \end{aligned}$$

Minimizing this inequality over all feasible $\tilde{x}$ gives us $g(\lambda_1, \ldots, \lambda_m, \nu) \leqslant p^*$. □

Definition (Dual problem)**.** The dual problem in the generalized inequalities case is:

$$\underset{\nu}{\text{maximize}}\ g(\lambda_1, \ldots, \lambda_m, \nu) \quad \text{subject to} \quad \forall i \in [\![1, m]\!],\ \lambda_i \preccurlyeq_{K_i^*} 0$$

Similarly to the real case, weak duality $p^* \geqslant d^*$ always holds, and strong duality $d^* = p^*$ holds for convex problems with constraint qualification.

Example (Semidefinite program)**.** The **Primal SDP** problem for $G_i, H \in \mathbb{S}_k$ is:

$$\underset{x}{\text{minimize}}\ c^\top x \quad \text{subject to} \quad \sum_{i=1}^{n} x_i G_i \preccurlyeq H$$

The Lagrange multiplier is a matrix $Z \in \mathbb{S}_k$. The Lagrangian itself is:

$$\mathcal{L}(x, Z) = c^\top x + \operatorname{Tr}(Z(x_1 G_1 + \cdots + x_n G_n - H))$$

The dual function is:

$$g(Z) = \inf_x \mathcal{L}(x, Z) = \begin{cases} -\operatorname{Tr}(GZ) & \text{if } \forall i \in [\![1, n]\!],\ \operatorname{Tr}(G_i Z) + c_i = 0 \\ -\infty & \text{otherwise} \end{cases}$$

The **Dual SDP** is therefore:

$$\underset{Z}{\text{maximize}}\ -\operatorname{Tr}(HZ) \quad \text{subject to} \quad \begin{cases} Z \succcurlyeq 0 \\ \forall i \in [\![1, n]\!],\ \operatorname{Tr}(G_i Z) + c_i = 0 \end{cases}$$

We have strong duality if the primal SDP is strictly feasible, that is that there exists some x with $\sum_i x_i G_i \prec H$.