Optimization

P. Gourdel[1]

December 9, 2024

[1]gourdel@univ-paris1.fr, University of Paris 1

Chapter 1

Presentation of Optimization

Notations

On the euclidean space $\mathbb{R}^n$, the inner product will be denoted by $\langle u,v\rangle = \sum_{i=1}^n u_i v_i$, the norm is $\|u\| = \sqrt{\langle u,u\rangle}$. We denote by $\overline{B}(x,r)$ and $B(x,r)$ the closed ball and the open ball of center x and radius r.

If A is a subset of $\mathbb{R}^n$, $\operatorname{int} A$ is the intérior of A and $\operatorname{cl} A$ the closure of A. We denote $x \geq y$ (respectively $x > y$, $x \gg y$) if for all $h = 1,\dots,n$ $x_h \geq y_h$ (respectively for all $h = 1,\dots,n$ $x_h \geq y_h$ and there exists at least one indice for which the inequality is strict, respectively for all $h = 1,\dots,n$ $x_h > y_h$).

If f is a linear form from $\mathbb{R}^n$ to $\mathbb{R}$, there exists a unique vector $u \in \mathbb{R}^n$ such that $f(v) = \langle u,v\rangle$ for all $v \in \mathbb{R}^n$.

We denote

$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_1 \geq 0,\dots,x_n \geq 0\} = \{x \in \mathbb{R}^n \mid x \geq 0\}$$
$$\mathbb{R}^n_+ \setminus \{0\} = \{x \in \mathbb{R}^n \mid x > 0\}$$
$$\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n \mid x_1 > 0,\dots,x_n > 0\} = \{x \in \mathbb{R}^n \mid x \gg 0\}$$

1.1 Mathematical presentation

1.1.1 Definitions

Let us consider $f : A \to \mathbb{R}$ (where $A \subset \mathbb{R}$ is the domain of function f) and $C \subset A$. The problem consists in finding the maximum (respectively the minimum of f on C). The function f is the objective function and the set C is the set of feasible points (admissible points), it is often described by a finite list of constraints.

We will note

$$(\mathcal{P}) \max_{x\in C} f(x) \qquad \textit{resp.}\ (\mathcal{Q}) \min_{x\in C} f(x)$$

Definition 1.1 The point $\overline{x}$ is solution of $(\mathcal{P})$ (respectively of $(\mathcal{Q})$) if $\overline{x} \in C$ and if for all x in C, $f(x) \leq f(\overline{x})$ (respectively $f(x) \geq f(\overline{x})$).

Definition 1.2 The point $\overline{x}$ is a local solution of $(\mathcal{P})$ (respectively of $(\mathcal{Q})$) if $\overline{x} \in C$ and if there exists $\varepsilon > 0$ such that for all x in $C \cap B(\overline{x},\varepsilon)$, $f(x) \leq f(\overline{x})$ (respectively $f(x) \geq f(\overline{x})$).

Definition 1.3 We define the value of Problem $(\mathcal{P})$ (respectively of $(\mathcal{Q})$) the supremum (respectively the infimum) of the set $\{f(x) \mid x \in C\}$. This value is either finite or infinite.

If the domain C is empty, we will let by convention $\operatorname{val}(P) = -\infty$ and $\operatorname{val}(\mathcal{Q}) = +\infty$. We will denote $\mathcal{S}ol(P)$ for the set of solutions of $(\mathcal{P})$.

One should distinguish between c a solution of $(\mathcal{P})$ which is a vector of $\mathbb{R}^n$ and $v = f(c)$ the corresponding value which is an element of $[-\infty,+\infty]$. There might exist several solutions c while the value v is unique.

Example 1.1

$$(\mathcal{P}) \max_{x\in\mathbb{R}} \sin x$$

$\text{val}(\mathcal{P}) = 1$ while $\mathcal{S}ol(\mathcal{P}) = \{\pi/2 + 2k\pi \mid k \in \mathbb{Z}\}$

Example 1.2

$$(\mathcal{P}) \max_{x\in\mathbb{R}} \frac{x^2}{1+x^2}$$

$\text{val}(\mathcal{P}) = 1$ while $\mathcal{S}ol(\mathcal{P}) = \emptyset$

Exercise 1.1 (*) Determine the set of solutions, the set of local solutions, and the value of the following problems:

1. $(\mathcal{P}_Y)\begin{cases} \max 1 - x^2 \\ x \in \mathbb{R} \end{cases} \quad (\mathcal{P}_X)\begin{cases} \max 1 - x^2 \\ x \in]2, 3[\end{cases}$

2. $(\mathcal{P}_Z)\begin{cases} \min x^2 + x + 1 \\ x \in \mathbb{R} \end{cases}$

3. $(\mathcal{P}_T)\begin{cases} \min x^2 + x + 1 \\ x \in \mathbb{N} \end{cases}$

Proposition 1.1 *Let $f : A \to \mathbb{R}$. Let us consider the optimization problem $(\mathcal{P})$, $\max_{x\in C} f(x)$ whose value is α, we have*

$$\mathcal{S}ol(\mathcal{P}) = f^{-1}(\{\alpha\}) \cap C \text{ and } \mathcal{S}ol(\mathcal{P}) = \{x \in A \mid f(x) \geq \alpha\} \cap C.$$

In particular, this set is closed in A if f is continuous and C is a closed set.

In particular, if $\text{val}(\mathcal{P}) \notin \mathbb{R}$, then $\mathcal{S}ol(\mathcal{P}) = \emptyset$.

Definition 1.4 If $(\mathcal{P})$ and $(\mathcal{Q})$ are two optimization problems, we say that they are equivalent if their sets of solution are equal (but in general, their values are not equal).

Exercise 1.2 (*) Let f be a function defined on C. Let us suppose that $\varphi : X \to \mathbb{R}$ is an increasing function, where X contains $f(C)$.

$$(\mathcal{P}_1)\begin{cases} \max f(x) \\ x \in C \end{cases} \quad (\mathcal{P}_2)\begin{cases} \max \varphi(f(x)) \\ x \in C \end{cases} \quad (\mathcal{P}_3)\begin{cases} \min -\varphi(f(x)) \\ x \in C \end{cases}$$

1. Prove that the three following problems are equivalents.

2. Prove with a counter-exemple that $(\mathcal{P}_1)$ and $(\mathcal{P}_2)$ are not necessarily equivalent if φ is only non decreasing.

Exercise 1.3 (**) Let f be a function defined on C. Let us suppose that $\varphi : X \to \mathbb{R}$ is an non-decreasing function, where X contains $\text{cl}(f(C))$. We assume in addition that $(\mathcal{P}_1)$ has a finite value.

$$(\mathcal{P}_1)\begin{cases} \max f(x) \\ x \in C \end{cases} \quad (\mathcal{P}_2)\begin{cases} \max \varphi(f(x)) \\ x \in C \end{cases}$$

1. Prove that $\text{val}(\mathcal{P}_2) \leq \varphi(\text{val}(\mathcal{P}_1))$.

2. Prove that if φ is continuous then $\text{val}(\mathcal{P}_2) = \varphi(\text{val}(\mathcal{P}_1))$.

3. Show that if there exists a solution, then $\text{val}(\mathcal{P}_2) = \varphi(\text{val}(\mathcal{P}_1))$.

4. Let us consider $f(x) = x$, $C =]0, 2[$ and $\varphi : R \to \mathbb{R}$ defined by $\varphi(x) = int(x) + x$ where $int(x)$ denotes the ceiling function (the ceiling fonction of x is $\min\{n \in \mathbb{Z} \mid x \leq n\}$). Show that φ is increasing. Compute $\text{val}(\mathcal{P}_1)$, $\text{val}(\mathcal{P}_2)$ and $\varphi(\text{val}(\mathcal{P}_1))$. Prove that $\text{val}(\mathcal{P}_2) < \varphi(\text{val}(\mathcal{P}_1))$

Exercise 1.4 (*) Let $X \subset Y$ be two subsets of $\mathbb{R}^n$ and f be a function from X to $\mathbb{R}$. Let us consider two optimization problems:

$$(\mathcal{P}_X)\begin{cases} \max f(x) \\ x \in X \end{cases} \quad (\mathcal{P}_Y)\begin{cases} \max f(x) \\ x \in Y \end{cases}$$

1. Show that $\text{val}(\mathcal{P}_X) \leq \text{val}(\mathcal{P}_Y)$.

2. If $\overline{y} \in \mathcal{S}ol(\mathcal{P}_Y)$, and if $\overline{y} \in X$, show that $\overline{y} \in \mathcal{S}ol(\mathcal{P}_X)$. This means :

$$\mathcal{S}ol(\mathcal{P}_Y) \cap X \subset \mathcal{S}ol(\mathcal{P}_X)$$

3. If $\overline{x} \in \mathcal{S}ol(\mathcal{P}_X)$, and if $\text{val}(\mathcal{P}_X) = \text{val}(\mathcal{P}_Y)$, then prove that $\overline{x} \in \mathcal{S}ol(\mathcal{P}_Y)$. Deduce that

$$\text{val}(\mathcal{P}_X) = \text{val}(\mathcal{P}_Y) \Rightarrow \mathcal{S}ol(\mathcal{P}_X) \subset \mathcal{S}ol(\mathcal{P}_Y)$$

4. We assume that for all y in the set Y, there exists $x \in X$ such that $f(x) \geq f(y)$, show that $\mathrm{val}(\mathcal{P}_X) = \mathrm{val}(\mathcal{P}_Y)$.

5. In the following example, show that $\mathcal{S}ol(\mathcal{P}_Y) \not\subset \mathcal{S}ol(\mathcal{P}_X)$.

$$(\mathcal{P}_X) \left\{ \begin{array}{l} \max 1 - x^2 \\ x \in]2, 3[\end{array} \right. \qquad (\mathcal{P}_Y) \left\{ \begin{array}{l} \max 1 - x^2 \\ x \in \mathbb{R} \end{array} \right.$$

6. Same question for $\mathcal{S}ol(\mathcal{P}_X) \not\subset \mathcal{S}ol(\mathcal{P}_Y)$.

$$(\mathcal{P}_X) \left\{ \begin{array}{l} \max e^x(\sin x + 2) \\ x \in [0, \pi] \end{array} \right. \qquad (\mathcal{P}_Y) \left\{ \begin{array}{l} \max e^x(\sin x + 2) \\ x \in \mathbb{R} \end{array} \right.$$

Exercise 1.5 (*) Let X be a nonempty subset of $\mathbb{R}^n$ and f, g be continuous functions from X to $\mathbb{R}$. Let us consider two optimization problems:

$$(\mathcal{P}_X) \left\{ \begin{array}{l} \max f(x) \\ x \in X \\ g(x) < 0 \end{array} \right. \qquad (\mathcal{P}_Y) \left\{ \begin{array}{l} \max f(x) \\ x \in X \end{array} \right.$$

Let us assume that $\overline{x}$ is a local solution of $\mathcal{P}_Y$ that satisfies $g(\overline{x}) < 0$. Prove that $\overline{x}$ is a local solution of $\mathcal{P}_X$.

Exercise 1.6 (**) Let $(\mathcal{P})$ be the optimization problem $\max_{x \in C} f(x)$. Let us define the set

$$D = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} \mid x \in C \text{ and } y \leq f(x)\}$$

Let us define g on $\mathbb{R}^n \times \mathbb{R}$ par $g(x, y) = y$ and the optimization problem $(\mathcal{Q})$, $\max_{(x,y) \in D} g(x, y)$.

1. Show that $(\mathcal{P})$ and $(\mathcal{Q})$ have the same value.

2. Show that if $\overline{x} \in \mathcal{S}ol(\mathcal{P})$, then $(\overline{x}, f(\overline{x})) \in \mathcal{S}ol(\mathcal{Q})$.

3. Prove that if $(\overline{x}, \overline{y}) \in \mathcal{S}ol(\mathcal{Q})$, then $\overline{x} \in \mathcal{S}ol(\mathcal{P})$, and $\overline{y} = f(\overline{x})$.

4. Are the two problems equivalent ?

Definition 1.5 Let us consider the optimization problem $(\mathcal{P})$, $\max_{x \in C} f(x)$ (respectively $\min_{x \in C} f(x)$. The sequence $(x_k)_k$ is said to be a maximizing (respectively minimizing) sequence for $(\mathcal{P})$ if for all k, $x_k \in C$ and if the limit of $f(x_k)$ exists (either finite or infinite) and is equal to the value of the problem $(\mathcal{P})$.

Exercise 1.7 (*) Let us consider the following optimization problem where α is a given parameter:

$$(\mathcal{P}_X) \left\{ \begin{array}{l} \min \alpha x^2 \\ x \in \mathbb{R} \end{array} \right.$$

1) If $\alpha > 0$, determine a minimizing sequence.
2) If $\alpha = 0$, determine a minimizing sequence.
3) If $\alpha < 0$, does there a minimizing sequence.

Exercise 1.8 (**) Let X be a nonempty subset of $\mathbb{R}^n$, and f be a function from X to $\mathbb{R}$. Let us consider the following optimization problem:

$$(\mathcal{P}_X) \left\{ \begin{array}{l} \max f(x) \\ x \in X \end{array} \right.$$

1. Prove that there exists a maximizing sequence $(x_k)_k$, i.e. a sequence of elements in X such that $f(x_k) \to \max_{x \in X} f(x)$.

2. Prove that there exists a maximizing sequence $(y_k)_k$ such that $f(y_k) \leq (f(y_{k+1})$.

3. Let us suppose moreover that $\mathcal{S}ol(\mathcal{P}) = \emptyset$, prove that there exists a maximizing sequence $(z_k)_k$ such that $f(z_k) < (f(z_{k+1})$.

1.1.2 Geometric interpretation

Depending wether the optimal point belongs or not to the boundary of the set of feasible points, we get using the level sets two kinds of pictures.

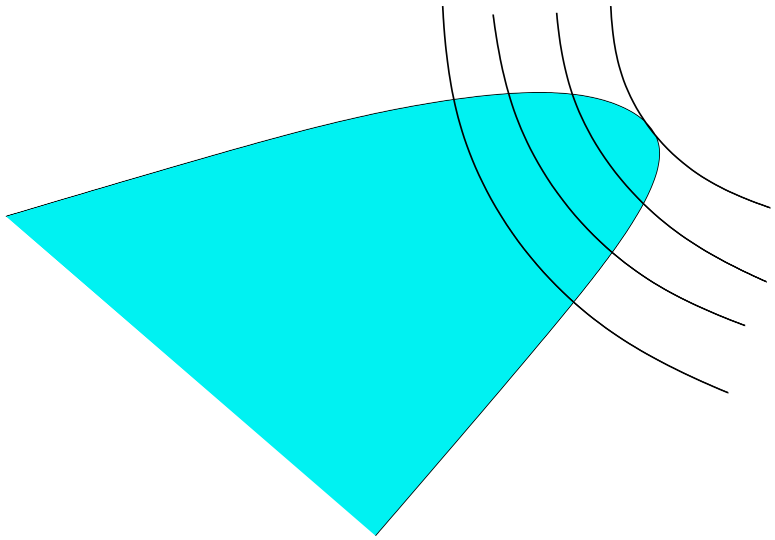

Figure 1.1: The optimal point belongs to the boundary

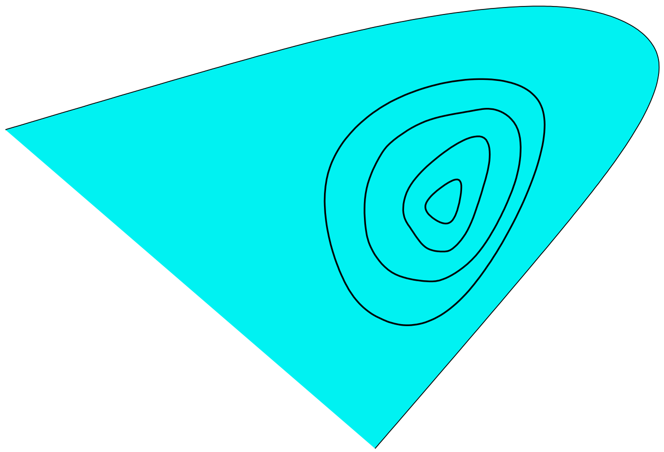

Figure 1.2: The optimal point belongs to the interior

1.2 Examples of economic problems

1.2.1 Consumer theory

In microeconomics, we suppose that u is a utility function from $\mathbb{R}^\ell_+$ to $\mathbb{R}$. let us give a price vector $p = (p_1, \ldots, p_\ell)$ and a wealth $w \geq 0$. The consumer's demand is the set of solutions of

$$(\mathcal{P}_X) \begin{cases} \max u(x) \\ p_1 x_1 + \ldots + p_\ell x_\ell \leq w \\ x_1 \geq 0, \ldots, x_\ell \geq 0 \end{cases}$$

1.2.2 Producer Theory

In microeconomics, we consider a firm which produces the good ℓ, using goods $(1, \ldots, \ell - 1)$ as inputs. We describe the production set with f, production function from $\mathbb{R}^{\ell-1}_+$ to $\mathbb{R}$. Let us give a price vector $p = (p_1, \ldots, p_{\ell-1})$ of the inputs and a level of production $y_\ell \geq 0$, The cost function $c((p_1, \ldots, p_{\ell-1}), y_\ell)$ of the firm is the value fo the problem

$$(\mathcal{P}) \begin{cases} \min p_1 y_1 + \ldots + p_{\ell-1} y_{\ell-1} \\ y_\ell = f(y_1, \ldots, y_{\ell-1}) \\ y_1 \geq 0, \ldots, y_{\ell-1} \geq 0 \end{cases}$$

The firm's demand of inputs corresponds to the set of solutions of $(\mathcal{P})$. In addition, if we consider the price p_ℓ of the unique output, the total offer (with usual signs' convention) of the firm is the set of solutions of

$$(\mathcal{Q}) \begin{cases} \max p_\ell y_\ell - c((p_1, \ldots, p_{\ell-1}), y_\ell) \\ y_\ell \geq 0 \end{cases}$$

1.2.3 Finance theory

In finance, there are S possible states of the world tomorrow with the corresponding probabilities $\pi_1, \ldots \pi_S$. Today, we can buy or sell J assets with price $q_1, \ldots, q_J$. If we own one unit of asset j, we will receive if state s occurs, the amount (possibly negative) a^j_s. The investor will try to maximize the expected value of his stochastic income with respect to his initial capital w. He will buy a portfolio $(z_1, \ldots, z_J)$, solution of

$$\begin{cases} \max \sum_{s=1}^{S} \pi_s \sum_{j=1}^{J} a_s^j z_j \\ \sum_{j=1}^{J} q_j z_j \leq w \end{cases}$$

1.2.4 Statistics

In statistics, we determine an estimator using the maximum of the likehood, and we determine the regression's lines by minimizing the sum of the squares of the "distance to the line" among all possible lines.

1.2.5 Transportation problems

Let us consider a firm with m units of production $P_1, \ldots, P_m$, which produce quantities $q_1, \ldots, q_m$ of a certain good. There are n markets $M_1, \ldots, M_n$ to provide whose respective demands are $\delta_1, \ldots, \delta_n$. In order to transport one unit of good from the the unity i to the market j, there is a cost γ_{ij}. We try to provide all the markets at the lowest transportation cost. We have to determine all the flows x_{ij} (quantity moved from P_i to M_j) solution of

$$\begin{cases} \min \sum_{i=1}^{m} \sum_{j=1}^{n} \gamma_{ij} x_{ij} \\ \sum_{i=1}^{m} x_{ij} \geq \delta_j \text{ for all } j, \\ \sum_{j=1}^{n} x_{ij} \leq q_i \text{ for all } i, \\ x_{ij} \geq 0, \text{ for all } i \text{ and all } j \end{cases}$$

1.2.6 Constant returns to scale

Let us consider a firm using m processes $P_1, \ldots, P_m$ of production. The process P_j is characterized by a vector $\alpha^j \in \mathbb{R}^\ell$. For a single level of activity, there are α_h^j units of good h produced by the firm if $\alpha_h^j \geq 0$ and α_h^j units of good h used by the firm in the process if $\alpha_h^j \leq 0$. The total amount of activity of Process P_j will be denoted by $x_j \geq 0$.

A first class of problem consists in furnishing the demand at minimal cost. There are ℓ markets (one for each good) with respective demands $\delta_1, \ldots, \delta_n$ and the marginal cost of Process P_j is γ_j. The problem consists in determining all activity levels x_j solutions of

$$\begin{cases} \max \sum_{j=1}^{m} \gamma_j x_j \\ \sum_{i=1}^{m} \alpha_h^j x_j \geq \delta_i \text{ for all } i, \\ x_j \geq 0, \text{ for all } j \end{cases}$$

A second class of problem consists in maximizing the income. We assume that the planer owns an initial stock $\sigma_1, \ldots, \sigma_\ell$ of inputs and that the marginal income of process j is r_j. The problem consists in determining all activity levels x_j solutions of

$$\begin{cases} \max \sum_{j=1}^{m} r_j x_j \\ \sum_{i=1}^{m} \alpha_h^j x_j \leq \sigma_h \text{ for all } h = 1, \ldots, \ell, \\ x_j \geq 0, \text{ for all } j \end{cases}$$

Chapter 2

Convexity of sets

2.1 Definition

2.1.1 Definition of a convex set

Let E be a vector space, for all couple of elements (x, y) of E, we will denote by $[x, y]$, the "segment" which is the subset of E defined by

$$[x, y] = \{tx + (1 - t)y \mid t \in [0, 1]\}$$

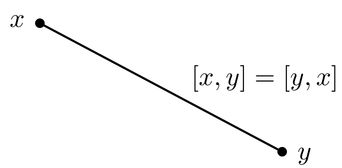

Figure 2.1: segment

Definition 2.1 A subset C of E is convex if for all $(x, y) \in C \times C$, the set $[x, y]$ is contained in C.

Examples: For all couple of elements (x, y) in E, $[x, y]$ is a convex subset of E. All vector subspace is affine (see the next part), all affine subspace is convex. All open (respectively closed) balls are convex. All set of solutions of linear system (involving equalities, large or strict inequalities) is convex, in particular any affine half-space. If $E = \mathbb{R}$, we can characterize the convex subsets which are the intervals.

Exercise 2.1 (*) We recall that a set J included in $\mathbb{R}$ is an intervall if it satisfies the following property :

$$\forall x \in J,\ \forall y \in J,\ \forall z \in \mathbb{R},\quad x \leq z \leq y \Rightarrow z \in J.$$

Show that every compact interval of $\mathbb{R}$ is a line segment of $\mathbb{R}$ and that the only nonempty convex subsets of $\mathbb{R}$ are the intervals.

Exercise 2.2 (*) Use Definition 2.1 in order to show that the following sets are convex:

a) $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 4\}$,

b) $\{(x, y) \in \mathbb{R}^2 \mid |x| + |y| \leq 2\}$,

c) $\{(x, y) \in \mathbb{R}^2 \mid \max\{|x|, |y|\} \leq 2\}$,

d) the set of the $(n \times n)$-matrices with elements ≥ 0.

Note that there is no notion of concave set. It is important to emphasize that it not a topologic concept, it can be defined even if the vector space is not embedded with a topology. Among convex sets, some of them are open, closed or even neither closed nor open.

Exercise 2.3 (**) Let A be a subset of $\mathbb{R}^n$. We will introduce the following property

$$\text{Forall } x \in A,\ y \in A,\ \tfrac{1}{2}(x + y) \in A \qquad (\mathcal{P})$$

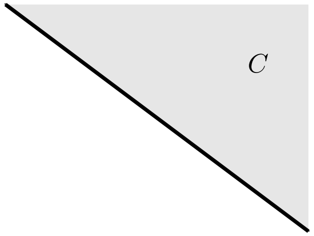

Figure 2.2: A convex and closed half space

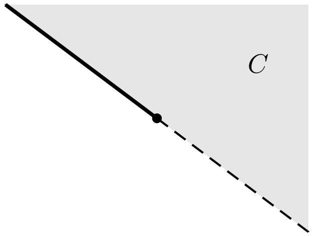

Figure 2.3: A convex half space which is neither closed nor open

1. Prove that this property is satisfied when A is convex.

2. Prove that $\mathbb{Q}$ satisfies this property though it is not a convex set.

3. We assume that A is a closed set that satisfies Property $(\mathcal{P})$. We want to prove that it is a convex set. Let us fix x and y in A, and let us introduce the set

$$J = \{t \in [0,1], tx + (1-t)y \in A\}.$$

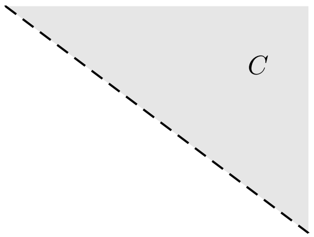

Figure 2.4: A convex and open half space

   a) Prove that J is a closed set containing both 0 and 1.

   b) Prove that when t and t' are in J, then $(t+t')/2$ is in J.

   c) Let us define for each $p \in \mathbb{N}$, the set J_p

   $$J_p = \left\{\frac{k}{2^p} \mid k \in \mathbb{N},\ k \leq 2^p\right\}$$

   Prove by induction that for each $p \in \mathbb{N}$, $J_p \subset J$. (**Hint:** note that $(1/2)(J_p + J_p) = J_{p+1}$.

   d) Prove that $J = [0,1]$ and that A is convex.

4. Summarize the exercise.

Exercise 2.4 (**) In $\mathbb{R}^2$, consider the triangle with vertices x^0, x^1, x^2 (non-collinear points).

1. Prove that for all $u \in \mathbb{R}^2$, there exists a unique $(\lambda_0, \lambda_1, \lambda_2) \in \mathbb{R}^3_+$ such that $u = \lambda_0 x_0 + \lambda_1 x_1 + \lambda_2 x_2$ and $\lambda_0 + \lambda_1 + \lambda_2 = 1$. The coefficients $\lambda_0, \lambda_1, \lambda_2$ are called the barycentric coordinates of u with respect to x^0, x^1, x^2.

2. Let y be the midpoint of the side opposite x^0 and let z be the intersection of the three medians. What are the barycentric coordinates (with respect to $\{x^0, x^1, x^2\}$) of respectively: x^0, x^1, x^2, y, z?

2.1.2 Definition of an affine subspace

Definition 2.2 A subset A of E is affine if for all couple of distinct points of A, the line defined by those two points is still in A. Formally, if for all $(x,y) \in A \times A$, the set $= \{tx + (1-t)y \mid t \in \mathbb{R}\}$ is included in A.

Note that the empty set is affine. It is easy to check that any translation of a vector subspace is an affine set. Exercise 2.8 will show that the converse is true when the set is nonempty.

Let us introduce as a complement the definition of a strictly convex set (the definition given here is not the most general).

Definition 2.3 Let C be a subset of E with a nonempty interior, C is strictly convex if for all $(x, y) \in \operatorname{cl} C \times \operatorname{cl} C$, such that $x \neq y$, and for all $\lambda \in]0, 1[$, $\lambda x + (1 - \lambda)y \in \operatorname{int} C$.

Exercise 2.5 (**) Let C be a subset contained dans E.

1. Prove that C is convex if and only if for all $(x, y) \in C \times C$, such that $x \neq y$, and for all $\lambda \in]0, 1[$, $\lambda x + (1 - \lambda)y \in C$.

2. Prove that if C is strictly convex, then C is convex.

3. If C is strictly convex and D be a subset of E such that $\operatorname{int}(C) \subset D \subset \operatorname{cl}(C)$, prove that D is convex.

4. Give an example of C and D, subsets of $\mathbb{R}^2$, such that C is convex, $\operatorname{int}(C) \subset D \subset \operatorname{cl}(C)$ but D is not convex.

2.1.3 Definition of the unit simplex of $\mathbb{R}^n$

Let us now introduce a very important example of convex set in $\mathbb{R}^n$, the unit simplex (or "usual simplex") denoted by S^{n-1} which is defined by:

$$S^{n-1} = \{\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n_+ \mid \sum_{i=1}^{n} \lambda_i = 1\}$$

We can remark that S^{n-1} is convex, closed and bounded, consequently it is a compact set.

2.2 First properties

Definition 2.4 Let $(x_i)_{i=1}^k$, be k points of $\mathbb{R}^n$. A convex combination (of length k) of $(x_i)_{i=1}^k$ is an element x of $\mathbb{R}^n$ such that there exists $\lambda \in S^{k-1}$ satisfying $x = \sum_{i=1}^k \lambda_i x_i$.

Exercise 2.6 (*) If (x, y) is a couple of elements in $\mathbb{R}^n$, show that the set of convex combinations of x and y is $[x, y]$.

Definition 2.5 Let $(x_i)_{i=1}^n$, be n points of $\mathbb{R}^n$. An affine combination of $(x_i)_{i=1}^n$ is an element x of $\mathbb{R}^n$ such that there exists $\lambda \in \mathbb{R}^n$ satisfying $x = \sum_{i=1}^n \lambda_i x_i$ and $\sum_{i=1}^n \lambda_i = 1$.

Proposition 2.1 *Let C be a subset of E. The set C is convex if and only if C contains all the convex combinations of finite families of elements of C.*

Proof of Proposition 2.1. It is obvious that if C contains all the convex combinations of finite families of elements of C then C is a convex subset of C.

Reciprocally, we prove the result by induction on the cardinal of the family. if the family has one or two elements, the definition of a convex subset show that all convex combination of this family remain in C. Let us assume that this is true for all the families that have at most n elements. Let $(x_1, \dots, x_n, x_{n+1})$ be a family of elements of C. Let $\lambda \in S^n$ and let $x := \sum_{i=1}^{n+1} \lambda_i x_i$. Since $\sum_{i=1}^{n+1} \lambda_i = 1$, there exists at least one λ_i which is not equal to 0. Let us assume with no loss of generality that $\lambda_1 \neq 0$. Then

$$x = \left(\sum_{i=1}^{n} \lambda_i\right) \left(\sum_{i=1}^{n} \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} x_i\right) + \lambda_{n+1} x_{n+1}$$

using the induction hypothesis, if we define

$$x' = \sum_{i=1}^{n} \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} x_i = \sum_{i=1}^{n} \mu_i x_i,$$

where $\mu_i = \dfrac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}$, we can remark that x' is an element of C since it is a convex combination of a family of n elements of C. In order to conclude, if we let $\lambda = \sum_{i=1}^n \lambda_i$, $\lambda \in [0, 1]$, this allows us to see x as a convex combination of x' and x_{n+1} since $x = \lambda x' + (1-\lambda)x_{n+1}$. Consequently x is in C, which ends the proof. □

Proposition 2.2 *Let A be a subset of E. The set A is affine if and only if A contains all the affine combinations of finite families of elements of A.*

Exercise 2.7 (*) Show Proposition 2.2.

Exercise 2.8 (*) Let A be a nonempty affine subset of E (vector space) and $a \in A$.

- For any a in A, let us define the translated set $B_a := A - \{a\} = A + \{-a\}$. This can reformulated as, $x \in B_a$ if and only if $x + a \in A$. Check that $0 \in B_a$.
- Deduce from Proposition 2.2 that B_a is a vector subspace of E.
- Deduce from Proposition 2.2 for if a and $\overline{a}$ are in A, then $B_a = B_{\overline{a}}$, which means that the set B is independent of the particular choice of a.

  Consequently, the set B is called the direction of A and we call affine the dimension of A the dimension of B as a vector space. We do not care about the dimension of the emptyset.

2.2.1 Stability by intersection

Note that the union of two convex sets is not convex.

Proposition 2.3 *Let E be a vector space and let $(C_i)_{i\in I}$ be a family of convex subsets of E. Then $\cap_{i\in I} C_i$ is convex.*

Exercise 2.9 (**) Let $(C_i)_{i\in\mathbb{N}}$ (respectively $(D_i)_i$) be a family of convex subsets of some vector space E.

1) If for all integer i, $C_i \subset C_{i+1}$, then $\cup_{i\in\mathbb{N}} C_i$ is convex.

2) Show that $\bigcup_{k=0}^{\infty} \bigcap_{j=k}^{\infty} D^j$ is a convex set.

Exercise 2.10 (**) Let $(C_i)_{i\in I}$ be a family of convex subsets of some vector space E.

Let us now assume that I is any set of indices and that the family of convex subsets satisfies that for all $(i,j) \in I \times I$, there exists $k \in I$ such that $C_i \cup C_j \subset C_k$. then $\cup_{i\in I} C_i$ is convex.

2.2.2 Stability by the sum operation

Proposition 2.4 *Let E be a vector space. Let $(C_i)_{i\in I}$, be a finite family of convex subsets of E. Then*

$$\sum_{i\in I} C_i := \{\sum_{i\in I} c_i \mid (c_i) \in \prod_{i\in I} C_i\}$$

is a convex subset of E.

The proof is left to the reader, it is worth to notice that for a convex set $2C = C + C$ and that this is not true in general.

Exercise 2.11 (**) We recall that if X is a subset of some vector space E, $2X = \{z \in E \mid \exists x \in X,\ z = 2x\}$.

1. Let C be a convex subset of E. Prove that $2C = C + C$.
2. Let us consider in $\mathbb{R}^2$, the set $A = \{(x,y) \in \mathbb{R}^2 \mid xy = 0\}$, draw A, prove that $2A = A$, $A + A = \mathbb{R}^2$, and deduce that $2A \neq A + A$.

2.3 Stability with respect to affine functions

Definition 2.6 Let A and B be two affine subspace, and $f : A \to B$. The mapping f is affine if for all couple of points (x,y) of A, and for all $t \in \mathbb{R}$,

$$f(tx + (1-t)y) = tf(x) + (1-t)f(y).$$

Exercise 2.12 (**) Let A and B be two affine spaces with respective directions E and F. Let f $A \to B$ and $a \in A$.

1. Prove that f is an affine mapping if and only if there exists a linear mapping $\varphi : E \to F$ such that for all x of A, $f(x) = f(a) + \varphi(x - a)$.

2. Show moreover that φ does not depend on the choice of a.

Consequently, φ is called the linear mapping associated to the affine mapping f.

Proposition 2.5 *Let E be a vector space.*

(i). Let C be a convex subset of E and let $\alpha \in \mathbb{R}$. Then $\alpha C := \{\alpha c \mid c \in C\}$ is convex in E.

(ii). Let f, be an affine mapping from E to some affine subspace A (contained in some vector space F), and let C be a convex subset of E. Then $f(C)$ is a convex subset of F.

The proof is left to the reader.

Proposition 2.6 *Let f be an affine mapping from A to B. Let us assume that the affine subspace A is contained in the vector space E while the affine subspace B is contained in the vector space F. Let C be a convex subset convex of F. Then $f^{-1}(C)$ is a convex subset de E.*

The proof is left to the reader.

Proposition 2.7 *Let $(C_i)_{i\in I}$ be a finite family of sets such that each C_i is a convex subset of the vector space E^i. Then $\prod_{i\in I} C_i$ is a convex subset of $\prod_{i\in I} E^i$.*

The proof is left to the reader.

2.4 Convex hull, affine hull and conic hull

2.4.1 Convex hull and affine hull

Definition 2.7 Let A be a subset of some vector space E.

(i). The convex hull of A, denoted by $\mathrm{co}(A)$, is the intersection of all convex subsets of E containing A.

(ii). The affine hull of A, denoted by $\mathrm{Aff}(A)$, is the intersection of all affine subspaces of E containing A.

Remark 2.1 Note that it follows from the definition that $\mathrm{co}(\emptyset) = \emptyset$ and that $\mathrm{Aff}(\emptyset) = \emptyset$. Since E is convex (and affine), if A is nonempty, $\mathrm{co}(A)$ (respectively $\mathrm{Aff}\, A$) is well defined and nonempty. Since the set of convex subsets (respectively affine subsets) is stable by intersection, $\mathrm{co}(A)$ (respectively $\mathrm{Aff}(A)$) is the smallest (in the sense of inclusion) convex (respectively affine) subset containing A. this implies for example that if C is a convex subset containing A, then $\mathrm{co}(A) \subset C$.

Proposition 2.8 *Let A be a subset of some vector space E.*

(i). The set $\mathrm{co}(A)$ is the set of all convex combinations of elements from A.

(ii). The set $\mathrm{Aff}(A)$ is the set of all affine combinations of elements from A.

Proof of Proposition 2.8. (i) Let us denote by B (respectively D), the set of all convex combinations of elements from A (respectively $\mathrm{co}(A)$). One has $B \subset D$ since $A \subset (\mathrm{co}(A))$. Moreover, we can apply Proposition 2.1 together with the convexity of $\mathrm{co}(A)$ (cf. the previous remark) in order to get $D = \mathrm{co}(A)$. Consequently $B \subset \mathrm{co}(A)$.

Let us now show that $\mathrm{co}(A) \subset B$. It is clear that $A \subset B$. Hence, in order to prove this inclusion, il suffices to prove that B is convex. Let x and y, be two elements of B and $t \in [0, 1]$. By definition of B, there exists two families $(x_1 \dots, x_n)$ and $(y_1, \dots, y_p)$ of elements from A, $\lambda \in S^{n-1}$ and $\mu \in S^{p-1}$ such that $x = \sum_{i=1}^n \lambda_i x_i$ and $y = \sum_{j=1}^p \mu_j y_j$. One has

$$tx + (1-t)y = \sum_{i=1}^n t\lambda_i x_i + \sum_{j=1}^p (1-t)\mu_j y_j.$$

Then $tx + (1-t)y$ is a convex combination of $(x_1, \ldots, x_n, y_1 \ldots, y_p)$ since

$$\sum_{i=1}^{n} t\lambda_i + \sum_{j=1}^{p}(1-t)\mu_j = t + (1-t) = 1$$

and therefore, $(t\lambda_1, \ldots, t\lambda_n, (1-t)\mu_1 \ldots, (1-t)\mu_p)$ belongs to S^{n+p-1}.

(ii) Similar to first part. □

Definition 2.8 Let us call (convex) polytope, the convex hull of a finite set.

Example 2.1 Let us define the simplex (denoted by S^{n-1}) as the subset of $\mathbb{R}^n$ which is the polytope generated by the elements of the canonic basis. The affine subspace generated by $\mathrm{Aff}(S^{n-1})$, is the hyperplane $\{x \in \mathbb{R}^n \mid x_1 + \ldots + x_n = 1\}$.

Exercise 2.13 (**) Let U and V be two subsets of $\mathbb{R}^n$. We will denote by $\overline{\mathrm{co}}\, U$ the closure of the convex set U.
Show that: $\mathrm{co}\, U + \mathrm{co}\, V \subset \mathrm{co}(\mathrm{co}\, U + V)$; $\mathrm{co}\, U + V \subset \mathrm{co}(U + V)$.
Prove then that $\mathrm{co}(U + V) = \mathrm{co}\, U + \mathrm{co}\, V$.
If $\overline{\mathrm{co}}\, U$ is compact, show that $\overline{\mathrm{co}}(U + V) = \overline{\mathrm{co}}\, U + \overline{\mathrm{co}}\, V$.

Exercise 2.14 (*) Let A and B be convex subsetsof $\mathbb{R}^n$, Let us define $D = \mathrm{co}(A \cup B)$ and $E = \{z = \lambda x + (1-\lambda)y \mid x \in A,\ y \in B,\ 0 \leq \lambda \leq 1\}$.

- Show that $E \subset D$.
- Check that E is convex and contains A. Deduce that

$$\mathrm{co}(A \cup B) = \{z = \lambda x + (1-\lambda)y \mid x \in A,\ y \in B,\ 0 \leq \lambda \leq 1\}.$$

- More generally, prove that for a finite collection $(A^i)_{i=1}^p$ of convex subsets of $\mathbb{R}^n$,
$\mathrm{co}(\bigcup_{i=1}^p A^i) = \{z = \sum_{i=1}^p \lambda_i x^i \mid \lambda_i \geq 0,\ x^i \in A^i, i = 1, \ldots p,\ \sum_{i=1}^p \lambda_i = 1\}$.

Definition 2.9 Let E be a vector space which has a finite dimension, and A be a nonempty subset, either finite or infinite. We call the dimension of A the dimension of the affine subspace generated by A.

With this definition, it is easy to show that the unit simplex S^{n-1} is $(n-1)$-dimensional.

2.4.2 Cone and Conic hull

Definition 2.10 A subset K of E is a blunt cone of vertex 0 if for all $x \in K$ and for all $t > 0$, tx belongs to K. If in addition, it contains 0, the set is a pointed cone of vertex 0.

By translation, we may define the notion of cone of vertex z.

Here, a cone will mean ici pointed cone of vertex 0.

Example, any vector subspace is a cone of vertex 0. If A is an affine subspace and if $a \in A$, then A is a cone of vertex a. A cone can be non-convex, $K = \{(x, y) \in \mathbb{R}^2 \mid xy = 0\}$. A bounded nonempty cone is reduced to its vertex. A nonempty cone K (blunt or pointed) of vertex 0 satisfies $0 \in \partial K$.

Proposition 2.9 *Let K be a subset of E.*

(i). K is a pointed cone of vertex 0 if and only if for all $x \in K$ and for all $t \geq 0$, tx belongs to K.

(ii). A one K (pointed or blunt) is convex if and only if it is stable by addition.

Proof of Proposition 2.9. Assertion (i) is trivial.

(ii) Let K be a convex cone. Let x an y be two elements of K. Then $\frac{1}{2}(x + y)$ belongs to K since it is convex and $x + y$ can be written as $2(\frac{1}{2}(x + y))$ which belongs to K since it is a cone. Consequently K is stable by addition.

Let K be a cone stable by addition and let x and y be two elements of K. One has to show that for all $t \in]0, 1[$, $tx + (1-t)y$ belongs to K.

The vectors tx and $(1-t)y$ are elements of K since it is a cone. Since K is stable by addition, $tx + (1-t)y$ belongs to K. $\square$

Examples: All space subspace of E is a convex cone of vertex 0. All set of solutions of a linear system of homogeneous linear equations and homogeneous linear inequations is a convex cone. The image of a convex cone of vertex z by an affine mapping f is a convex cone which vertex is equal to $f(z)$. The inverse image of a convex cone of vertex 0 by a linear mapping f is a convex cone which vertex is equal to 0.

Definition 2.11 Let A be a nonempty subset of E. The convex conic hull of A is the intersection of all the convex pointed cone of vertex 0 containing A. It is denoted by $\mathrm{K}(A)$.

Since E is a convex cone, the set $\mathrm{K}(A)$ is well defined. It is easy to prove that all intersection of convex cones is also a convex cone. Consequently, $\mathrm{K}(A)$ is the smallest (in the sense of inclusion) convex cone containing A. It is also obvious that $\mathrm{K}(A)$ is the set of all linear combinations with non-negative coefficients of elements from A.

Chapter 3

The one dimensional results

3.1 Existence results

The main starting point of existence results is Weierstrass's Theorem and its corollaries.

Theorem 3.1 (Weierstrass) *If the domain is compact, and if f is continuous then f is bounded and the extremum are reached.*

The proof is based on the notion of maximizing sequence (respectively minimizing). When the domain is not bounded, numerous problems can be solved using the coercivity condition.

Definition 3.1 The function $f : \mathbb{R} \to \mathbb{R}$ is said coercive if

$$\lim_{x\to-\infty} f(x) = \lim_{x\to+\infty} f(x) = +\infty.$$

Exercise 3.1 (*) Let us assume that $f : \mathbb{R} \to \mathbb{R}$ is coercive and continuous, show that f has a lower bound and that there exists at least a minimum.

3.2 First order condition

3.2.1 Necessary condition

Proposition 3.1 *Let $-\infty \leq a < b \leq \infty$, we let $I = [a,b] \cap \mathbb{R}$ and we consider $f : I \to \mathbb{R}$ differentiable on I, if we assume that the problem $\max_{x\in I} f(x)$ has a local solution $\overline{x} \in \mathbb{R}$, then*

1. *when $\overline{x} \in]a,b[$, one has $f'(\overline{x}) = 0$,*
2. *when $\overline{x} = a$, one has $f'(\overline{x}) \leq 0$,*
3. *when $\overline{x} = b$, one has $f'(\overline{x}) \geq 0$.*

Exercise 3.2 (*) Let us use the notations of Proposition 3.1.

1. Prove Proposition 3.1.
2. Let us fix $a = \pi$. Prove that π is the solution (even it is a global solution) of $\max_{x\in[\pi,\pi]} e^x$. What is the value of the derivative $f'(a)$?

3.2.2 Alternative formulation in terms of multipliers

Let us focus on the previous problem when the domain is bounded (a and b finite). The problem $\max_{x\in[a,b]} f(x)$ can be formulated as

$$(\mathcal{P}) \begin{cases} \max f(x) \\ x - b \leq 0 \\ a - x \leq 0 \end{cases} \quad \text{i.e } (\mathcal{P}_1) \begin{cases} \max f(x) \\ g_1(x) \leq 0 \\ g_2(x) \leq 0 \end{cases}$$

with the notations $g_1(x) = x - b$ and $g_2(x) = a - x$. We can state Proposition 3.1 under the following version.

Proposition 3.2 *Let $-\infty < a < b < \infty$, we let $I = [a, b]$ and we consider $f : I \to \mathbb{R}$ differentiable on I, we assume that the problem $(\mathcal{P}_1)$ has a local solution $\overline{x}$, then there exist multipliers $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ such that*

$$f'(\overline{x}) = \lambda_1 g_1'(\overline{x}) + \lambda_2 g_2'(\overline{x})$$

and $\lambda_1 g_1(\overline{x}) = \lambda_2 g_2(\overline{x}) = 0$.

Exercise 3.3 (*) Deduce Proposition 3.2 from Proposition 3.1.

Be careful, if we consider the optimization problem,

$$(\mathcal{P}) \begin{cases} \max x \\ g_1(x) = x^2 \leq 0 \end{cases} \tag{3.1}$$

It is easy to prove that $\overline{x} = 0$ is the unique solution of $(\mathcal{P})$ but nevertherless the system

$$\begin{cases} f'(\overline{x}) = \lambda_1 g_1'(\overline{x}) \\ \lambda_1 \geq 0 \\ \lambda_1 g_1(\overline{x}) = 0 \end{cases}$$

has no solution.

Even with a single variable, the optimality does not imply the existence of multipliers, we need an additional condition ("qualification condition"). Solving the previous system dealing with the existence of multipliers at point $\overline{x}$ makes sense only if this point is qualified.

We will now introduce a first version of the notion of qualification. This is not the better one, though it allow to give an intuition on the notion.

Definition 3.2 Let us consider the following set of feasible points, the set O is open and g_i are continuous on O.

$$\begin{cases} x \in O \\ g_1(x) \leq 0 \\ \dots \\ g_p(x) \leq 0 \end{cases}$$

We say that the Slater condition is satisfied for this list of constraints if there exists a point $\hat{x}$ such that $\hat{x} \in O$, and for each $i = 1, \dots, p$,

$$\begin{cases} \text{either } g_i \text{ is an affine function } \textbf{and } g_i(\hat{x}) \leq 0 \\ \textbf{or} \\ g_i \text{ is a convex function } \textbf{and } g_i(\hat{x}) < 0 \end{cases}$$

Note that in the previous definition, the point $\hat{x}$ is <u>feasible</u>.

Exercise 3.4 (*) Let $f : [a, b] \to \mathbb{R}$, let us consider the optimization problems

$$(\mathcal{P}_1) \begin{cases} \max f(x) \in \mathbb{R} \\ x - b \leq 0 \\ a - x \leq 0 \end{cases} \qquad (\mathcal{P}_2) \begin{cases} \max f(x) \\ (x - b)^3 \leq 0 \\ a - x \leq 0 \end{cases}$$

1) Compare the two problems.
2) Show that the Slater condition is satisfied for $(\mathcal{P}_1)$.
3) Show that the Slater condition is NOT satisfied for $(\mathcal{P}_2)$.

Exercise 3.5 (*) Let us consider again the problem

$$(\mathcal{P}) \begin{cases} \max x \\ g_1(x) = x^2 \leq 0 \end{cases}$$

Show that the Slater qualification condition does not hold.

Proposition 3.3 *Let $\overline{x}$ be a local solution of the optimization problem:*

$$(\mathcal{P}) \begin{cases} \max f(x) \\ x \in O \\ g_1(x) \leq 0 \\ \dots \\ g_p(x) \leq 0 \end{cases}$$

Let us assume that the Slater condition is satisfied and that each function is differentiable on a neighborhood of $\overline{x}$, then, there exists $(\lambda_i)_{i=1}^p \in \mathbb{R}_+^p$ such that

$$\begin{cases} f'(\overline{x}) = \lambda_1 g_1'(\overline{x}) + \dots + \lambda_p g_p'(\overline{x}) \\ \lambda_i g_i(\overline{x}) = 0 \text{ for all } i = 1, \dots, p \end{cases}$$

Proof. Let us define the (possibly empty) sets, $K = \{i \in \{1, \dots, p\} \mid g_i(\overline{x}) = 0\}$, $K_+ = \{i \in K \mid g_i'(\overline{x}) > 0\}$, $K_0 = \{i \in K \mid g_i'(\overline{x}) = 0\}$and $K_- = \{i \in K \mid g_i'(\overline{x}) < 0\}$. Note that we only need to prove the existence of $(\lambda_i)_{i \in K}$ since when $x \notin K$, the "complementary condition" $\lambda_i g_i(\overline{x}) = 0$ implies that $\lambda_i = 0$. let us prove the existence of multipliers with respect to the value of $f'(\overline{x})$.

If $f'(\overline{x}) = 0$, it suffices to let in addition $\lambda_i = 0$, for each $i \in K$.

Let us now consider for example the case where $f'(\overline{x}) > 0$, (a symmetric argument allows to treat the case where $f'(\overline{x}) < 0$). We can introduce the alternative problem

$$(\mathcal{P}_1) \begin{cases} \max f(x) \\ g_i(x) \leq 0, \text{ for all } i \in K \end{cases}$$

Using Exercise 1.4, we know that $\overline{x}$ is also a local solution of $(\mathcal{P}_1)$.

Step 1 Let us prove that $K_+ \cup K_0$ is nonempty. First, one can notice that there exists some $\varepsilon > 0$, such that for each $x \in]\overline{x}, \overline{x} + \varepsilon]$, $g_k(x) < 0$, for each $k \in K_-$, and $f(x) > f(\overline{x})$. The value of ε can be chosen such that that $\overline{x}$ is a global solution of

$$(\mathcal{P}_2) \begin{cases} \max f(x) \\ g_i(x) \leq 0, \text{ for all } i \in K \\ x \in [\overline{x} - \varepsilon, \overline{x} + \varepsilon] \end{cases}$$

Since $f(\overline{x}+\varepsilon) > f(\overline{x})$, there exists some $k \in K$ such that $g_k(\overline{x}+\varepsilon) > 0$. In view of the choice of ε, we already know that $k \notin K_-$.

Step 2 Let us know prove that K_+ is nonenmpty. If K_+ is empty, then $k \in K_0$, and the convexity of g_k implies that $\overline{x}$ is a global minimum for g_k. The Slater condition assumes that either g_k is affine or $g_k(\hat{x}) < 0 = g_k(\overline{x})$. The second case is excluded and therefore g_k is affine and more precisely constant. This is not possible since $g_k(\overline{x} + \varepsilon) > 0 = g_k(\overline{x})$.

Step 3 Since K is nonempty, we can choose some $k \in K$, and let

$$\begin{cases} \lambda_i = 0 & \text{if } i \neq k \\ \lambda_k = \dfrac{f'(\overline{x})}{g_k'(\overline{x})} & \end{cases}$$

It is obvious that (λ_i) satisfies all the conditions.

3.3 Convex and concave functions

3.3.1 Definitions

In this part, f is a function defined on U convex subset (interval) of $\mathbb{R}$ to $\mathbb{R}$.

Definition 3.3 The function f is convex (resp. concave) if for all $(x, y) \in U \times U$ and for all $t \in [0, 1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

$$(respectively)\ f(tx + (1-t)y) \geq tf(x) + (1-t)f(y).$$

The function f is strictly convex if for all $(x, y) \in U \times U$ such that $x \neq y$ and for all $t \in]0, 1[$, $f(tx + (1-t)y) < tf(x) + (1-t)f(y)$.

A function f is convex if and only if $-f$ is concave. Consequently, the results obtained for convex functions can be translated in terms of concave functions. These notions can be defined locally.

Definition 3.4 The epigraph and the hypograph (denoted by epi(f) and hypo(f)) of a function f are the sets (see figure 3.1) defined by

$$\text{epi}(f) = \{(x, t) \in U \times \mathbb{R} \mid t \geq f(x)\}.$$

$$\text{hypo}(f) = \{(x, t) \in U \times \mathbb{R} \mid t \leq f(x)\}.$$

Theorem 3.2 *The three following properties are equivalent:*

(i) f is convex (resp. concave);

(ii) for all $k \geq 2$, $(x_i) \in U^k$ and $\lambda \in S^{k-1}$,
$f(\sum_{i=1}^k \lambda_i x_i) \leq$ (resp. $\geq$) $\sum_{i=1}^k \lambda_i f(x_i)$;

(iii) The epigraph (resp. the hypograph) of f is a convex subset of $\mathbb{R}^2$.

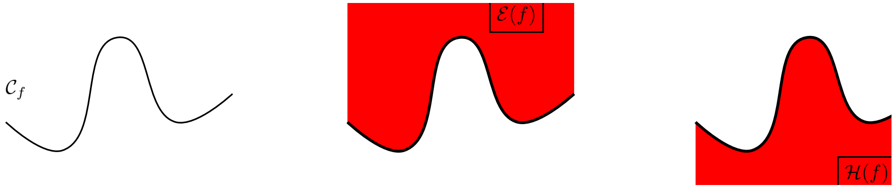

Figure 3.1: graph, epigraph and hypograph

Proof of Theorem 3.2. We will only prove the result in the convex case. It is obvious that (ii) implies (i). Let us now show that (i) implies (iii). Let (x, λ) and (y, μ) two elements of epi(f) and let $t \in [0, 1]$. It means that $f(x) \leq \lambda$ and $f(y) \leq \mu$. Since f is convex, $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$. Then, $f(tx + (1-t)y) \leq t\lambda + (1-t)\mu$. This is equivalent to $t(x, \lambda) + (1-t)(y, \mu) = (tx + (1-t)y, t\lambda + (1-t)\mu)$ belongs to the epigraph of f. Therefore, this set is convex.

We will end the proof by showing that (iii) implies (ii). Let $k \geq 2$, $(x_i) \in (\mathbb{R}^n)^k$ and $\lambda \in S^k$. then, $(x_i, f(x_i))$ is an element of the epigraph of f and since this set is convex,

$$\sum_{i=1}^{k} \lambda_i (x_i, f(x_i)) = (\sum_{i=1}^{k} \lambda_i x_i, \sum_{i=1}^{k} \lambda_i f(x_i))$$

is an element of epi(f). Then, by definition of the epigraph, $f(\sum_{i=1}^{k} \lambda_i x_i) \leq \sum_{i=1}^{k} \lambda_i f(x_i)$. $\square$

Examples:

1. Let us recall a function $\mathbb{R} \to \mathbb{R}$, is an affine function if there exists α and β such that for all x in $\mathbb{R}$,

$$f(x) = \alpha x + \beta.$$

   Any affine function is both convex and concave, but not strictly. It can be easily proved that if f is convex and concave on U, then it is the restriction on U of an affine function.

2. $|.|$ is a convex function.

3. If C is a nonempty convex convex subset of $\mathbb{R}$, the distance to C defined by $d_C(x) = \inf\{|x - c| \text{ such that } c \in C\}$ is convex.

Proposition 3.4 *(i) A finite sum of convex functions (resp. concave) defined on U is convex (resp. concave);*

(ii) if f is convex (resp. concave) and $\lambda > 0$, λf is convex (resp. concave);

(iii) The supremum (resp. infimum) of a family of convex functions (resp. concave) defined on U is convex (resp. concave) on its domain (when the supremum is finite);

(iv) If f is a convex function (resp. concave) from I to J, intervals of $\mathbb{R}$, and if φ is a convex function (resp. concave) non-decreasing from I to $\mathbb{R}$ then $\varphi \circ f$ is convex (resp. concave).

(v) if g is an affine function from $\mathbb{R}$ to $\mathbb{R}$ and f a convex function on $U \subset \mathbb{R}$, then $f \circ g$ is a convex function on $g^{-1}(U)$.

The proof of this proposition is left to the reader.

Exercise 3.6 (*) Let f be the function defined by $f(x) = -|x|$ on $\mathbb{R}$,

Show that f is convex on $[0, +\infty[$ and on $]-\infty, 0]$ but not on $\mathbb{R}$.

3.3.2 quasi-convex functions

Definition 3.5 Let f be a real-valued function defined on an interval I, we say that f is quasi-concave if for all $\alpha \in \mathbb{R}$, the set $\{x \in I \mid f(x) \geq \alpha\}$ is convex. We say that f is quasi-convex if the sets $\{x \in I \mid f(x) \leq \alpha\}$ are convex.

Exercise 3.7 (*) Let f be a real-valued function defined on an interval U,

Show that f is quasi-convex if and only if for all x, y of U and all $\lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \leq \max(f(x), f(y))$$

Proposition 3.5 *Let f be a real-valued function defined on an interval,*

1) if f is convex, then f is quasi-convex.

2) The function f is quasi-convex if and only if $(-f)$ is quasi-concave.

3) if f is weakly monotone, then f is both quasi-concave and quasi-convex.

For example, the exponential function is convex and quasi-concave.

Exercise 3.8 (*) Show Proposition 3.5.

Exercise 3.9 (*) Let U be an open interval of $\mathbb{R}$ and f be a function $\mathcal{C}^1$ from U to $\mathbb{R}$. Show that f is quasi-convex if and only if

$$\forall x, \in U,\ \forall y \in U, f(y) \leq f(x) \Rightarrow f'(x)(x-y) \leq 0.$$

Exercise 3.10 (*) Let f be the function defined by $f(x) = -|x|$ on $\mathbb{R}$,
Show that f is quasi-convex on $[0, +\infty[$ and on $]-\infty, 0]$ but not on $\mathbb{R}$.

There exists several notions of strict quasi-convexity, and one should be cautious and check the definition used. Here we will use the following definition.

Definition 3.6 Let f be a real-valued function defined on an interval U, We say that f is strictly quasi-convex if f is quasi-convex and if for all x, y from U satisfying $x \neq y$ and for all $\lambda \in]0, 1[$,

$$f(\lambda x + (1-\lambda)y) < \max(f(x), f(y))$$

When the function is continuous, we can propose another characterization of strict convexity (see exercise 3.11).

Exercise 3.11 (**) Let f be a continuous real-valued function defined on an interval U, we assume that for all x, y from U satisfying $x \neq y$, $f(x) = f(y)$ and for all $\lambda \in]0, 1[$,

$$f(\lambda x + (1-\lambda)y) < f(x)$$

Show that f is strictly quasi-convex.

Exercise 3.12 (**) Let U be an open interval of $\mathbb{R}$ and f be a strictly quasi-convex function:

1. Show that if $\overline{x}$ is a local solution of $\min_{x \in U} f(x)$, then it is also a global solution.

2. If there exists a solution to the minimization problem, then it is unique.

Exercise 3.13 (**) Let U be an open interval of $\mathbb{R}$ and f a function $\mathcal{C}^2$ from $U \to \mathbb{R}$. We assume that for all $x \in U$, $f'(x) = 0 \Rightarrow f''(x) > 0$.

1. The goal of this question is to show by contradiction that f is strictly quasi-convex :

   a) Let $x_0 \neq x_1$ be two elements of U, such that there exists $x_\lambda \in]x_0, x_1[$ satisfying $f(x_\lambda) \geq \max(f(x_0), f(x_1))$. Show that the problem $\max_{u \in [x_0, x_1]} f(u)$ has at least one solution z satisfying $z \in]x_0, x_1[$.

   b) Show that $f'(z) = 0$, and that $f''(z) > 0$. Deduce the contradiction.

2. Let $\overline{x} \in U$, such that $f'(\overline{x}) = 0$, show that $\overline{x}$ is the unique global minimum (necessary and sufficient condition). (Hint, one may use exercise 3.12).

3. Let g be a function $U \to \mathbb{R}$ C^1, such that $g'(x) \neq 0$ for all $x \in U$, show that g is both strictly quasi-convex and strictly quasi-concave.

Exercise 3.14 1) Show with a counter-example that the sum of two quasi-convex functions is not necessarily quasi-convex.
2) Show with a counter-example that the sum of a quasi-convex function and a convex function is not necessarily quasi-convex.
3)Show with a counter-example that the sum of a strictly quasi-convex function and a strictly convex function is not necessarily quasi-convex.

3.3.3 Regularity of convex functions

We will give in this part important results on the continuity of convex functions.

Proposition 3.6 *if f is convex (resp. concave) and if $x < x' < x''$ are elements of I, then (figure 3.2),*

$$\frac{f(x')-f(x)}{x'-x} \le \frac{f(x'')-f(x)}{x''-x} \le \frac{f(x'')-f(x')}{x''-x'}$$

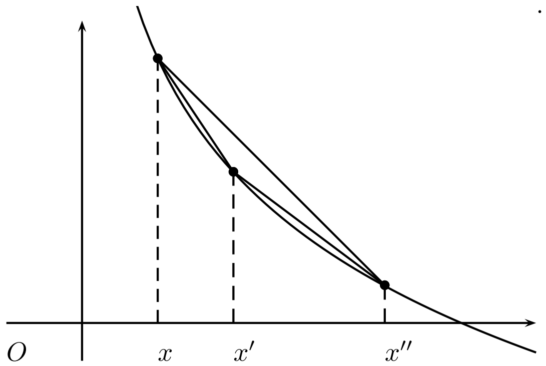

Figure 3.2: Monotonicity of the slope of a convex function

We can deduce from this that f is Lipschitz and even give an explicit bound to the Lipschitz constant.

Corollary 3.1 *1) Let $\alpha \le \beta$, f a convex function on $I = [\alpha, \beta]$, then f is bounded on I.*

2) Let $a \le b$, $\varepsilon > 0$ f be a convex function on $I = [a - \varepsilon, b + \varepsilon]$, we assume that on I, f is bounded from below by m and from above by M. then f is k-Lipschitz on $[a, b]$, for $k = (M - m)/\varepsilon$ and therefore continuous on $[a, b]$.

3) Let f be a convex function on U, then f is locally lipschitz and continuous on the interior of U.

Proof : 1) Since f is convex, it is also quasi-convex. For all $x \in [\alpha, \beta]$, $f(x) \le M = \max(f(\alpha), f(\beta))$.

If we denote $\gamma = (\alpha+\beta)/2$ and $m = 2f(\gamma) - M$, we want to prove that for all $x \in [\alpha, \beta]$, $f(x) \ge m$. Let us denote by $y = 2\gamma - x$, by definition γ is the middle of $\{x, y\}$. The convexity of f allows us to deduce that $f(\gamma) \le (f(x) + f(y))/2$. Since $f(x) \le 2f(\gamma) - f(y)$, then m is a lower bound.

2) Let us recall that f is Lipschitz on X if for all $(x, x') \in X \times X$, $|f(x) - f(x')| \le k|x - x'|$. It is easy to prove that a Lipschitz function is continuous.

If f is bounded from below by m and from above by M, we can write for all $x < y$ in $[a, b]$,

$$\frac{f(x-\varepsilon)-f(x)}{(x-\varepsilon)-x} \le \frac{f(x)-f(y)}{x-y} \le \frac{f(y)-f(y+\varepsilon)}{y-(y+\varepsilon)}$$

Then

$$\frac{M-m}{-\varepsilon} \le \frac{f(x)-f(y)}{x-y} \le \frac{M-m}{\varepsilon}$$

This allows us to deduce that $|f(x) - f(y)| \le \left(\frac{M-m}{\varepsilon}\right)|x - y|$.

3) Easy deduction of 1) and of 2). □

Corollary 3.2 *Let f be a convex function on some interval U,*

1) if $x_1 < x_2 < x_3 < x_4 < x_5$ are elements of U, then

$$\frac{f(x_1)-f(x_2)}{x_1-x_2} \le \frac{f(x_2)-f(x_3)}{x_2-x_3} \le \frac{f(x_3)-f(x_4)}{x_3-x_4} \le \frac{f(x_4)-f(x_5)}{x_4-x_5}$$

2) The right and left derivatives exist on the interior of U and moreover if $x_1 < x_2$ are interior points, then

$$f'_r(x_1) \le \frac{f(x_2)-f(x_1)}{x_2-x_1} \le f'_\ell(x_2).$$

3) The right and left derivatives are non decreasing on the interior of U, and more over if x is an interior point, $f'_\ell(x_1) \le f'_r(x_1)$.

Remark 3.1 Be careful, f may be discontinuous on the boundary of the domain. For example,

$$f(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x > 0 \end{cases}$$

is convex on $\mathbb{R}_+$.

Exercise 3.15 (***) Let f be a function defined on some interval I, and let $\overline{x}$ be a point of I. Let us define the set

$$\partial f(\overline{x}) = \{\alpha \in \mathbb{R} \mid f(x) \geq f(\overline{x}) + \alpha(x - \overline{x}), \text{ for all } x \in I\}.$$

1. Show that if $\overline{x}$ is an interior point of I, and if the right and the lef derivatives exists at point $\overline{x}$, then for all $\alpha \in \partial f(\overline{x})$, we have $f'_r(\overline{x}) \geq \alpha$ and $\alpha \leq f'_l(\overline{x})$.

2. We assume that f is convex. Deduce from Proposition 3.2 that if $\overline{x}$ is an interior point of I, $\partial f(\overline{x}) = [f'_l(\overline{x}), f'_r(\overline{x})]$, in particular, that the set is nonempty.

3. Let $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = min(0, -x)$, determine for all $\overline{x}$ in $\mathbb{R}$, the set $\partial f(\overline{x})$

4. We assume that f is convex and that $\partial f(\overline{x}) = \{a\}$, show that f is differentiable at point $\overline{x}$ and that $f'(\overline{x}) = a$.

3.3.4 Characterization with derivatives

Let us end this part by studying the possible characterizations when the first (respectively the second) derivative exists.

Proposition 3.7 *Let f be a differentiable function defined on some convex set $U \subset \mathbb{R}$. The function f is convex if and only if for all $(x, y) \in U \times U$, $f(y) - f(x) \geq f'(x)(y - x)$.*

Proof of Proposition 3.7. Let us first consider the implication $\Rightarrow$. Let $(x, y) \in U \times U$. If $x = y$, there is nothing to prove. Let us assume now that $x \neq y$, using mean value Theorem, there exists some c in $]x, y[$ if $x < y$ and in $]y, x[$ if $y < x$ such that

$$\frac{f(y) - f(x)}{y - x} = f'(c)$$

The weak monotony of f' (See Corollary 3.2) allows us to conclude.

Conversely, we assume that the property on the derivative of f is satisfied. Let $(x, y) \in U \times U$ and let $t \in]0, 1[$. We have

$$f(x) - f(x + t(y - x)) \geq f'(x + t(y - x))(-t(y - x))$$

and

$$f(y) - f(x + t(y - x)) \geq f'(x + t(y - x))((1 - t)(y - x)).$$

If we multiply the first inequality by $(1 - t)$, the second inequality by t, and if we sum, then we get that $(1-t)(f(x) - f(x+t(y-x))) + t(f(y) - f(x+t(y-x))) \geq (-t(1-t) + t(1-t))f'(x+t(y-x))(y-x) = 0$. Then $f(x + t(y - x)) = f((1 - t)x + ty) \leq (1 - t)f(x) + tf(y)$, i.e. f is convex. $\square$

The next exercise propose an alternative proof of the converse implication.

Exercise 3.16 Let f be a differentiable mapping defined on the convex set U, we assume that for all $(x, x') \in U \times U$, $f(x) - f(x') \geq f'(x')(x - x')$. 1) Show that

$$\text{epi}(f) = \bigcap_{x' \in U} \{(x, y) \in U \times \mathbb{R} \mid y \geq f(x') + f'(x')(x - x')\}.$$

2) Using the characterization given by Theorem 3.2, deduce that f is convex.

The monotony of f' characterize entirely the convexity, this can be formalized in the next proposition.

Proposition 3.8 *Let f be a differentiable function on the interval U. The function f is convex if and only if for all $(x, y) \in U \times U$, $(f'(y) - f'(x))(y - x) \geq 0$.*

Proof of Proposition 3.8. The implication $\Rightarrow$ is a consequence of the weak monotony of f' (see Corollary 3.2).

For the converse implication, let us consider x and y two elements of U. Using the proposition 3.7, it suffices to show that $f(y) - f(x) - (x -$

$y)f'(x) \geq 0$ in order to prove that f is convex. If $x = y$, there is nothing to prove.

We will first consider the case $x > y$. Since $y - x < 0$, we have by assumption that for all t of $]y, x[$, $(f'(t) - f'(x))(t - x) \geq 0$, and consequently $f'(t) - f'(x) \leq 0$. If we integrate on $[y, x]$ this inequality, we get $\int_y^x (f'(t) - f'(x))dt \leq 0$, this means $f(y) - f(x) - (x - y)f'(x) \geq 0$,

The proof of the case $x < y$ is similar. □

Let us now consider the case where f is twice differentiable on U.

Corollary 3.3 *Let f be a twice differentiable function on the interval U, f is convex if and only if for all $x \in U$, $f''(x) \geq 0$.*

Proof of Proposition 3.3. Indeed, f is convex, if and only if f' is weakly increasing, therefore if and only if $f'' \geq 0$ on U. □

In order to get a result involving strict convexity, we may use the results of the next exercise.

Exercise 3.17 1) Show that the function f defined by $f(x) = x^4$ is strictly convex on $\mathbb{R}$ but nevertheless the second derivative is not always positive on $\mathbb{R}$ since $f''(0) = 0$.

2) Let f be a twice differentiable function on the interval U. Show that if for all $x \in U$, $f''(x) > 0$, then f is strictly convex on U.

3) Let f be a $\mathcal{C}^2$ function on U a neighborhood of $\overline{x}$, and assume that $f''(\overline{x}) > 0$. Show that there exists V a (possibly smaller) neighborhood of $\overline{x}$ such that f is strictly convex on V.

4) Let f be a $\mathcal{C}^1$ function on some interval U such that f' is increasing (strictly). Prove that f is strictly convex on V.

3.4 Sufficient conditions for optimality

When the problem is not convex, the only case where we can conclude is the following:

Proposition 3.9 *Let $-\infty \leq a < b \leq \infty$, we let $I = [a, b] \cap \mathbb{R}$ and we consider $f : I \to \mathbb{R}$ differentiable at point a. If $f'(a) > 0$, a is a local solution of minimization problem.*

When the problem is "convex", we can state:

Proposition 3.10 *Let us consider the optimization problem :*

$$\begin{cases} \max f(x) \\ x \in O \\ g_1(x) \leq 0 \\ \dots \\ g_p(x) \leq 0 \end{cases}$$

We assume that $\overline{x}$ is a feasible point for this system. If the function f is globally concave, (respectively locally concave at point $\overline{x}$), and if the functions g_i are globally convex, (respectively locally convex at point $\overline{x}$) and if there exists $(\lambda_i)_{i=1}^p \in \mathbb{R}_+^p$ such that

$$\begin{cases} f'(\overline{x}) = \lambda_1 g_1'(\overline{x}) + \ldots + \lambda_p g_p'(\overline{x}) \\ \lambda_i g_i(\overline{x}) = 0 \text{ for all } i = 1, \ldots, p. \end{cases}$$

then, the condition is sufficient, $\overline{x}$ is a global solution (respectively local) of the maximization problem.

Proof of the proposition 3.10. Let us call the "Lagrangian" of the problem $\mathcal{L}(x) = f(x) - \sum_{i=1}^p \lambda_i g_i(x)$. Since this function is globally concave, (respectively locally concave at point $\overline{x}$) and $\mathcal{L}'(\overline{x}) = 0$, we can write that for all point x feasible (respectively feasible and in a neighborhood of $\overline{x}$), $\mathcal{L}(\overline{x}) \leq \mathcal{L}(x)$. we can conclude if we note that for each feasible point x, $\mathcal{L}(x) \geq f(x)$ and that $\mathcal{L}(\overline{x}) = f(\overline{x})$ (in view of the conditions on the multipliers). □.

Exercise 3.18 Let us consider the problem:

$$(\mathcal{P}) \begin{cases} \min x^3 \\ -1 - x \leq 0 \end{cases}$$

1) Check that the Slater condition is satisfied, that the objective function is strictly quasi-concave, that the constraint function is convexe.

2) Write the conditions for the existence of a multiplier at some feasible point x and show that there exists two solutions: "$x = 0$ associated to the multiplier $\lambda = 0$" and "$x = -1$ associated to the multiplier $\lambda = 3$".

3) Show that the point $x = 0$ is not a solution (even a local solution) of our optimization problem.

3.5 Introduction to sensibility analysis

Let us consider in this part an optimization problem with a single inequality constraint involving one real parameter β

$$(\mathcal{P}_\beta)\begin{cases}\max f(x)\\ g(x)\le\beta\end{cases}$$

We assume that the functions f and g are $\mathcal{C}^1$ on $\mathbb{R}$ and that for all β, there exists a unique solution x_β. We denote by $\varphi(\beta)$ the value of $(\mathcal{P}_\beta)$, we want to study the behavior of the function φ. Note that it is obvious that when β increases, the set of feasible points is bigger or equal, which implies that the function φ will be non-decreasing.

Theorem 3.3 *Let us assume that the function f is strictly concave, g convex, that for all x of $\mathbb{R}$, $g'(x)>0$, and that the limit of f in $-\infty$ is $-\infty$, then for all β,*

1. *there exists a unique solution denoted by $h(\beta)$.*
2. *the function h is continuous at point β and has a right and left derivative at each point.*
3. *the function φ is $\mathcal{C}^1$ at point β and $\varphi'(\beta)=\lambda_\beta$ where λ_β is the unique multiplier associated to the optimal solution $h(\beta)$.*

Proof of Theorem 3.3. According to the question of existence, since g is increasing, there exists an inverse and the problem can be equivalently stated as $\max_{x\in]-\infty,g^{-1}(\beta)]} f(x)$. Since $\lim_{x\to-\infty} f(x)=-\infty$, and since f is continuous (in view of its concavity on $\mathbb{R}$), this problems has a solution which is unique since f is strictly concave.

Let us first remark that the assumption on g' implies that each point of the domain will be qualified. Therefore x is solution of $(\mathcal{P}_\beta)$ if and only if there exists a multiplier λ. Then, (x,λ) is solution of $(\mathcal{S}_\beta)$

$$(\mathcal{S}_\beta)\begin{cases}\lambda\ge 0\\ f'(x)=\lambda g'(x)\\ \lambda(g(x)-\beta)=0\\ g(x)-\beta\le 0\end{cases}\quad\Leftrightarrow\quad\begin{cases}\lambda=f'(x)/g'(x)\ge 0\\ (f'(x)/g'(x))(g(x)-\beta)=0\\ g(x)-\beta\le 0\end{cases}$$

$$\Leftrightarrow\quad(\mathcal{S}'_\beta)\begin{cases}f'(x)\ge 0\\ f'(x)(g(x)-\beta)=0\\ g(x)-\beta\le 0\end{cases}$$

One should note that the associated multiplier is unique. In order to study the regularity of the value function, we will distinguish several case depending on the sign of the multiplier.

First case: $g(h(\beta))<\beta$.
In this case, the associated multiplier denoted by λ_β is equal to 0. We have,

$$\begin{cases}f'(x_\beta)=0\\ 0.(g(x_\beta)-\beta)=0\end{cases}$$

For all β' close enough to β, $g(h(\beta))<\beta'$ and the couple $(h(\beta),0)$ is solution of $(\mathcal{P}_{\beta'})$, i.e. that $h(\beta')=h(\beta)$. The solution is here constant. Since $\varphi(\beta')=f(x_{\beta'})$, we can deduce that $\varphi'(\beta)=0=\lambda_\beta$.

Second case : $g(x_\beta)=\beta$ **and** $\lambda_\beta>0$.
We have, $(\mathcal{S}'_\beta)\begin{cases}f'(h(\beta))>0\\ g(h(\beta))-\beta=0\end{cases}$

In view of the implicit function theorem, there exists $\varepsilon>0$, and a function $\gamma:[\beta-\varepsilon,\beta+\varepsilon]\to[h(\beta)-\varepsilon,h(\beta)+\varepsilon]$, such that for all $\beta'\in[\beta-\varepsilon,\beta+\varepsilon]$, $g(\gamma(\beta'))=\beta'$.

It is obvious that locally $\gamma(\beta')$ is solution of $(\mathcal{S}'_{\beta'})$, this means that locally $h=\gamma$, in particular h is $\mathcal{C}^1$ in a neighborhood of β. Since $\varphi(\beta')=f(h(\beta'))$, we can deduce that

$$\varphi'(\beta)=\gamma'(\beta)f'(\gamma(\beta))=\frac{1}{g'(h(\beta))}f'(h(\beta))=\lambda_\beta$$

Third case : $g(x_\beta)=\beta$ **and** $\lambda_\beta=0$.
We have,

$$\begin{cases}f'(x_\beta)=0\\ g(x_\beta)-\beta=0\end{cases}$$

- As in the first case, for all $\beta'>\beta$ and close enough to β, the couple $(h(\beta),0)$ is solution of $(\mathcal{P}_{\beta'})$, i.e. that $h(\beta')=h(\beta)$. We can deduce that h is continuous on the right, has a right derivative, at point β, moreover the right derivative $\varphi'_r(\beta) = 0 = \lambda_\beta$.

- As in the second case, for all $\beta' < \beta$ close enough to β, in view of the implicit function theorem, there exists $\varepsilon > 0$, and a function $\gamma : [\beta-\varepsilon, \beta+\varepsilon] \to [x_\beta - \varepsilon, x_\beta + \varepsilon]$, such that for all $\beta' \in [\beta-\varepsilon, \beta+\varepsilon]$, $g(\gamma(\beta')) = \beta'$.

  It remains to study the sign of the multiplier. Since $f'(\gamma(\beta)) = 0$, γ is increasing, we can deduce that the strict concavity of f implies that $f'(\gamma(\beta')) \geq 0$ for all $\beta' < \beta$ close enough to β. Then $\gamma(\beta')$ is solution of $(\mathcal{S}'_{\beta'})$, i.e. $h = \gamma$. We can deduce that h is continuous on the left, has a left derivative at point β, moreover the left derivative

$$\varphi'_\ell(\beta) = \gamma'(\beta) f'(\gamma(\beta)) = \frac{1}{g'(x_\beta)} f'(x_\beta) = \lambda_\beta$$

In this theorem, h is not necessarily $\mathcal{C}^1$, while $\varphi = f \circ h$ is $\mathcal{C}^1$. $\square$.

Exercise 3.19 Let β be a real parameter, we consider the problem

$$(\mathcal{P}_\beta) \left\{ \begin{array}{l} \max -x^2 \\ x \leq \beta \end{array} \right.$$

Show that

$$\mathcal{S}ol(\mathcal{P}_\beta) = \{0\} \text{ if } \beta \geq 0 \text{ and } \{\beta\} \text{ if } \beta \leq 0$$

Deduce that in particular the solution is not differentiable at point 0 with respect to β.

Exercise 3.20 Let h defined by $h(x) = g(-x)$.

1) Compare the sets of solutions and the values of the problems

$$(\mathcal{P}_\beta) \left\{ \begin{array}{l} \max f(x) \\ g(x) \leq \beta \end{array} \right. \qquad (\mathcal{Q}_\beta) \left\{ \begin{array}{l} \max f(-x) \\ h(x) \leq \beta \end{array} \right.$$

2) Deduce, that in Theorem 3.3, the conclusion remains true if we replace the condition “for all x, $g'(x) > 0$” by “for all x, $g'(x) < 0$”.

Chapter 4

Finite dimensional optimization

4.1 Existence results

As in dimension 1, the main starting point of existence results is Weierstrass's Theorem and its corollaries.

Theorem 4.1 (Weierstrass) *Let f be a function whose domain is compact, and if f is continuous then f is bounded and the extrema are reached.*

The proof uses the notion of maximizing sequence (cf. Definition 1.5) and is based on the following lemma.

Lemma 4.1 *Let us consider a maximizing sequence $(x_k)_k$. If the domain is closed, the objective function f is continuous, then any cluster point of $(x_k)_k$ is a solution to the maximization problem.*

When the domain is not bounded, numerous problems can be solved using the coercivity condition.

Definition 4.1 The function $f : A \to \mathbb{R}$ is said coercive if

$$\lim_{\|x\| \to +\infty} f(x) = +\infty.$$

Proposition 4.1 *Let us assume that $f : A \to \mathbb{R}$ is coercive and continuous, and that A is closed, then f has a lower bound and there exists at least a minimum.*

Exercise 4.1 (*) Let us consider the following optimization problem.

$$(\mathcal{P}) \quad \begin{cases} \min & x + y \\ s.t. & x \geq 0 \\ & y \geq 0 \\ & (1+x)(1+y) \leq 2 \end{cases}$$

1. Draw the set of feasible points compact.
2. Prove that we can apply 4.1 to this problem.

Exercise 4.2 (*) Let us consider the following optimization problem.

$$(\mathcal{P}) \begin{cases} \max & 3x_1x_2 - x_2^3 \\ s.t. & x_1 \geq 0 \\ & x_2 \geq 0 \\ & x_1 - 2x_2 = 5 \\ & 2x_1 + 5x_2 \geq 20 \end{cases}$$

1. Is the set of feasible points compact ?
2. Prove that we can apply Proposition 4.1 to the "modified" problem

$$(\mathcal{Q}) \begin{cases} \min & -3x_1x_2 + x_2^3 \\ s.t. & x_1 \geq 0 \\ & x_2 \geq 0 \\ & x_1 - 2x_2 = 5 \\ & 2x_1 + 5x_2 \geq 20 \end{cases}$$

3. conclude

4.2 Convex and quasi-convex functions

4.2.1 Definitions

In this part, f is a function defined on U convex subset of $\mathbb{R}^n$ with values in $\mathbb{R}$.

Definition 4.2 The function f is convex (resp. concave) if for all $(x, y) \in U \times U$ and for all $t \in [0, 1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

$$(respectively)\ f(tx + (1-t)y) \geq tf(x) + (1-t)f(y).$$

The function f is strictly convex if for all $(x, y) \in U \times U$ such that $x \neq y$ and for all $t \in]0, 1[$, $f(tx + (1-t)y) < tf(x) + (1-t)f(y)$.

Exercise 4.3 (**)

1. Let U be an open convex set on which the function f is assumed to be convex, We assume moreover that f is continuous on the closure of U. Show that f is convex on $\overline{U}$.

2. Let $U = \mathbb{R}^2_{++} = \{x \in \mathbb{R}^2 \mid x_1 > 0 \text{ and } x_2 > 0\}$, let us consider f defined on $\overline{U}$ by $f(x_1, x_2) = \sqrt[4]{x_1 x_2}$. Is f strictly concave on U? on $\overline{U}$?

A function f is convex if and only if $-f$ is concave. Consequently, the results obtained for convex functions can be translated in terms of concave functions. These notions can be defined locally.

Definition 4.3 Let f be a real-valued function defined on a convex set U, we say that f is quasi-concave if for all $\alpha \in \mathbb{R}$, the set $\{x \in U \mid f(x) \geq \alpha\}$ is convex. We say that f is quasi-convex if for all $\alpha \in \mathbb{R}$, the set $\{x \in U \mid f(x) \leq \alpha\}$ is convex.

There exist several notions of strict quasi-convexity, and one should be cautious and check the definition used. Here we will use the following definition.

Definition 4.4 Let f be a real-valued function defined on a convex set U, We say that f is strictly quasi-convex if f is quasi-convex and if for all x, y from U satisfying $x \neq y$ and for all $\lambda \in]0, 1[$,

$$f(\lambda x + (1-\lambda)y) < \max(f(x), f(y))$$

Exercise 4.4 (**)

1. Let U be an open convex set on which the function f is assumed to be quasi-convex, We assume moreover that f is continuous on the closure of U. Show that f is quasi-convex on $\overline{U}$.

2. Let $U = \mathbb{R}^2_{++} = \{x \in \mathbb{R}^2 \mid x_1 > 0 \text{ and } x_2 > 0\}$, let us consider f defined on $\overline{U}$ by $f(x_1, x_2) = \sqrt[4]{x_1 x_2}$. Is f strictly quasi-concave on U? on $\overline{U}$?

Remark 4.1 Let U be a convex subset of $\mathbb{R}^n$. One has the equivalence

- The function f is convex (respectively strictly convex, quasi-convex, strictly quasi convex) on U,

- for all $x \neq y$ in U, the function $\varphi_{x,y}$ defined from $[0, 1]$ to $\mathbb{R}$ by,

  $$\varphi_{x,y}(t) = f(ty + (1-t)x) = f(x + t(y-x))$$

  is convex (respectively strictly convex, quasi-convex, strictly quasi convex) on $[0, 1]$.

Note that $\varphi_{x,y}$ is differentiable (resp. twice differentiable) when f is differentiable (resp. twice differentiable) then, $\varphi'_{x,y}(t) = \langle \nabla f(x+t(y-x)), y-x\rangle$ and $\varphi''_{x,y}(t) = \langle H_f(x + t(y-x)))(y-x), y-x\rangle$.

Definition 4.5 The epigraph and the hypograph (denoted by epi(f) and hypo(f)) of a real-valued function f are the sets (see figure 4.1) defined by

$$\text{epi}(f) = \{(x, t) \in U \times \mathbb{R} \mid t \geq f(x)\}.$$

$$\text{hypo}(f) = \{(x, t) \in U \times \mathbb{R} \mid t \leq f(x)\}.$$

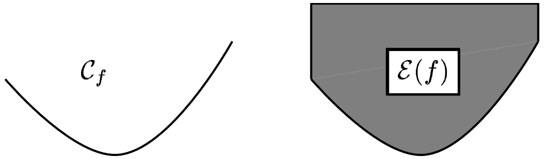

Figure 4.1: graph and epigraph

Theorem 4.2 *The three following properties are equivalent:*

1. *f is convex (resp. concave);*
2. *for all $k \geq 2$, $(x_i) \in U^k$ and $\lambda \in S^{k-1}$, (unit simplex of $\mathbb{R}^k$)* $f(\sum_{i=1}^k \lambda_i x_i) \leq$ (resp. $\geq$) $\sum_{i=1}^k \lambda_i f(x_i)$;
3. *The epigraph (resp. the hypograph) of f is a convex subset of $\mathbb{R}^{n+1}$.*

Proof of Theorem 4.2. Simple adaptation of the proof of Theorem 3.2

Examples:

1. Any affine real-valued function is both convex and concave, but not strictly. It can be easily proved that if f is convex and concave on U, then it is the restriction to U of an affine function, (still called affine function by an abuse of language).
2. Each norm is a convex function.
3. If C is a nonempty convex subset of $\mathbb{R}^n$, the distance to C defined by $d_C(x) = \inf\{\|x - c\| \text{ such that } c \in C\}$ is convex.

Proposition 4.2 *1. A finite sum of convex functions (resp. concave) defined on U is convex (resp. concave);*

2. if f is convex (resp. concave) and $\lambda > 0$, λf is convex (resp. concave);

3. The supremum (resp. infimum) of a family of convex functions (resp. concave) defined on U is convex (resp. concave) on its domain (when the supremum is finite);

4. If f is a convex function (resp. concave) from U (convex set of $\mathbb{R}^n$) to I (interval of $\mathbb{R}$), and if φ is a convex function (resp. concave) non-decreasing from I to $\mathbb{R}$ then $\varphi \circ f$ is convex (resp. concave) on U.

5. if g is an affine function from $\mathbb{R}^n$ to $\mathbb{R}^p$ and f a convex function on $U \subset \mathbb{R}^p$, then $f \circ g$ is a convex function on $g^{-1}(U)$.

The proof of this proposition is left to the reader.

Exercise 4.5 Let U be a convex set of $\mathbb{R}^n$ and f be a convex function: Show that if $\overline{x}$ is a local solution of $\min_{x \in U} f(x)$, then it is also a global solution.

Exercise 4.6 (*) Let f be a real-valued function defined on a convex set U, show that f is quasi-convex if and only if for all x, y of U and all $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \max(f(x), f(y))$$

Exercise 4.7 (*) Let U be an open convex set of $\mathbb{R}^n$ and f be a strictly quasi-convex function:

1. Show that if $\overline{x}$ is a local solution of $\min_{x \in U} f(x)$, then it is also a global solution.
2. If there exists a solution to the minimization problem, then it is unique.

When the function is regular, we can propose characterizations of (strict) quasi-convexity (see exercises 4.8 and 4.9).

Exercise 4.8 (*) Let U be an open set of $\mathbb{R}^n$ and f be a function $\mathcal{C}^1$ on U. Show that f is quasi-convex if and only if

$$\forall x, \in U,\ \forall y \in U, f(y) \leq f(x) \Rightarrow \langle \nabla f(x), x - y \rangle \leq 0.$$

Exercise 4.9 (*) Let f be a continuous real-valued function defined on a convex set U, we assume that for all x, y of U satisfying $x \neq y$, $f(x) = f(y)$ and for all $\lambda \in]0, 1[$,

$$f(\lambda x + (1 - \lambda)y) < f(x)$$

Show that f is strictly quasi-convex.

Exercise 4.10 (*) Let U be a convex subset of $\mathbb{R}^n$ and f be a quasi-convex function. We want to study $\min_{x\in U} f(x)$.

1. Show that the set of solution to the minimization problem is convex.
2. Prove with a counter-example that a local solution is not necessarily a global solution. **Hint:** consider $f : \mathbb{R} \to \mathbb{R}$, defined by

$$f(x) = \begin{cases} -x-1 & \text{if } x \leq -1, \\ 0 & \text{if } x \in [-1,1], \\ x-1 & \text{if } x \geq 1. \end{cases}$$

Exercise 4.11 (**) Let U be an open convex set of $\mathbb{R}^n$ and f a function $\mathcal{C}^2$ from $U \to \mathbb{R}$. We assume that for all $x \in U$, and all $v \in \mathbb{R}^n \setminus \{0\}$, $\langle \nabla f(x), v\rangle = 0 \Rightarrow v^t H_f(x) v > 0$. Show that f is strictly quasi-convex.

Hint: use Exercise 3.13 and Remark 4.1.

4.2.2 Regularity of convex functions

We will give in this part important results on the continuity of convex functions.

Proposition 4.3 *Let U be a convex set, $\varepsilon > 0$ and f be a convex function on $V = U + B(0,\varepsilon)$,*

1. *We assume that on V, f is bounded from below by m and from above by M. then f is k-Lipschitz on U, for $k = (M-m)/\varepsilon$ and therefore continuous on U.*
2. *Let f be a convex function on U, then f is locally lipschitz and continuous on the interior of U.*

Note that this property is false in an infinite dimension setting. For example, if we consider $E = \mathbb{R}[X]$ (polynomial functions), embedded with the norm

$$\|P\| = \sum_{k=0}^{deg(P)} |a_k|, \text{ when } P = \sum_{k=0}^{deg(P)} a_k X^k.$$

It is easy to see that the linear φ(and consequently convex) is NOT continuous, where $\varphi(P) = P'(1)$.

4.2.3 Characterization of convexity with derivatives

Let us end this part by studying the possible characterizations when the first (respectively the second) derivative exists.

Proposition 4.4 *Let f be a differentiable function defined on some convex set $U \subset \mathbb{R}^n$. The function f is convex if and only if for all $(x,y) \in U \times U$,*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle.$$

Proof of Proposition 4.4.

Let us first consider the implication $\Rightarrow$. Let $(x,y) \in U \times U$ and $x \neq y$. In view of Remark 4.1, we know that the function $\varphi_{x,y}$ is convex. Since this function has a derivative, we know that $\varphi_{x,y}(1) - \varphi_{x,y}(0) \geq \varphi'_{x,y}(0)(1-0)$ which leads to the conclusion.

Conversely, we assume that the property on the gradient of f is satisfied. Let $x_0 \neq x_1$ in U, let us consider the function φ_{x_0,x_1}. In view of Proposition 3.7, it suffices to prove that for all $t \neq t'$ (in $[0,1]$),

$$\varphi_{x_0,x_1}(t') - \varphi_{x_0,x_1}(t) \geq \varphi'_{x_0,x_1}(t)(t'-t)$$

in order to prove that φ_{x_0,x_1} is convex. If we denote by $x = tx_1 + (1-t)x_0$ and $y = t'x_1 + (1-t')x_0$, the left side of the previous inequality can be interpreted as $f(y) - f(x)$, while the right side is equal to $\langle \nabla f(x), y - x\rangle$. Since φ_{x_0,x_1} is convex, it follows from Remark 4.1 that f is convex. $\square$

The next exercise propose an alternative proof of the converse implication.

Exercise 4.12 (**) Let f be a differentiable mapping defined on the convex set U, we assume that for all $(x,x') \in U \times U$, $f(x) - f(x') \geq \langle \nabla f(x'), x - x'\rangle$.

1. Show that

$$\mathrm{epi}(f) = \bigcap_{x' \in U} \{(x,y) \in U \times \mathbb{R} \mid y \geq f(x') + \langle \nabla f(x'), x - x'\rangle\}.$$

2. Using the characterization given by Theorem 4.2, deduce that f is convex.

Exercise 4.13 (***) Let f be a function defined on some convex set U, and let $\overline{x}$ be an interior point. Let us define the set (cf. Exercise 3.15)

$$\partial f(\overline{x}) = \{\alpha \in \mathbb{R}^n \mid f(x) \geq f(\overline{x}) + \langle \alpha, x - \overline{x} \rangle, \text{ for all } x \in U\}$$

In order to simplify the notations, it may be easier to consider the case $\overline{x} = 0$.

1. Apply Proposition 4.4 in order to show that if f is convex and differentiable at point $\overline{x}$, then $\nabla f(\overline{x}) \in \partial f(\overline{x})$.
2. If f is differentiable at point $\overline{x}$, write the Taylor expansion of degree 1 at a neighborhood of $\overline{x}$, and deduce that $\alpha \in \partial f(\overline{x}) \Rightarrow \alpha = \nabla f(\overline{x})$.
3. We assume that f is convex and that $\partial f(\overline{x}) = \{\alpha\}$, show that f is differentiable at point $\overline{x}$ and that $\nabla f(\overline{x}) = \alpha$.

Proposition 4.5 *Let f be a differentiable function on some convex set U. The function f is convex if and only if for all $(x, y) \in U \times U$,*

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0.$$

Proof of Proposition 4.5. The implication $\Rightarrow$ is a consequence of Remark 4.1. Indeed, we know that the function $\varphi_{x,y}$ is convex, which implies $\varphi'_{x,y}(1) - \varphi'_{x,y}(0) \geq 0$. This leads to the conclusion.

For the converse implication, let us consider x and y two elements of U. We know that for all t of $]0, 1[$,

$$\langle \nabla f(ty + (1-t)x) - \nabla f(x), (ty + (1-t)x) - x \rangle \geq 0,$$

and consequently $\langle \nabla f(ty + (1-t)x), y - x \rangle \geq \langle \nabla f(x), y - x \rangle$. If we integrate on $[0, 1]$ this inequality, we get

$$\int_0^1 (\langle \nabla f(ty + (1-t)x), y - x \rangle dt \geq \langle \nabla f(x), y - x \rangle,$$

this means $f(y) - f(x) - \langle \nabla f(x), x - y \rangle \geq 0$. Using Proposition 4.4, we can deduce that f is convex. □

Let us now consider the case where f is twice differentiable on U. We will denote by $H_f(x)$ the hessian matrix at point x, we recall that this matrix is symmetric when f is $\mathcal{C}^2$.

Definition 4.6 Let M be a symmetric matrix (n, n).

- We say that M is positive semidefinite (respectively negative semidefinite) if for all $v \in \mathbb{R}^n$, $\langle v, Mv \rangle \geq 0$ (respectively $\langle v, Mv \rangle \leq 0$).
- We say that M is positive definite (respectively negative definite) if for all $v \in \mathbb{R}^n \setminus \{0\}$, $\langle v, Mv \rangle > 0$ (respectively $\langle v, Mv \rangle < 0$).

Proposition 4.6 *Let M be a symmetric matrix (n, n).*

- *the matrix M has n eigenvalues (possibly equal),*
- *the matrix M is positive semidefinite (respectively negative semidefinite) if and only if all eigenvalues are non-negative, (respectively non-positive),*
- *the matrix M is positive definite (respectively negative definite) if and only if all eigenvalues are positive, (respectively negative).*

Exercise 4.14 (*) Let M be a symmetric $(n, n)-$matrix, then

- if M is positive semidefinite, then $\det M \geq 0$ and $Tr(M) \geq 0$.
- if M is positive definite, then $\det M > 0$ and $Tr(M) > 0$.
- if M is negative semidefinite, then $Tr(M) \leq 0$ and $\det M \leq 0$ when n is odd, $\det M \geq 0$ when n is even.
- if M is negative definite, then $Tr(M) < 0$ and $\det M < 0$ when n is odd, $\det M > 0$ when n is even.

Exercise 4.15 (*) When the dimension is equal to 2, we can reenforce the conclusion of the previous exercise. Let $M = \begin{pmatrix} r & s \\ s & t \end{pmatrix}$ be a symmetric $(2, 2)-$matrix, then

- the matrix M is positive semidefinite if and only if $\det M \geq 0$ and $Tr(M) \geq 0$.

- the matrix M is positive definite if and only if $\det M > 0$ and $Tr(M) > 0$.

Exercise 4.16 (*) Let M be the following matrix,

$$M = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Show that $\det M > 0$ and $Tr(M) > 0$, and that M is not positive semidefinite.

Proposition 4.7 *Let M be a symmetric matrix, then M is positive definite if and only if for all $k = 1, \ldots, n$, $\Delta_k > 0$, where*

$$\Delta_k = \begin{vmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \ldots & a_{kk} \end{vmatrix}$$

Exercise 4.17 (*) 1) Let M be a symmetric matrix, such that M is positive semidefinite, then for all $k = 1, \ldots, n$, $\Delta_k \geq 0$, where

$$\Delta_k = \begin{vmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \ldots & a_{kk} \end{vmatrix}$$

2) Let M be the following matrix,

$$M = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Prove that for all $k = 1, \ldots, 3$, $\Delta_k \geq 0$, but that M is not positive semidefinite.

Proposition 4.8 *Let f be a twice continuously differentiable function on U, which is a convex and open subset of $\mathbb{R}^n$. Then, f is convex if and only if for all $x \in U$, $H_f(x)$ is positive semidefinite.*

Proof of Proposition 4.8. Indeed, if f is convex, then for any x in U and v in $\mathbb{R}^n$, there exists $\varepsilon > 0$, such that $y = x + \varepsilon v \in U$. We can consider the function $\varphi_{x,y}$ which is convex. We can deduce that $\langle \varepsilon v, H_f(x) \varepsilon v \rangle \varphi''_{x,y}(0) \geq 0$. This implies that $H_f(x)$ is positive semidefinite.

Conversely, let us assume that at each point of U, $H_f(x)$ is positive semidefinite. Then the computation of $\varphi''_{x,y}$ (cf. Remark 4.1) allows us to deduce that $\varphi_{x,y}$ is convex. □

A simple adaptation of the previous proof allows us to deduce the following proposition:

Proposition 4.9 *Let f be a twice continuously differentiable function on some convex set U,*

1) If for all $x \in U$, $H_f(x)$ is positive definite, then f is strictly convex.

2) If at point $\overline{x}$, the hessian matrix is positive definite, then f is strictly convex in some neighborhood of $\overline{x}$.

4.3 Unconstrained optimization

4.3.1 Classical case

The problem $\max_{x \in U} f(x)$ is said unconstrained if the set U is an open subset of $\mathbb{R}^n$.

Remark 4.2 Let us consider the unconstrained problem $\min_{x \in U} f(x)$ and $\overline{x} \in U$. The point $\overline{x}$ is a local solution of this problem if and only if it is an interior point of the set $\{x \in U \mid f(x) \geq f(\overline{x})\}$.

Remark 4.3 The previous remark can only be used for unconstrained optmization: for example, $\overline{x} = (0, 0)$ is a local solution of the problem $\min_{x \in U} f(x)$ (where $U = \mathbb{R}^2_+$ NOT open and $f(x) = x_1^2 + x_2^2$), but the corresponding set $\{x \in U \mid f(x) \geq f(\overline{x})\}$ is reduced to $\{\overline{x}\}$ and the interiory property does not hold.

Proposition 4.10 *Let us consider the unconstrained problem* $\min_{x \in U} f(x)$*.*

- *If* $\overline{x}$ *is a local solution, and if* f *is differentiable at point* $\overline{x}$*, then* $\nabla f(\overline{x}) = 0$*, ("critical point")*

- *If* $\overline{x}$ *is a local solution, and if* f *is twice continuously differentiable at point* $\overline{x}$*, then* $H_f(\overline{x})$ *is positive semidefinite,*

- *If* $\overline{x}$ *is a critical point, if* f *is twice continuously differentiable at point* $\overline{x}$*, and if* $H_f(\overline{x})$ *is positive definite, then* $\overline{x}$ *is a local solution. In addition, there exists some* $\varepsilon > 0$*, such that on* $B(\overline{x}, \varepsilon) \subset U$*,* $\overline{x}$ *is the unique solution.*

- *If* $\overline{x}$ *is a critical point and if* f *is convex, then* $\overline{x}$ *is a global solution.*

Proof of Proposition 4.10

It suffices to remark that if $\overline{x}$ is a local solution, then for any direction $u \in \mathbb{R}^n$, there exists some positive real number $\varepsilon > 0$, such that the function

$$\begin{array}{rl} \varphi: &]-\varepsilon, \varepsilon[\to \mathbb{R} \\ & \varphi(t) = f(\overline{x} + tu) \end{array}$$

is defined, and that 0 is a (global) minimum.

If f is differentiable at point $\overline{x}$, then φ has a derivative and $\varphi'(0) = 0$. In view of the chain rule theorem, $\varphi'(t) = \langle \nabla f(\overline{x} + tu), u \rangle$. This leads to $\langle \nabla f(\overline{x}), u \rangle \geq 0$. Since this is true for any $u \in \mathbb{R}^n$, we can conclude that $\nabla f(\overline{x}) = 0$.

If f is twice differentiable at point $\overline{x}$, then φ has a second derivative and $\varphi''(0) \geq 0$. In view of the chain rule theorem, $\varphi''(t) = \langle u, H_f(\overline{x} + tu)(u) \rangle$. This leads to $\langle u, H_f(\overline{x})(u) \rangle = 0$. Since this is true for any $u \in \mathbb{R}^n$, we can conclude that $H_f(\overline{x})$ is positive semidefinite.

Note that some results of unconstrained optimization can be used even if the domain is not open (cf. Exercise 1.5).

Remark 4.4 Let $\overline{x}$ is a local solution of the problem $\min_{x \in X} f(x)$. When X is a neighborhood of $\overline{x}$, there exists some V open subset of C containing $\overline{x}$ such that $\overline{x}$ is a local solution of the problem $\min_{x \in V} f(x)$. This allows to apply the first and second order necessary conditions.

Exercise 4.18 We can reenforce the third result of the previous proposition. For all k strictly smaller than the smallest eigenvalue of the Hessian matrix $H_f(\overline{x})$ (in particular, k can be chosen positive), there exists some $\varepsilon > 0$ such that for all $x \in B(\bar{x}, r)$, $f(x) \geq f(\bar{x}) + \frac{k}{2}\|x - \bar{x}\|^2$.

4.3.2 Extension to affine constraints

Let us consider the following problem

$$\left\{ \begin{array}{l} \min f(x) \\ x \in U, Ax = b \end{array} \right.$$

The set of feasible points C is $\{x \in U \mid Ax = b\}$, where A is a $(n \times p)$-matrix and $b \in \mathbb{R}^p$. This correspond to the case of p affine constraints represented by the p rows of A and the vector b. If we denote by a_j the vector de $\mathbb{R}^n$ corresponding to the j-row of A, the set C is defined by:

$$x \in C \text{ if and only if } \langle a_j, x \rangle - b_j = 0 \text{ for all } j = 1, \ldots, p.$$

If we denote by $g_j(x) = \langle a_j, x \rangle - b_j$ (which are affine functions), and $J = \{1, \ldots, p\}$, the problem can be written as

$$\left\{ \begin{array}{l} \min f(x) \\ g_j(x) = 0, \forall j \in J \\ x \in U \end{array} \right.$$

Note that C can be written as $U \cap \Lambda^{-1}(b)$ where Λ is the linear mapping defined by $x \to \Lambda(x) = Ax$. Let $\overline{x}$ be a local solution of the problem:

$$\left\{ \begin{array}{l} \min f(x) \\ Ax = b \\ x \in U \end{array} \right.$$

For all $u \in \operatorname{Ker}\Lambda$, note that for all $t \in \mathbb{R}$, $A(\bar{x} + tu) = b$.

Since U is open et $\overline{x} \in U$, there exists some $\varepsilon > 0$ such that for all $t \in]-\varepsilon, \varepsilon[$, $\bar{x} + tu \in U$. Once again, we can consider

$$\begin{array}{rl} \varphi_u : &]-\varepsilon, \varepsilon[\to \mathbb{R} \\ & \varphi_u(t) = f(\overline{x} + tu) \end{array}$$

Consequently, for all $t \in]-\varepsilon, \varepsilon[$, $\varphi_u(t) := f(\overline{x}+tu) \geq \varphi_u(0) = f(\overline{x})$, which means that 0 is a minimum de φ_u. This implies that $\varphi'_u(0) = 0$. We can compute $\varphi'_u(0) = \nabla f(\bar{x}) \cdot u = 0$.

Let $(u_1, \ldots u_k)$ be a basis of the kernel

Note that when U is equal to $\mathbb{R}^n$, we can entirely describe C which is equal to $\{\overline{x}\} + \operatorname{Ker}\Lambda$. This implies that 0 is a solution of

$$\begin{cases} \min f(\overline{x} + \lambda_1 u_1 + \ldots + \lambda_k u_k) \\ \lambda_1 \in \mathbb{R}, \ldots, \lambda_k \in \mathbb{R} \end{cases}$$

In the general case, C is equal to $U \cap (\{\overline{x}\} + \operatorname{Ker}\Lambda)$. This implies that there exists $\varepsilon > 0$, such that 0 is a solution of

$$\begin{cases} \min h(\lambda) \\ \lambda_1 \in \mathbb{R}, \ldots, \lambda_k \in \mathbb{R} \\ \lambda \in B(0, \varepsilon) \end{cases}$$

where $h(\lambda) = f(\overline{x} + \lambda_1 u_1 + \ldots + \lambda_k u_k)$.

This leads to $\nabla h(0) = 0$, and we can formulate this result as $\nabla f(\bar{x})$ is orthogonal to $\operatorname{Ker}\Lambda$. It follows that $\nabla f(\overline{x})$ is in the image of Λ^t, the transposed of de φ. We recall that the matrix of Λ^t is the matrix where the rows are the columns of A, the vectors a_j. We have the result, there exists some vector $\overline{\lambda}$ in $\mathbb{R}^p$ (vector of Lagrange multipliers) such that:

$$\nabla f(\bar{x}) = \sum_{j=1}^{p} \bar{\lambda}_j a_j$$

Since $a_j = \nabla g_j(\overline{x})$, this means that

$$\nabla f(\bar{x}) = \sum_{j=1}^{p} \bar{\lambda}_j \nabla g_j(\overline{x})$$

4.4 Karush-Kuhn-Tucker

4.4.1 Necessary condition

Let us consider the domain defined by the list of constraints

$$\begin{cases} f_i(x) = 0, \ \forall i \in I \\ g_j(x) \leq 0, \ \forall j \in J \\ x \in U \end{cases}$$

Among the inequality constraints, we will distinguish those that correspond to affine function. This can be formulated as J is equal to the partition $J_a \cup J_{na}$, where g_j is affine when $j \in J_a$. The set U is open which may hide strict inequalities.

Definition 4.7 We say that the Slater condition is satisfied if

- all f_i are affine,
- all g_j are convex
- there exists $\overline{x}$ be a feasible point of the previous system, satisfying moreover $g_j(\overline{x}) < 0$, for all $j \in J_{na}$.

Remark 4.5 When all the constraints are affine, then the Slater's condition is equivalent to to the existence of a feasible point.

Theorem 4.3 (Karush-Kuhn-Tucker) *Let us consider the optimization problem:*

$$\begin{cases} \min f(x) \\ f_i(x) = 0, \ \forall i \in I \\ g_j(x) \leq 0, \ \forall j \in J \\ x \in U \end{cases}$$

where U is an open set. Let $\overline{x}$ be a local solution of this problem. We assume that there exists V open neighborhood of $\overline{x}$ such that f, f_i, g_j are continuously differentiable on V. If the Slater condition is satisfied then there exists $\overline{\lambda} \in \mathbb{R}^I$ and $\overline{\mu} \in \mathbb{R}_+^J$ such that

$$\nabla f(\bar{x}) + \sum_{i \in I} \bar{\lambda}_i \nabla f_i(\overline{x}) + \sum_{j \in J} \overline{\mu}_j \nabla g_j(\overline{x}) = 0 \tag{4.1}$$

and $\overline{\mu}_j g_j(\overline{x}) = 0$*, for all* $j \in J$*.*

Note that if we distinguish among the inequalities, the set of binding constraints at point $\overline{x}$ by $J(\overline{x}) = \{j \in J \mid g_j(\overline{x})\}$, then the condition 4.1 can be rewritten as,

$$\nabla f(\overline{x}) + \sum_{i \in I} \overline{\lambda}_i \nabla f_i(\overline{x}) + \sum_{j \in J(\overline{x})} \overline{\mu}_j \nabla g_j(\overline{x}) = 0. \tag{4.2}$$

Note that when $j \notin J(\overline{x})$, in view of the "complementary condition" ($\overline{\mu}_j g_j(\overline{x}) = 0$), we can deduce that the multiplier $\overline{\mu}_j$ is equal to zero.

Definition 4.8 We will associate to the previous minimization problem, the following system "Karush-Kuhn-Tucker conditions"

$$\begin{cases} f_i(\overline{x}) = 0, \forall i \in I \\ g_j(\overline{x}) \leq 0, \forall j \in J \\ \overline{\lambda}_i \in \mathbb{R}, \forall i \in I \\ \overline{\mu}_j \geq 0, \forall j \in J \\ \overline{\mu}_j g_j(\overline{x}) = 0, \text{ for all } j \in J \\ \nabla f(\overline{x}) + \sum_{i \in I} \overline{\lambda}_i \nabla f_i(\overline{x}) + \sum_{j \in J} \overline{\mu}_j \nabla g_j(\overline{x}) = 0 \\ \overline{x} \in U \end{cases}$$

Exercise 4.19 1. Prove that for all $z \in \mathbb{R}^2$,

$$\begin{cases} z_1 \geq 0, \\ z_2 \geq 0, \\ z_1 + z_2 = 0, \end{cases} \Leftrightarrow \begin{cases} z_1 = 0, \\ z_2 = 0. \end{cases}$$

2. Prove that for any $p \geq 1$, for any $\alpha \in \mathbb{R}^p_+$, $\beta \in \mathbb{R}^p_+$,

$$\sum_{i=1}^{p} \alpha_i \beta_i = 0 \Leftrightarrow \alpha_i \beta_i = 0, \text{ for all } i = 1, \ldots, p$$

3. Prove that the Karush-Kuhn-Tucker system is equivalent to there exists $(\overline{x}, \overline{\lambda}, \overline{\mu}) \in U \times \mathbb{R}^I \times \mathbb{R}^J_+$ such that

$$\begin{cases} f_i(\overline{x}) = 0, \forall i \in I \\ g_j(\overline{x}) \leq 0, \forall j \in J \\ \sum_{j \in J} \overline{\mu}_j g_j(\overline{x}) = 0, \\ \nabla f(\overline{x}) + \sum_{i \in I} \overline{\lambda}_i \nabla f_i(\overline{x}) + \sum_{j \in J} \overline{\mu}_j \nabla g_j(\overline{x}) = 0 \end{cases}$$

4.4.2 Sufficient conditions for optimality

Definition 4.9 We say that the problem

$$\begin{cases} \min f(x) \\ x \in U, \\ h_1(x) = 0 \\ \ldots \\ h_q(x) = 0 \\ g_1(x) \leq 0 \\ \ldots \\ g_p(x) \leq 0 \end{cases}$$

is convex if the functions h_j are affine, the functions g_i are convex, and if f is convex [1].

When the problem is convex, we can state:

Proposition 4.11 *Let us consider the optimization problem:*

$$\begin{cases} \min f(x) \\ g_1(x) \leq 0 \\ \ldots \\ g_p(x) \leq 0 \\ x \in U \end{cases}$$

Let us assume that $\overline{x}$ *be a feasible point and that* U *is an open set. If the function* f *is convex on* U *(globally convex), (respectively locally convex at point* $\overline{x}$*), and if the functions* g_i *are convex on* U*, (respectively locally convex at point* $\overline{x}$*) and if there exists* $(\overline{\lambda}_i)_{i=1}^p \in \mathbb{R}^p_+$ *such that*

$$\begin{cases} \nabla f(\overline{x}) + \overline{\lambda}_1 \nabla g_1(\overline{x}) + \ldots + \overline{\lambda}_p \nabla g_p(\overline{x}) = 0 \\ \overline{\lambda}_i g_i(\overline{x}) = 0 \text{ for all } i = 1, \ldots, p. \end{cases}$$

then, the condition is sufficient, $\overline{x}$ *is a global solution (respectively local) of the minimization problem.*

Proof of the proposition 4.11. In order to simplify the notation, we

[1] for a maximization problem, the objective function is concave.

will focus on the global case. Let us call the "Lagrangian" of the problem

$$\mathcal{L}(x,\lambda) = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x).$$

We can consider the partial function $H(x) = \mathcal{L}(x, \overline{\lambda})$, The function H is convex on U, and $\overline{x}$ is a critical point of H. So for each $x \in U$ (feasible or not), $H(\overline{x}) \leq H(x)$.

We can conclude, if in addition we notice that for each feasible point x, $g_i(x) \leq 0 \Rightarrow \overline{\lambda}_i g_i(x) \leq 0 \Rightarrow \mathcal{L}(x, \overline{\lambda}) \leq f(x)$ and that $\mathcal{L}(\overline{x}, \overline{\lambda}) = f(\overline{x})$ (in view of the complementary relations). □

Corollary 4.1 *Let us consider the optimization problem:*

$$\begin{cases} \min f(x) \\ x \in U \\ f_i(x) = 0,\ i \in I \\ g_j(x) \leq 0,\ j \in J \end{cases}$$

Let us assume that $\overline{x}$ be a feasible point and that U is an open set. If the function f is (respectively locally at point $\overline{x}$) convex, the functions f_i are affine, if for all $j \in J(\overline{x})$ functions g_j are convex on U (respectively locally at point $\overline{x}$), and if there exists $(\lambda) \in \mathbb{R}^I$ and $\mu \in \mathbb{R}_+^{J(\overline{x})}$ such that

$$\begin{cases} \nabla f(\overline{x}) - \sum_{i\in I} \lambda_i \nabla f_i(\overline{x}) + \sum_{j \in J(\overline{x})} \mu j \nabla g_j(\overline{x}) = 0 \\ \mu_j g_j(\overline{x}) = 0 \text{ for all } j \in J. \end{cases}$$

then, the condition is sufficient, (respectively sufficient) $\overline{x}$ is a solution (repectively local solution) of the minimization problem.

We can extend in the next exercise Corollary 4.1 by assuming that the objective function is only quasi-concave but only if the point is not a critical point (cf. Exercise 3.18).

Exercise 4.20 Let us consider $(\overline{x}, \overline{\lambda}, \overline{\mu})$ solution of Karush-Kuhn-Tucker system (Definition 4.8) associated to the minimization problem

$$\begin{cases} \min f(x) \\ f_i(x) = 0,\ \forall i \in I \\ g_j(x) \leq 0,\ \forall j \in J \\ x \in U \end{cases}$$

We assume in addition that the objective function is quasi-concave on U and continuous on U, the functions f_i are affine, and for all $j \in J$, the functions g_j are convex on U, The goal of the exercise is to show by contradiction that if $\nabla f(\overline{x}) \neq 0$, then $\overline{x}$ is a solution of the minimization problem. With no loss of generality, we may assume that $\overline{x} = 0$.

1. Let us assume that there exists some feasible point y such that $f(y) < f(0)$, prove the existence of some $\varepsilon > 0$ such that for all $z \in B(y, \varepsilon)$, $f(z) < f(\overline{x})$.

2. Let us fix some $z \in B(y, \varepsilon)$. Show that the function $\varphi := \varphi_{0,z}$ is quasi-concave and consequently $\varphi'(0) \leq 0$.

3. Deduce from $\nabla f(\overline{x}) + \sum_{i \in I} \lambda_i \nabla f_i(\overline{x}) + \sum_{j \in J(\overline{x})} \mu j \nabla g_j(\overline{x}) = 0$ that $\varphi'(0) \geq 0$.

4. Compute $\varphi'(0)$ and deduce that $\nabla f(0) \perp B(y, \varepsilon)$. Conclude.

4.5 Polarity and orthogonality

4.5.1 Separation theorems

Theorem 4.4 *Let A and B be two nonempty disjoint convex subsets of $\mathbb{R}^n$. If A is compact and B is closed, then there exists $y \in E$, $y \neq 0$ such that:*

$$\sup\{y \cdot a \mid a \in A\} < \inf\{y \cdot b \mid b \in B\}$$

Theorem 4.5 *Let A and B be two nonempty disjoint convex subsets of $\mathbb{R}^n$. Then there exists a hyperplane separating A and B.*

$y \in E$, $y \neq 0$ such that

$$\sup\{y \cdot a \mid a \in A\} \leq \inf\{y \cdot b \mid b \in B\}$$

4.5.2 definitions

Definition 4.10 Let A be a subset of $\mathbb{R}^n$, the polar cone of A, denoted by A° is defined by

$$A^\circ = \{y \in E \mid \text{ for all } a \in A, y \cdot a \leq 0\}$$

The orthogonal of A, denoted by $A^\perp$ is defined by

$$A^\perp = \{y \in E \mid \text{ for all } a \in A, y \cdot a = 0\}$$

Proposition 4.12 *Let A be a subset of $\mathbb{R}^n$*

1. *A° is a closed convex cone (of vertex 0),*
2. *$A^\perp$ is a linear subspace,*
3. *$A^\perp = A^\circ \cap (-A)^\circ$;*
4. *If $A_1 \subset A_2$, then $A_2^\circ \subset A_1^\circ$ and $A_2^\perp \subset A_1^\perp$;*
5. *If $A \neq \emptyset$, $A^\circ = (\mathrm{cl}(\mathrm{K}(A)))^\circ$ and $A^\perp = (\mathrm{vect}(A))^\perp$ where $\mathrm{vect}(A)$ is the linear subspace spanned by A.*

Exercise 4.21 (*) Prove Proposition 4.12.

Exercise 4.22 Let A be a nonempty subset of $\mathbb{R}^n$

1. $A \subset A^{\circ\circ} := (A^\circ)^\circ$
2. $A^\circ = A^{\circ\circ\circ} := ((A^\circ)^\circ)^\circ$
3. $A^{\circ\circ} = A^{\circ\circ\circ\circ} := (((A^\circ)^\circ)^\circ)^\circ$

Exercise 4.23 Let A be a nonempty subset of $\mathbb{R}^n$

1. If A is a cone subspace, then y belongs to A° if and only if $a \to y \cdot a$ is bounded from above on A.
2. If A is a linear subspace, then $A^\circ = A^\perp$; y belongs to $A^\perp$ if and only if $a \to y \cdot a$ is bounded from above on A. A vector y belongs to $A^\perp$ if and only if $a \to y \cdot a$ is bounded from below on A.

Theorem 4.6 (Bipolar Theorem) *If A be a nonempty subset of $\mathbb{R}^n$, then, $(A^\circ)^\circ = \mathrm{cl}(\mathrm{K}(A))$ and $(A^\perp)^\perp = \mathrm{vect}(A)$, where $\mathrm{vect}(A)$ is the linear subspace spanned by A.*

Corollary 4.2 *Let A be a nonempty subset of $\mathbb{R}^n$. The set A is a closed convex cone if and only if $A = (A^\circ)^\circ$. The set A is a linear subspace if and only if $A = (A^\perp)^\perp$.*

Corollary 4.3 *Let M be a linear subspace of $\mathbb{R}^n$, $M = E$ if and only if $M^\perp = \{0\}$.*

Exercise 4.24 Let M be a convex set of $\mathbb{R}^n$. M is a neighborhood of 0 if and only if $M^\circ = \{0\}$.

4.5.3 Farkas' lemma

Theorem 4.7 (Farkas' lemma) *Let $(a_i)_{i\in I}$ be a finite family of elements in $\mathbb{R}^n$. Let*

$$A = \{\sum_{i\in I} \lambda_i a_i \mid \lambda \in \mathbb{R}_+^I\}$$

and

$$B = \{y \in E \mid y \cdot a_i \leq 0, \quad \text{for all } i \in I\}$$

one has $A^\circ = B$ and $B^\circ = A$.

The key argument is Exercise 4.26.

Exercise 4.25 (**) Let $(a_1, \ldots, a_p)$ be a finite subset of $E = \mathbb{R}^n$.
1) Show that $\mathrm{co}(a_1, \ldots, a_p)$ is a compact set.
2) Extend this result if E is a normed vector space.

Exercise 4.26 (***) Let $(a_1, \ldots, a_p)$ be a finite subset of $E = \mathbb{R}^n$.

The goal of this exercise is to show by induction on p that the set $K(a_1, \ldots, a_p)$ which denotes $\{\sum_{i=1}^p \lambda_i a_i \mid \lambda \in R_+^p\}$ is closed.

A) 1) show that if $-a_{p+1} \in K(a_1, \ldots, a_p)$ then

$$K(a_1, \ldots, a_{p+1}) = \bigcup_{i=1}^{p+1} K(\{a_j \mid j \neq i\}).$$

2) show that if M and N are two closed convex cones of E, such that $M \cap (-N) = \{0\}$ then $M + N$ is closed.

B) 1) Show that the result is true when $p = 1$.

2) Write the proof (one will have to distinguish wether $-a_{p+1}$ belongs or not to $K(a_1, \ldots, a_p)$).

Corollary 4.4 (Second version of Farkas' lemma) *Let $(a_i)_{i\in I}$ and $(b_j)_{j\in J}$ be finite families of elements in $\mathbb{R}^n$. Let*

$$A = \{\sum_{i\in I} \lambda_i a_i + \sum_{j\in J} \mu_j b_j \mid \lambda \in \mathbb{R}^I_+,\ \mu \in \mathbb{R}^J\}$$

et

$$B = \{y \in E \mid y \cdot a_i \leq 0, \ \ \textit{for all } i \in I,\ y \cdot b_j = 0,\ \textit{for all } j \in J\}$$

one has $A^\circ = B$ and $B^\circ = A$.

Corollary 4.5 *Let $(a_i)_{i\in I}$ be a finite family of elements in $\mathbb{R}^n$. Let*

$$A = \{\sum_{i\in I} \lambda_i a_i \mid \lambda \in \mathbb{R}^I\}$$

and

$$B = \{y \in E \mid y \cdot a_i = 0,\ \text{pour tout } i \in I\}$$

One has, $A^\perp = B$ and $B^\perp = A$.

Exercise 4.27 Let $(a_i)_{i=1,\ldots,p}$ be a finite family of elements in $\mathbb{R}^n$. Let $b \in \mathbb{R}^n$ satisfying the following property:

For all $y \in \mathbb{R}^n$

$$\begin{cases} y \cdot a_1 \leq 0 \\ \ldots \\ y \cdot a_p \leq 0 \end{cases} \Rightarrow y \cdot b \leq 0.$$

Prove (using the hint) that there exists $\lambda \in \mathbb{R}^I_+$ such that $b = \sum_{i\in I} \lambda_i a_i$.

Hint: Prove that 0 is a solution of the following optimization problem and write the Karush-Kuhn-Tucker conditions.

$$\begin{cases} \max y \cdot b & y \cdot a_1 \leq 0 \\ & \ldots \\ & y \cdot a_p \leq 0 \end{cases}$$

4.6 Tangent and normal cones

Let us now introduce two definitions.

Definition 4.11 Let C be a convex set containing c. The tangent cone to the set C at point c denoted by $T_C(c)$ is

$$T_C(c) = \text{cl}\{t(c' - c) \mid t \geq 0,\ c' \in C\}$$

The normal cone to the set C at point c denoted by $N_C(c)$ is:

$$N_C(c) = \{y \in E \mid y \cdot c \geq y \cdot c'\ \forall c' \in C\}.$$

Exercise 4.28 (*) (i) Let C be a convex set containing c, and D denotes is translation $D := C + \{-c\}$. Show that $T_C(c) = T_D(0)$ and $N_C(c) = N_D(0)$.

(ii) Let $c \in A \subset B$, then $T_A(c) \subset T_B(c)$ and $N_B(c) \subset N_A(c)$.

Proposition 4.13 *(i) For all $c \in C$, $T_C(c)$ and $N_C(c)$ are nonempty, closed and convex cones.*

(ii) $T_C(c) = (N_C(c))^\circ$ and $N_C(c) = (T_C(c))^\circ$.

It is very important to emphasize that these notions are local notions, cf. next exercise.

Exercise 4.29 (*) Let C_1 and C_2 be convex sets containing c, if there exists $\varepsilon > 0$ such that $C_1 \cap B(c,\varepsilon) = C_2 \cap B(c,\varepsilon)$ then $N_{C_1}(c) = N_{C_2}(c)$ and $T_{C_1}(c) = T_{C_2}(c)$.

Exercise 4.30 (*) Let us consider $\min x^2 + y^2$ under the constraint $2x + y \leq -4$.

1) Is the Slater condition satisfied for this problem ?

2) Write the Karush-Kuhn-Tucker conditions of this problem.

3) Solve the Karush-Kuhn-Tucker system.

4) Discuss wether the conditions are necessary/sufficient in order to solve the optimization problem.

Exercise 4.31 (**)
let us consider the following optimization problems:

$$(\mathcal{P}) \begin{cases} \max & 3x_1x_2 - x_2^3 \\ s.c. & x_1 \geq 0 \\ & x_2 \geq 0 \\ & x_1 - 2x_2 = 5 \\ & 2x_1 + 5x_2 \geq 20 \end{cases} \qquad (\mathcal{Q}) \begin{cases} \max & 3x_1x_2 - x_2^3 \\ s.c. & x_1 > 0 \\ & x_2 > 0 \\ & x_1 - 2x_2 = 5 \\ & 20 - 2x_1 - 5x_2 \leq 0 \end{cases}$$

1) Draw the set of feasible points for $(\mathcal{P})$ and comment.

2) Prove that x is feasible for $(\mathcal{P})$ if and only if x is feasible for $(\mathcal{Q})$. Deduce that the two problems are equivalent.

3) Prove that $\mathcal{S}ol(\mathcal{P})$ is nonempty.

4) Write the Karush-Kuhn-Tucker conditions.

5) Solve the Karush-Kun-Tucker conditions.

6) Solve the optimization problem $(\mathcal{P})$

Chapter 5

Linear Programming

A *linear programming* problem consists in finding the maximum (resp. minimum) value of a linear functional, subject to a finite number of linear constraints. If c and a^i, $i = 1, \ldots, m$ are elements of $\mathbb{R}^n$, and if $b = (b_1, \ldots, b_m)$ belongs to $\mathbb{R}^m$, the most general form of a linear programming problem is the following:

Maximize (resp. minimize)

$$f(x) = c \cdot x$$

subject to the conditions

$$\begin{array}{ll} a^i \cdot x & \begin{pmatrix} \leq \\ = \\ \geq \end{pmatrix} \quad b_i, \quad i = 1, \ldots, m \\ x & \in \mathbb{R}^n. \end{array}$$

The linear functional $f(x)$ is called the *objective function*. The linear equations and inequations are called *constraints* of the problem. The set of points that satisfy these linear equations or inequations is called *feasible set*. An element of this set is a *feasible solution*; it is an *optimal solution* if it solves the maximization (resp. minimization) problem. Let (P) and (P') be two optimization problems, we recall that (P) and (P') are *equivalent* if their sets of solutions are equal.

Note that in general, we can not use Weierstrass theorem neither a coercivity property of the objective function.

5.1 The main results

5.1.1 existence

Note that in general, we can not use Weierstrass theorem neither a coercivity property of the objective function.

Proposition 5.1 *Given a linear programming problem (P) (maximization), one of the three following alternatives holds:*

- either there is no feasible point for Problem (P), the value is equal to $-\infty$, $\mathcal{S}ol(P) = \emptyset$.

- either the objective function is not bounded from above on the nonempty set of feasible points, the value is equal to $+\infty$, $\mathcal{S}ol(P) = \emptyset$.

- either the objective function is bounded from above on the nonempty set of feasible points, the value is finite and the set of solutions is nonempty.

5.1.2 Necessary and sufficient conditions of optimality

Let f^k (resp. f'^k) $k = 1, \ldots, q$, h^k (resp. h'^k) $k = 1, \ldots, r$, g (resp. g') be linear functional on $\mathbb{R}^n$, α_k (resp. α'_k) $k = 1, \ldots, q$, β_k (resp. β'_k) $k = 1, \ldots, r$ be real numbers. We now consider linear programming problems expressed in the following form:

$$(I)\qquad \begin{array}{l}\text{Minimize } g(x) \text{ subject to}\\ f^k(x) = \alpha_k,\ k = 1,\dots,q\\ h^k(x) \geq \beta_k,\ k = 1,\dots,r\\ x \in \mathbb{R}^n.\end{array}$$

Proposition 5.2 *For a feasible solution $\overline{x}$ to be an optimal solution of Problem (I), it is both necessary and sufficient that there exist some KKT multipliers $\mu_1,\dots,\mu_q \in \mathbb{R}$ and $\nu_1,\dots,\nu_r \in \mathbb{R}_+$.*

Note that the Slater condition is satisfied if and only if there exists a feasible point.

Corollary 5.1 *For a feasible solution $\overline{x}$ to be an optimal solution of Problem (I), it is both necessary and sufficient that there exist $\mu_1,\dots,\mu_q \in \mathbb{R}$ and $\nu_1,\dots,\nu_r \in \mathbb{R}_+$ such that $g = \sum_{k=1}^q \mu_k f^k + \sum_{k=1}^r \nu_k h^k$, with $\nu_k = 0$ if $h^k(\overline{x}) > \beta_k$.*

5.2 Saddle point properties

$$(P)\qquad \begin{array}{l}\min f(x) \text{ subject to}\\ f_i(x) \leq 0,\ i \in I\\ x \in \mathbb{R}^n.\end{array}$$

The *Lagrangian* of minimization Problem (P) is the function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^I_+ \to \mathbb{R}$ defined by for all $(x,\lambda) = (x,(\lambda_i)_{i\in I})$,

$$\mathcal{L}(x,\lambda) = f(x) + \sum_{i\in I} \lambda_i f_i(x).$$

We say that $(\overline{x},\overline{\lambda})$ is a *saddle-point* of $\mathcal{L}$ if for all $\forall\lambda \in \mathbb{R}^I_+, \forall x \in \mathbb{R}^n$,

$$\mathcal{L}(x,\overline{\lambda}) \geq \mathcal{L}(\overline{x},\overline{\lambda}) \geq \mathcal{L}(\overline{x},\lambda).$$

Theorem 5.1 *(i) If $(\overline{x},\overline{\lambda})$ is a saddle-point of $\mathcal{L}$, then $\overline{x}$ is a solution to (P) and $(\overline{\lambda})$ are the Karush-Kuhn-Tucker coefficients associated with $\overline{x}$. Moreover, $L(\overline{x},\overline{\lambda}) = f(\overline{x})$.*
(ii) If $(\overline{x},\overline{\lambda})$ satisfies Karush-Kuhn-Tucker Conditions for Problem (P), then $(\overline{x},\overline{\lambda})$ is a saddle-point of $\mathcal{L}$.

In this case, the dual problem can be defined as

$$(D)\qquad \begin{array}{c}\max G(\lambda) \text{ subject to}\\ \lambda \in \mathbb{R}^I_+.\end{array}$$

where

$$G(\lambda) = \inf_{x\in\mathbb{R}^n} \mathcal{L}(x,\lambda)$$

Note that $G(\overline{\lambda}) = \mathcal{L}(\overline{x},\overline{\lambda}) = f(\overline{x}) = \text{val}(P)$

5.3 The duality theorem of linear programming

5.3.1 Canonical form of the duality theorem

Let A be a $(m \times n)$-matrix, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$. Let us consider the following linear programming problem (P), here referred to as the *primal problem*,

$$(P)\qquad \begin{array}{l}\text{Minimize } c\cdot x \text{ subject to}\\ Ax \geq b\\ x \geq 0\\ x \in \mathbb{R}^n.\end{array}$$

and define its *dual problem* (D)

$$(D)\qquad \begin{array}{l}\text{Maximize } b\cdot p \text{ subject to}\\ {}^t pA \leq {}^t c\\ p \geq 0\\ p \in \mathbb{R}^m.\end{array}$$

Writing down the dual of the dual leads to the given primal problem. For this reason, the linear programming problems (P) and (D) are referred to as *primal problem* and *dual problem* in *canonical form*. Each one of both problems is said to be the *dual* of the other one.

Proposition 5.3 *Given the pair of dual linear programming problems (P) and (D), one of the two following alternatives holds:*

- either (P) and (D) have a couple $(\overline{x}, \overline{p})$ of optimal solutions satisfying: $\overline{p} \cdot b = c \cdot \overline{x}$ (obviously, the same relation $p \cdot b = c \cdot x$ holds then for every couple (x, p) of optimal solutions since the primal and the dual have the same value);

- or neither (P) nor (D) has an optimal solution and one of both feasible sets is empty.

In what follows, the duality theorem is used to establish an important result on linear inequalities which extends Exercise 4.27.

5.3.2 Application to nonhomogeneous Farkas' lemma

Exercise 5.1 (***) Let for every $i = 0, 1, \ldots, m$, $a^i \in \mathbb{R}^n$, $\alpha_i \in \mathbb{R}$ be such that the system $a^i \cdot x \leq \alpha_i,\ i = 1, \ldots, m$ is consistent. Then in order that the following implication is true

$$a^i \cdot x \leq \alpha_i,\ i = 1, \ldots, m \Rightarrow a^0 \cdot x \leq \alpha_0 \tag{5.1}$$

it is necessary and sufficient that there exist nonnegative real numbers $\lambda_i \geq 0$ such that

$$a^0 = \sum_{i=1}^{m} \lambda_i a^i \text{ and } \alpha_0 \geq \sum_{i=1}^{m} \lambda_i \alpha_i.$$

Exercise 5.2 Let $p_1 > 0$, $p_2 > 0$ and the optimization problem

$$(\mathcal{P}) \begin{cases} \min & -x_1 - x_2 \\ s.c. & x_1 \geq 0,\ x_2 \geq 0 \\ & p_1 x_1 + p_2 x_2 \leq 1 \end{cases}$$

1) Write the Karush-Kuhn-Tucker conditions of $(\mathcal{P})$, solve the primal.

2) Write the dual and solve it.

Chapter 6

Brief presentation of Scilab

6.1 The linear case

6.1.1 Economical motivation

Exemple 6.1 Let us consider a firm which is able to produce two kinds of output (we do not face indivisibility problems). Each output needs to spend some time on three workshops.

- Output A requires to spend 1 hour on the first workshop $P1$, 5 hours on $P2$ and 1 hour on $P3$.

- Output B requires to spend 4 hour on the first workshop $P1$, 2 hours on $P2$ and 2 hours on $P3$.

The actual renting contract for $P1$ allows us to use it during 10 hours per day, 20 hours for $P2$ and 6 hours for $P3$. The unit profit for output A is (respectively B) is 2 € (respectively 3 €).

Let us denote by x and y the quantities of output A and output B. The profit maximization can be modelled as

$$\left\{\begin{array}{ll} \max & 2x+3y \\ \text{s.t.} & x \geq 0 \\ & y \geq 0 \\ & x+4y \leq 10 \\ & 5x+2y \leq 20 \\ & x+2y \leq 6 \end{array}\right. \qquad (\mathcal{P}_1)$$

6.1.2 Graphic solving

In order to determine the set of feasible points, we will draw the line D_1 whose equation is $x + 4y = 10$, D_2 whose equation is $5x + 2y = 20$ and finally D_3 whose equation is $x + 2y = 6$. We can draw the figure 6.1.

The graphical analysis (figure 6.2) allows us to determine the optimal point of this problem. The optimum is unique ("classical case"), the point is M_3 defined by the intersection of D_2 and D_3. We determine its coordinates by solving the following system:

$$\left\{\begin{array}{rcrcr} 5x & + & 2y & = & 20 \\ x & + & 2y & = & 6 \end{array}\right.$$

We can solve and find the production (3,5, 1,25), which corresponds to a profit of $10,75$€.

At this point, we can determine the multipliers $\lambda \in \mathbb{R}^5_+$ which are solutions of

$$\left\{\begin{array}{l} \lambda_1(-\overline{x}) = 0 \\ \lambda_2(-\overline{y}) = 0 \\ \lambda_3(\overline{x}+4\overline{y}-10) = 0 \\ \lambda_4(5\overline{x}+2\overline{y}-20) = 0 \\ \lambda_5(\overline{x}+2\overline{y}-6) = 0 \\ \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \lambda_1 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ -1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 1 \\ 4 \end{pmatrix} + \lambda_4 \begin{pmatrix} 5 \\ 2 \end{pmatrix} + \lambda_5 \begin{pmatrix} 1 \\ 2 \end{pmatrix} \end{array}\right.$$

Since the constraints 1, 2 and 3 are not binding, the associated multipliers

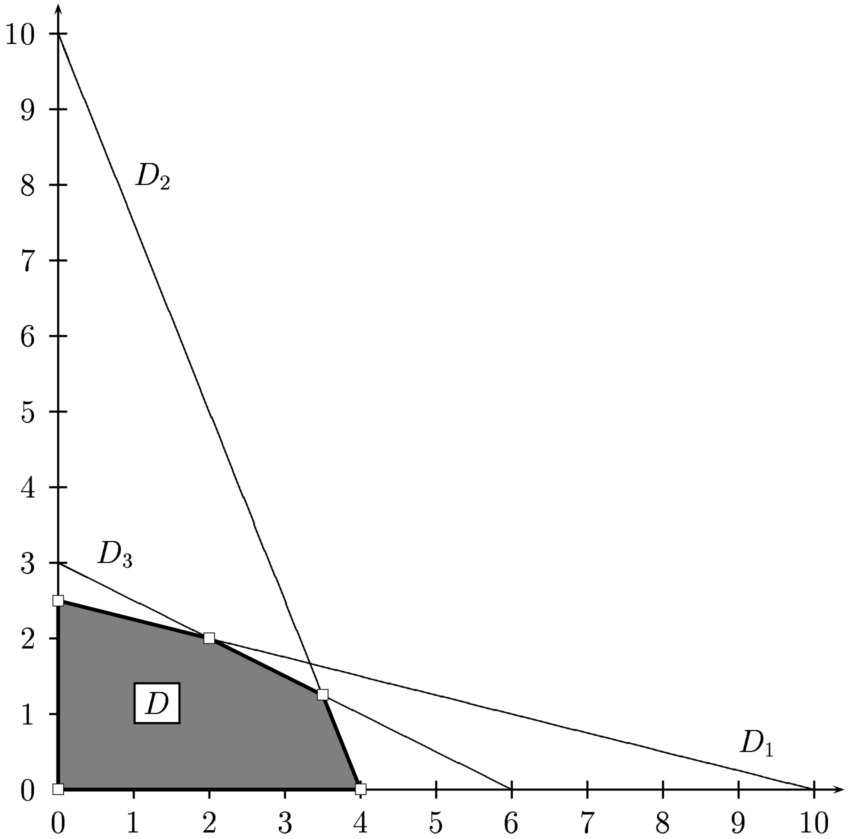

Figure 6.1: feasible set of Problem $(\mathcal{P}_1)$.

are equal to zero.

$$\begin{cases} \lambda_1 = 0 \\ \lambda_2 = 0 \\ \lambda_3 = 0 \\ \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \lambda_4 \begin{pmatrix} 5 \\ 2 \end{pmatrix} + \lambda_5 \begin{pmatrix} 1 \\ 2 \end{pmatrix} \end{cases}$$

One gets $\lambda_4 = 0.125$ and $\lambda_5 = 1.375$.

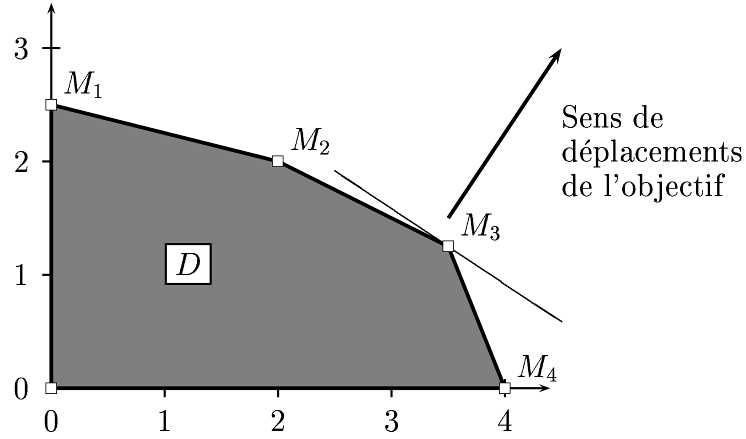

Figure 6.2: Optimal point of Example 6.1

6.1.3 Scilab Solver

Let us use the scilab software (`www.scilab.org`). "Scilab is a scientific software package for numerical computations providing a powerful open computing environment for engineering and scientific applications. Scilab is an open source software. Since 1994 it has been distributed freely along with the source code via the Internet. It is currently used in educational and industrial environments around the world."

Let us transform our problem as

$$\begin{cases} \min -2x_1 - 3x_2 \\ \text{s.t.} \quad x \geq 0 \\ \qquad Ax \geq b \end{cases} \qquad (\mathcal{P}_1)$$

where A is a matrix with three rows and 2 columns and b a column vector

$$A = \begin{pmatrix} 1 & 4 \\ 5 & 2 \\ 1 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 10 \\ 20 \\ 6 \end{pmatrix}.$$

Let us write the following text file "exemple1.sci"

```
// My first Scilab program
P=-[2;3];
b=[10;20;6];
A=[1,4;5,2;1,2]
Z_sup=[]; //max value for x, empty here which means that
there is not such constraints
Z_inf=[0;0]; //min value for x

[Zopt,lag,CA]=linpro(P,A,b,Z_inf,Z_sup)
```

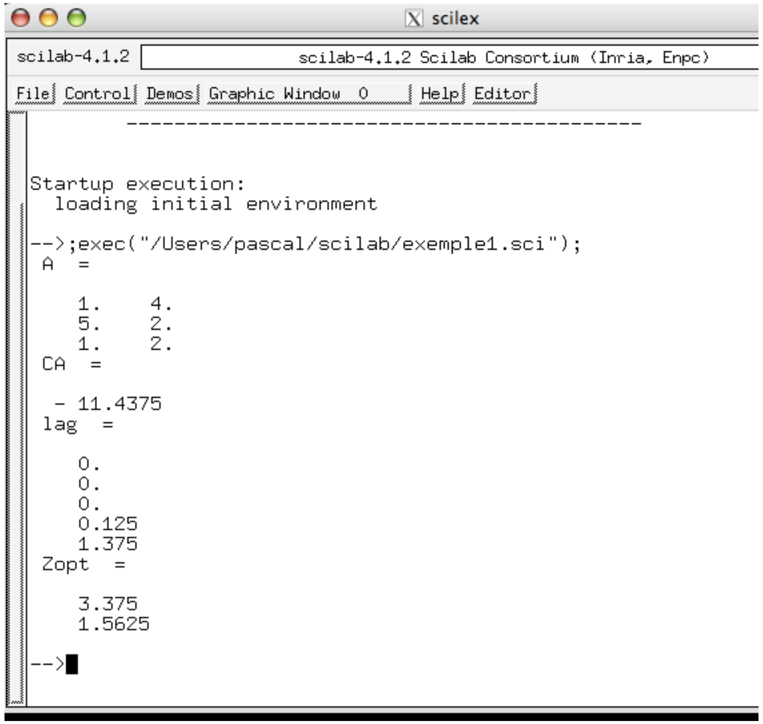

Figure 6.3: Numerical solution

It is easy to understand the syntax of Scilab for defining matrices, the coefficients are written by rows and separated by a semicolon. The characters `//` explains that what follows has to be treated as comments. At the end of a line, when there is a semicolon, the result will not be printed on the screen. The syntax of the `linpro` instruction can be found with the help (cf. figure 6.4).

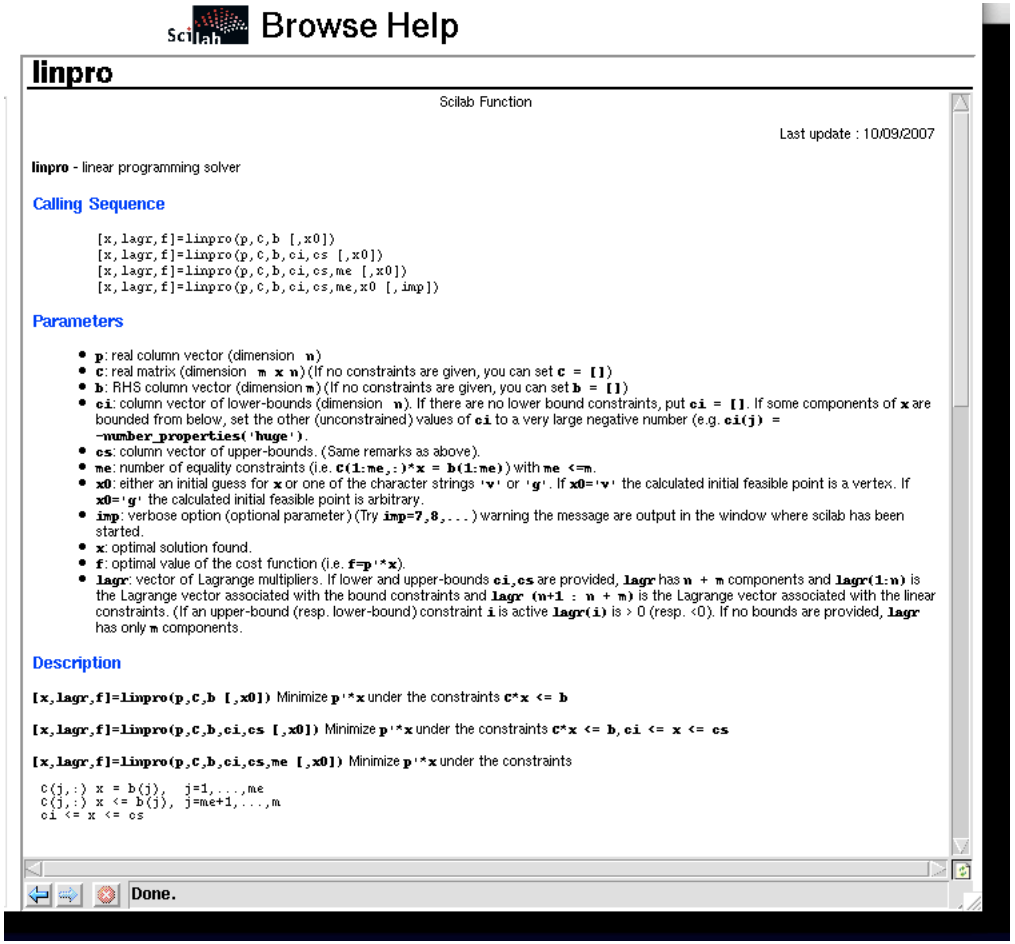

Figure 6.4: help window for Instruction `linpro`.

In the software, we use the menu "File", in order to select the file "exemple1.sci". The software does not only give us the point $(\overline{x}, \overline{y})$ solution but also the value of the problem and a vector of multipliers (which are not necessarily unique).

6.1.4 Sensibility analysis

Let us consider now the modified problem ($\mathcal{P}_2$), when the workshop $P1$ is available during 11 hours. How will move the production (solution of the problem) and the profit (value of the problem)

$$\begin{array}{ll} \max 2x+3y & \\ \text{s.t.} & x \geq 0 \\ & y \geq 0 \\ & x+4y \leq 11 \\ & 5x+2y+ \leq 20 \\ & x+2y \leq 6 \end{array} \qquad (\mathcal{P}_2)$$

This problem is very similar to the original one ($\mathcal{P}_1$) (same objective function, almost same set of feasible points). On the figure 6.5, one can understand that the domain has been transform by a translation of one of the lines defining the original set of feasible points. On the picture 6.6,

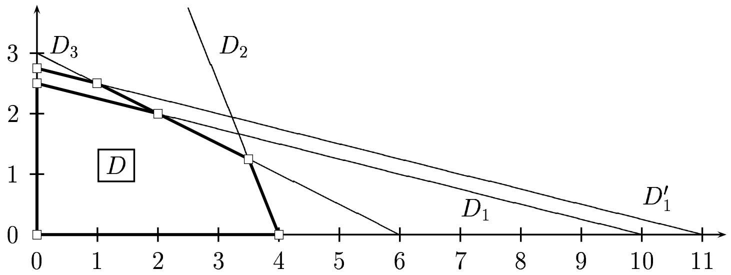

Figure 6.5: modification of the feasible corresponding to the translation of Line D_1

it is easy to remark that the change for the set of feasible points does not imply a change on the coordinates of the optimal point which is still M_3 (3,5, 1,25). Consequently the profit is unchanged.

Remark 6.1 For large values of the available time of Workshop 1, this constraint is not binding, while it is binding for small enough values.

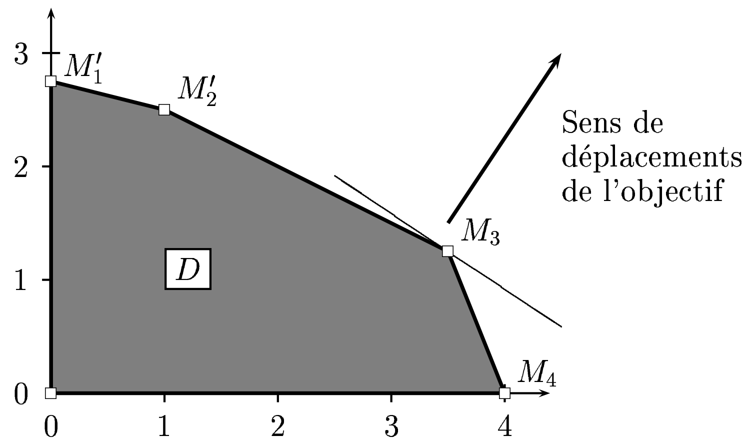

Figure 6.6: Graphical solving of Problem ($\mathcal{P}_2$)

6.1.5 Second modification of the constraints

Let us consider the same problem as ($\mathcal{P}_1$) except that now, the third workshop will be available during 6 hours and a half. How will move the production (solution of the problem) and the profit (value of the problem)

This problem can be written as:

$$\begin{array}{ll} \max 2x+3y & \\ \text{s.t.} & x \geq 0 \\ & y \geq 0 \\ & x+4y \leq 10 \\ & 5x+2y+ \leq 20 \\ & x+2y \leq 6,5 \end{array} \qquad (\mathcal{P}_3)$$

On the picture 6.7, we present the new set of feasible points.

On the picture 6.8, we can check that the new optimal point is the point defined by

$$\begin{cases} 5x + 2y = 20 \\ x + 2y = 6.5 \end{cases}$$

The solution is (3,375, 1,3125), and corresponds to a profit of 11, 4375. The additional profit is equal to 0, 6875 €. If the renting of this additional half hour costs less than 0, 6875 €, renting this additional half hour is a profitable operation. The “marginal” price of an additonal hour needs

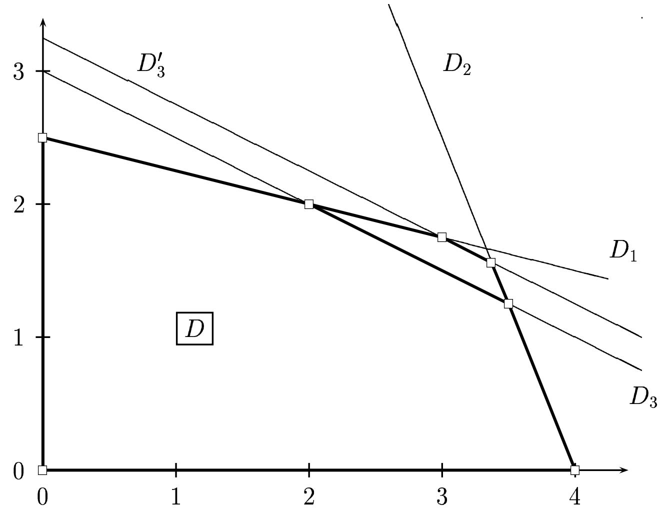

Figure 6.7: feasible set corresponding to $(\mathcal{P}_3)$

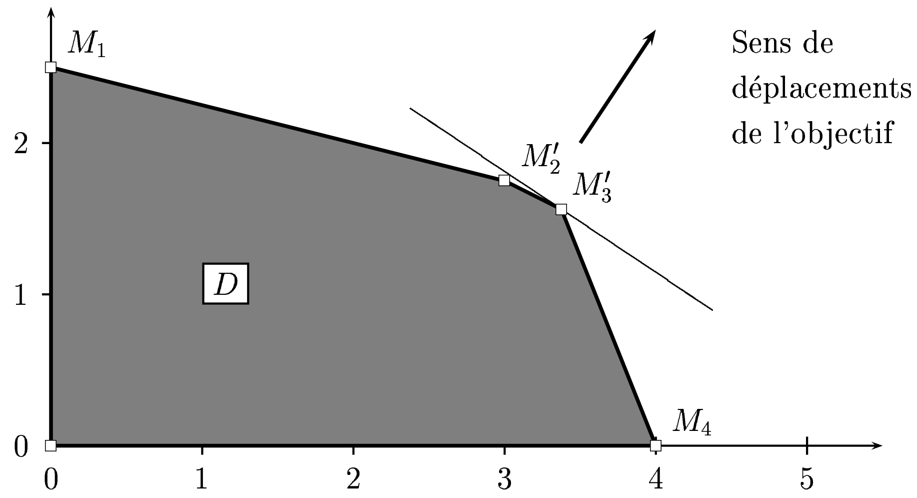

Figure 6.8: graphical solving of $(\mathcal{P}_3)$

to be less than $2 * 0.6875 = 1.375$. Note that this corresponds to the value of the multiplier associated with the third workshop (it is the case here because the multiplier is unique and because the relaxation of the constraint is not too large).

Each constraint has an implicit price, the shadow price, which is given by the multiplier.

6.1.6 Alternative program

In order to propose several simulations without changing our scilab program, we can consider a new program

```
// A second program
labels=["Workshop 1";"2";"3"];
[ok,u,v,w]=getvalue("define parameters",labels,...
list("vec",1,"vec",1,"vec",1),["10";"20";"6"]);
P=-[2;3];
A=[1,4;5,2;1,2];
Z_sup=[]; //max value for x, empty here which means that
there is not such constraints
Z_inf=[0;0]; //min value for x
b=[u;v;w]
// ------------------------
// Calcul de l'optimum
// ------------------------
[Zopt,lag,CA]=linpro(P,A,S,Z_inf,Z_sup)
```

The instruction "getvalue" allows to type in a window the values of the parameters (See Figure 6.9). This allows us to solve both $(\mathcal{P}_2)$ and

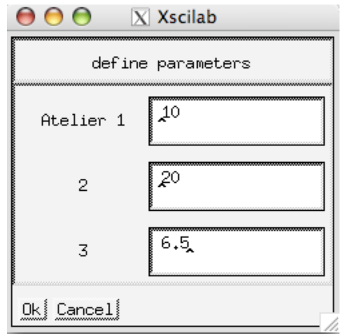

Figure 6.9: the parameters window

($\mathcal{P}_{3)}$ (See Figure 6.10).

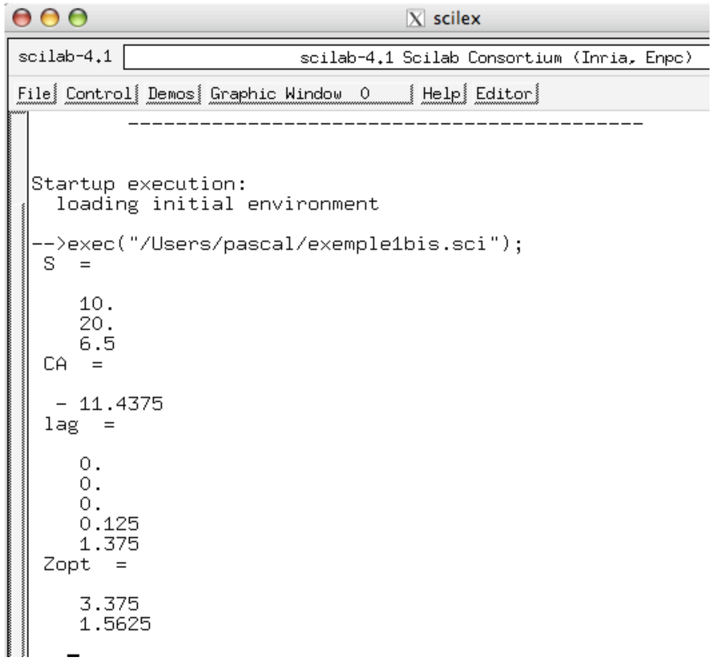

Figure 6.10: Solving ($\mathcal{P}_{3)}$

6.2 The non linear case

6.3 Unconstrained optimization

Let us consider the problem ($\mathcal{P}_5$) $\min x^2 + y^2$. We can use the following scilab program :

```
function [f,g,ind]=cost(x,ind)
f=x(1)2̂+x(2)2̂, g=[2*x(1);2*x(2)]
endfunction;

// g is the gradient of f
// here, ind is an unused parameter but which has to be
written
[f,xopt]=optim(cost,[1;2])
```

Let us start with initial point $(1, 2)$.

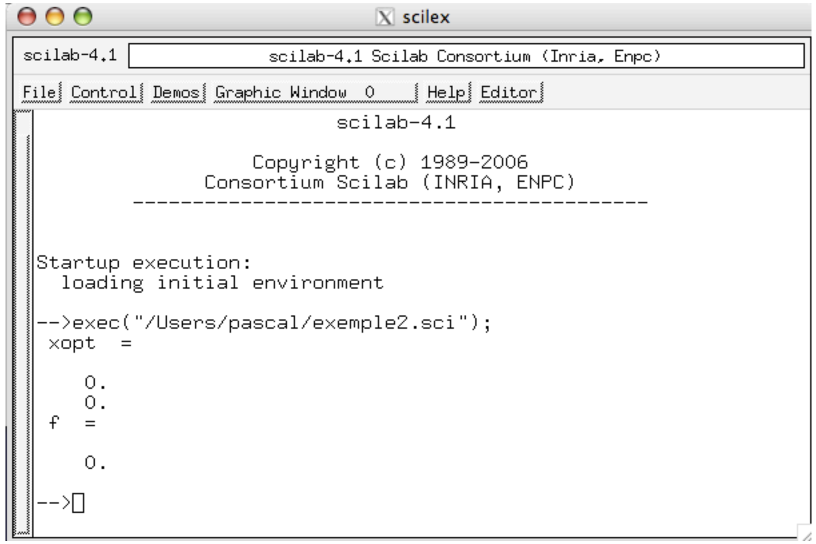

Figure 6.11: Solving ($\mathcal{P}_{5)}$

6.3.1 Unconstrained Optimization

Let us consider
$$\begin{array}{ll} \max 3x + 3y & \\ \text{sous contraintes} & x \in [2, 10] \\ & y \in [-10, 10] \end{array} \qquad (\mathcal{P}_6)$$

The scilab program will give us a 3-dimensional plotting and the solution.

```
   // programme exemple3.sci
function [f,g,ind]=cost(x,ind)
f=x(1)^2+x(2)^2, g=[2*x(1);2*x(2)]
endfunction;

function [z]=C(x,y) , z=x^2+y^2, endfunction;
x=2:10;y=-10:10;
z=feval(x,y,C);
xbasc();
plot3d(x,y,z);

[f,xopt,gopt]=optim(cost,'b',[2;-10],[10;10],[5;5])
// f n'est pas nul
// The gradient at the opt point is ``normal''
// to the boundary
```

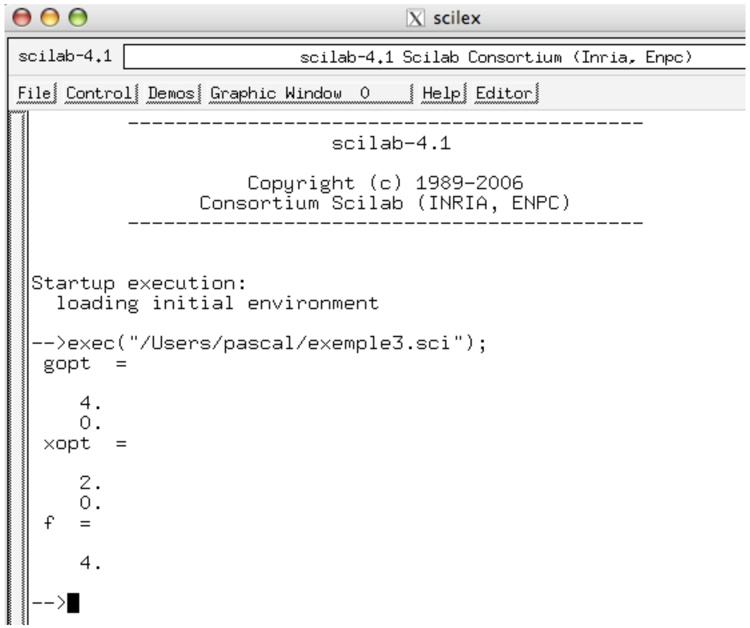

Figure 6.12: Solving $(\mathcal{P}_6)$

We remark that at the optimal point, the gradient of the objective function is not equal to zero, (it is $(4,0)$).

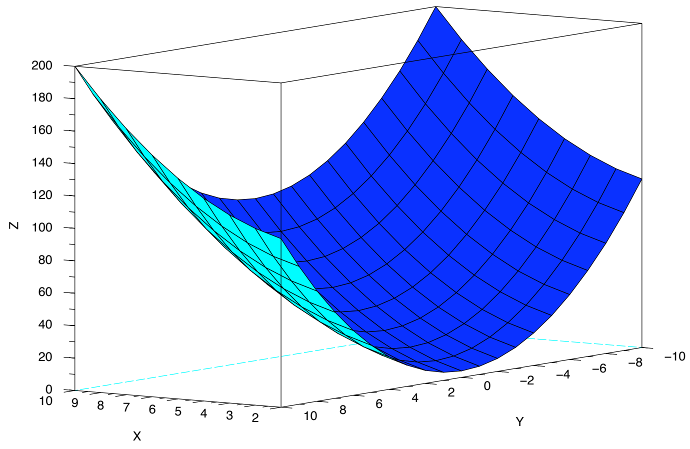

Figure 6.13: graph of objective function of $(\mathcal{P}_6)$

Contents

1 Presentation of Optimization **1**
- 1.1 Mathematical presentation . . . 1
- 1.2 Examples of economic problems . . . 4

2 Convexity of sets **6**
- 2.1 Definition . . . 6
- 2.2 First properties . . . 8
- 2.3 Stability with respect to affine functions . . . 9
- 2.4 Convex hull, affine hull and conic hull . . . 10

3 The one dimensional results **13**
- 3.1 Existence results . . . 13
- 3.2 First order condition . . . 13
- 3.3 Convex and concave functions . . . 15
- 3.4 Sufficient conditions for optimality . . . 20
- 3.5 Introduction to sensibility analysis . . . 21

4 Finite dimensional optimization **23**
- 4.1 Existence results . . . 23
- 4.2 Convex and quasi-convex functions . . . 24
- 4.3 Unconstrained optimization . . . 28
- 4.4 Karush-Kuhn-Tucker . . . 30
- 4.5 Polarity and orthogonality . . . 32
- 4.6 Tangent and normal cones . . . 34

5 Linear Programming **36**
- 5.1 The main results . . . 36
- 5.2 Saddle point properties . . . 37
- 5.3 The duality theorem of linear programming . . . 37

6 Brief presentation of Scilab **39**
- 6.1 The linear case . . . 39
- 6.2 The non linear case . . . 44
- 6.3 Unconstrained optimization . . . 44