Analyse Convexe et Applications

Polycopié de cours

Christophe Dutang

Support de cours de L3 – Mathématique à l'Université du Mans entre 2013 et 2017

Notes basées sur les cours de Robert Dalmasso, Geir Dahl et Anatoli Iouditski

Janvier 2022

CC BY NC SA

Table des matières

Polycopié de cours

Chapitre 1

Ensembles convexes (2,5 séances)

1.1 Ensembles convexes

1.1.1 Ensembles convexes

Définition 1.1.1 (Combinaison convexe)**.** *Soit m points $x_1, \ldots, x_m$ de $\mathbb{R}^n$. Une combinaison (finie) convexe des points $x_1, \ldots, x_m$ est le point $x = \sum_{i=1}^m \lambda_i x_i$ tel que pour tout $i = 1, \ldots, m$, $\lambda_i \in [0, 1]$ et $\sum_{i=1}^m \lambda_i = 1$.*

Remarque 1.1.1 (Cas particulier $m = 2$)**.** *Une combinaison convexe de deux points x_1, x_2 de $\mathbb{R}^n$ est $x = \lambda x_1 + (1 - \lambda)x_2$ et $\lambda \in [0, 1]$. Si $n = 1$, alors les points x pour $\lambda \in [0, 1]$ décrivent l'intervalle $[x_1, x_2]$ ou $[x_2, x_1]$. Si $n \geq 2$, alors les points x pour $\lambda \in [0, 1]$ décrivent le segment $[x_1, x_2]$.*

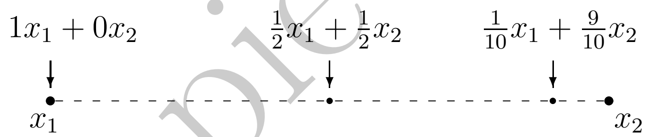


FIGURE 1.1 – Combinaison convexe de deux éléments

Remarque 1.1.2 (Cas particulier $m = 3$)**.** *Une combinaison convexe de trois points x_1, x_2, x_3 de $\mathbb{R}^m$ est $x = \lambda x_1 + \mu x_2 + (1 - \lambda - \mu)x_3$ et $\lambda, \mu \in [0, 1]$. Si $n = 1$, alors les points x pour $\lambda \in [0, 1]$ décrivent l'intervalle $[\min_i x_i, \max_i x_i]$. Si $n = 2$ et les points sont non-alignés, alors les points x pour $\lambda \in [0, 1]$ décrivent le triangle de sommet x_1, x_2 et x_3*

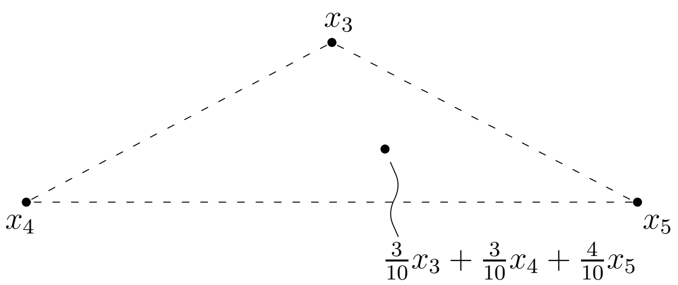


FIGURE 1.2 – Combinaison convexe de trois éléments

Polycopié de cours

Définition 1.1.2 (Ensemble convexe)**.** *Un ensemble $C \subset \mathbb{R}^n$ est un ensemble convexe si toute combinaison convexe de deux éléments de C est un point de C, i.e.*

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

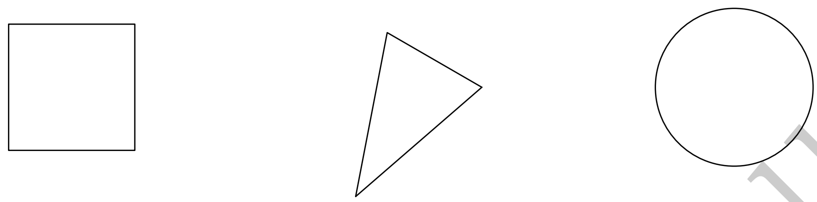


FIGURE 1.3 – Exemple d'ensembles convexes

Proposition 1.1.3. *C est un ensemble convexe $\Leftrightarrow \forall x_1, \ldots, x_m \in C, \forall \lambda_i \in [0,1], \sum_{i=1}^{m} \lambda_i = 1 \Rightarrow \sum_{i=1}^{m} \lambda_i x_i \in C$.*

Démonstration. $\Leftarrow$ évident
$\Rightarrow$ La propriété est vérifiée pour $m = 2$. Par récurrence sur m, si la propriété est vérifiée pour m, considérons $\lambda_1, \ldots, \lambda_{m+1} \in [0,1]$ tels que $\sum_{i=1}^{m+1} \lambda_i = 1$. On pose $\lambda = \lambda_{m+1}$ que l'on suppose $\lambda > 0$ sans perte de généralité. Le point

$$y = \sum_{i=1}^{m} \frac{\lambda_i}{1-\lambda} x_i$$

est un point de C puisque

$$\sum_{i=1}^{m} \frac{\lambda_i}{1-\lambda} = \frac{\sum_{i=1}^{m} \lambda_i}{1-\lambda_{m+1}} = \frac{1-\lambda_{m+1}}{1-\lambda_{m+1}} = 1.$$

Ainsi

$$z = \lambda x_{m+1} + (1-\lambda) y \in C \text{ et } z = \lambda_{m+1} x_{m+1} + \sum_{i=1}^{m} \lambda_i x_i.$$

$\square$

Proposition 1.1.4 (Opérations préservant la convexité)**.** *Si A et B sont convexes, alors $A + B$ et $A \cap B$ sont des ensembles convexes.*

Démonstration. Voir TD1. $\square$

1.1.2 Application en optique

La forme des lentilles peuvent vérifier la convexité (figures 1.5a et 1.5b) ou non (figures 1.5c et 1.5d).

Polycopié de cours

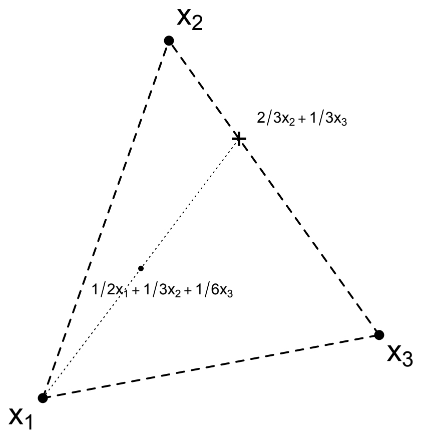
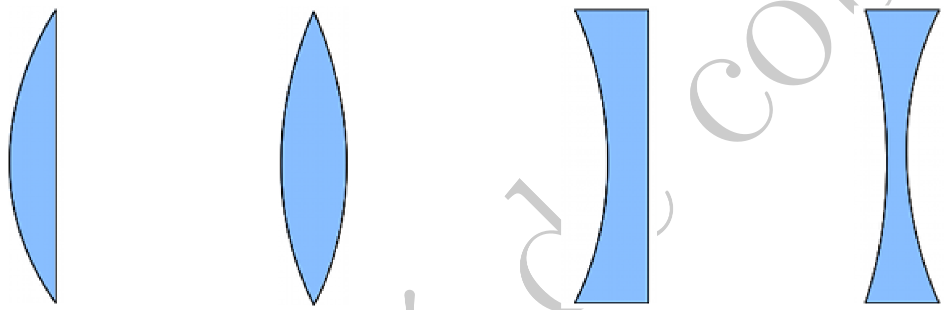


FIGURE 1.4 – Lien entre combinaison convexe de deux et trois éléments



(a) plan-convexe (b) biconvexe (c) plan-concave (d) biconcave

FIGURE 1.5 – Formes de lentilles

1.1.3 Cones, polyèdres, polytopes

Définition 1.1.3 (Cone convexe)**.** *Un ensemble $C \subset \mathbb{R}^n$ est un cone convexe si toute combinaison positive de deux éléments de C est un point de C, i.e.*

$$\forall x, y \in C, \forall \lambda, \mu \geq 0, \lambda x + \mu y \in C.$$

Définition 1.1.4 (polyèdre)**.** *Un polyèdre de $\mathbb{R}^n$ est l'ensemble $\{x \in \mathbb{R}^n, Ax \leq b\}$ pour $A \in \mathbb{R}^{p \times n}$ et $b \in \mathbb{R}^p$. Autrement dit c'est l'ensemble de points vérifiant des inégalités linéaires de leurs composantes, voir par exemple figure 1.10d.*

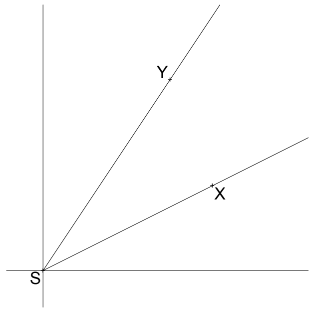
Exemple 1.1.5. *Si $n = 1$, les cones convexes sont les demi-droites du type $[0, +\infty[$ ou $]-\infty, 0]$. Si $n = 2$, les cones convexes sont les cones détermines par le sommet $S = (0,0) \in \mathbb{R}^2$ et deux droites SX et SY pour 2 points $X = (x_1, x_2)$ et $Y = (y_1, y_2)$. Sur la figure 1.6a, on a choisit $X = (2/3, 1/3), Y = (1/2, 3/4)$.*

Proposition 1.1.5. *Les cônes et les polyèdres sont des ensembles convexes.*

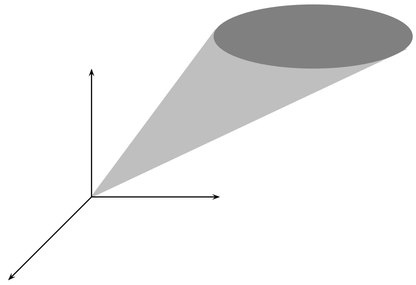
Démonstration. Soient $x, y \in C$ pour un cône C et $\lambda \in [0, 1]$. On a

$$\begin{cases} \lambda \geq 0 \\ \mu = 1 - \lambda \geq 0 \\ x \in C \\ y \in C \end{cases} \Rightarrow \lambda x + \mu y \in C.$$

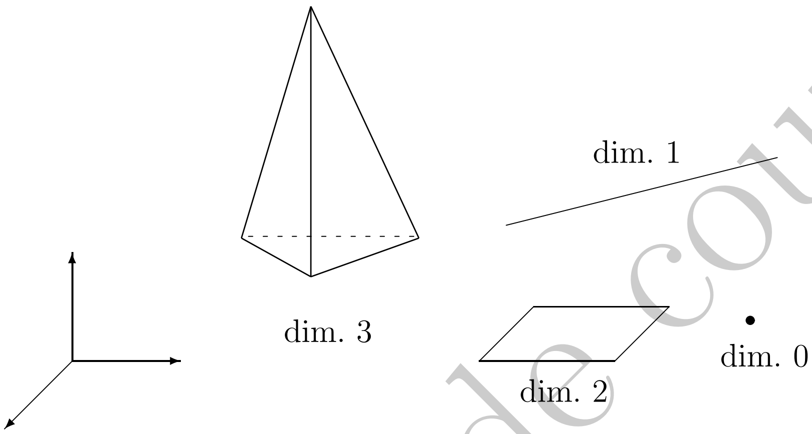
Polycopié de cours



(a) Cone dans le plan



(b) Cone dans l'espace



(c) Polytopes

Figure 1.6 – Cônes et polytopes du plan et de l'espace

Polycopié de cours

Donc $\lambda x + (1-\lambda)y \in C$.

Soient $x, y \in P$ pour un polyèdre $P = \{x, Ax \leq b\}$ et $\lambda \in [0,1]$. On a

$$\left\{\begin{array}{l} x \in P \\ y \in P \\ \lambda \geq 0 \\ 1-\lambda \geq 0 \end{array}\right. \Rightarrow \left\{\begin{array}{l} A(\lambda x) = \lambda Ax \leq \lambda b \\ A((1-\lambda)y) = (1-\lambda)Ay \leq (1-\lambda)b \end{array}\right. \Rightarrow A(\lambda x + (1-\lambda)y) \leq b$$

Donc $\lambda x + (1-\lambda)y \in P$. □

Proposition 1.1.6. *C est un cône convexe $\Leftrightarrow \forall x_1, \dots, x_m \in C, \forall \lambda_i \geq 0, \sum_{i=1}^m \lambda_i x_i \in C$.*

Démonstration. Preuve similaire à la proposition 1.1.6. □

Définition 1.1.6 (Cône polyhèdrique)**.** *Un cône polyhèdrique de $\mathbb{R}^n$ est l'ensemble $\{x \in \mathbb{R}^n, Ax \leq 0\}$. Il est de la forme*

$$\left\{\sum_{i=1}^m \lambda_i x_i, \forall i = 1, \dots, m, \lambda_i \geq 0\right\},$$

pour $x_i \in \mathbb{R}^n$.

1.1.4 Enveloppes convexes

Définition 1.1.7 (Enveloppe convexe)**.** *Pour un ensemble $S \subset \mathbb{R}^n$, l'enveloppe convexe de S notée* co(S) *est l'ensemble de toute combinaison convexe d'éléments de S, i.e.* co(S) $= \{s, \exists \lambda_1, \dots, \lambda_m \in [0,1], \exists x_1, \dots, x_m \in S, \sum_{i=1}^n \lambda = 1 \Rightarrow s = \sum_{i=1}^m \lambda_i x_i\}$.

Proposition 1.1.7. *L'enveloppe convexe de S est un ensemble convexe.*

Démonstration. Immédiat puisque il se construit sur les combinaisons convexes. □

Proposition 1.1.8. *L'enveloppe convexe de S est le plus petit ensemble convexe contenant S.*

Démonstration. $\text{co}(S)$ est un ensemble convexe et $S \subset \text{co}(S)$ puisque $\forall s \in S, s = 1 \times s + 0 \times s$. Donc l'intersection W de tous les ensembles convexes contenant S est telle que $W \subset \text{co}(S)$.

Soit C un ensemble convexe contenant S. C contient toute combinaison convexe de ses éléments, en particulier ceux de S puisque $S \subset C$. Donc $\text{co}(S) \subset C$. Ainsi $\forall C$ convexe, $\text{co}(S) \subset \bigcap C = W$. □

Définition 1.1.8 (Polytope)**.** *Un polytope de $\mathbb{R}^n$ est l'enveloppe convexe de m points $x_1, \ldots, x_m$ de $\mathbb{R}^n$, i.e.* $\text{co}(\{x_1, \ldots, x_m\})$.

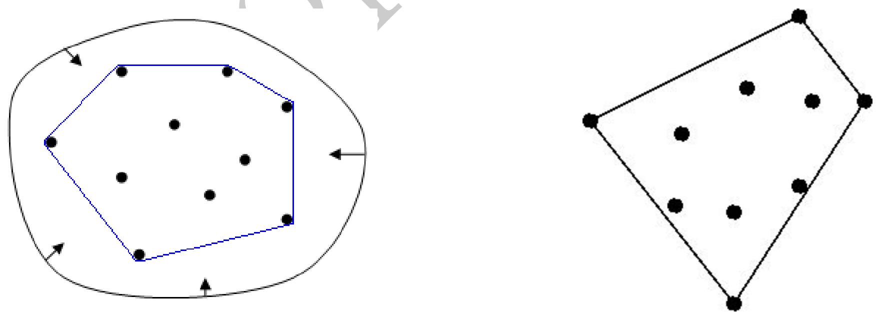
Exemple 1.1.9. *Pour $x_1, x_2 \in \mathbb{R}^2$, l'enveloppe convexe de $S = \{x_1, x_2\}$ est le segment x_1, x_2. Pour $x_1, x_2, x_3 \in \mathbb{R}^2$, l'enveloppe convexe de $S = \{x_1, x_2, x_3\}$ est le triangle x_1, x_2, x_3. Pour $x_1, x_2, x_3, x_4 \in \mathbb{R}^2$, l'enveloppe convexe de $S = \{x_1, x_2, x_3, x_4\}$ est soit un triangle contenant le quatrième point (qui n'est pas un sommet) soit un quadrilatère.*

Exemple 1.1.10. *Les ensembles S et T définis par $S = \{(0,0), (1,0), (0,1)\}$ et $T = \{(0,0), (1,0), (0,1), (0,1/2), (1/2,0), (1/2,1/2)\}$ ont la même enveloppe convexe à savoir le triangle rectangle de sommet* $(0,0)$.

Proposition 1.1.9. *Pour tout ensemble $S \subset T$, on a* $\text{co}(S) \subset \text{co}(T)$.

Démonstration. Soit $x \in \text{co}(S)$. On sait qu'il existe $\lambda_1, \ldots, \lambda_m \in [0,1]$ et $x_1, \ldots, x_m \in S$, tels que $\sum_{i=1}^{n} \lambda = 1 \Rightarrow s = \sum_{i=1}^{m} \lambda_i x_i$. Comme $S \subset T$, alors pour tout $i = 1, \ldots, m$, $x_i \in T$. Donc $\sum_{i=1}^{m} \lambda_i x_i = s \in \text{co}(T)$.

Autre preuve :
Par définition de l'enveloppe convexe, on a $T \subset \text{co}(T), S \subset \text{co}(S)$. En utilisant, $S \subset T$, on a $S \subset \text{co}(T)$. Comme $\text{co}(S)$ est le plus petit ensemble convexe contenant S. Donc $\text{co}(S) \subset \text{co}(T)$. □



(a) Exemple 1 (b) Exemple 2

FIGURE 1.7 – Enveloppes convexes

Définition 1.1.11 (Enveloppe conique)**.** *Pour un ensemble $S \subset \mathbb{R}^n$, l'enveloppe conique de S notée* $\text{cone}(S)$ *est l'ensemble de toute combinaison positive d'éléments de S, i.e.* $\text{cone}(S) = \{s, \exists \lambda_1, \ldots, \lambda_m \geq 0, \exists x_1, \ldots, x_m \in S, s = \sum_{i=1}^{m} \lambda_i x_i\}$.

Proposition 1.1.10. *L'enveloppe conique de S est un cone.*

Démonstration. Immédiat car il se construit sur les combinaisons positives. □

Polycopié de cours

Proposition 1.1.11. *L'enveloppe conique de S est le plus petit cone contenant S.*

Démonstration. $\text{cone}(S)$ est un ensemble conique et $S \subset \text{cone}(S)$ puisque $\forall s \in S, s = 1 \times s$. Donc l'intersection W de tous les ensembles coniques contenant S est telle que $W \subset \text{cone}(S)$.

Soit C un cone contenant S. C contient toute combinaison positive de ses éléments, en particulier ceux de S puisque $S \subset C$. Donc $\text{cone}(S) \subset C$. Ainsi $\forall C$ cone, $\text{cone}(S) \subset \bigcap C = W$.

□

Exemple 1.1.12 (Simplexe standard)**.** *Le simplexe (standard) de $\mathbb{R}^n$ donné par $S_n = \{x \in \mathbb{R}^n, \forall i = 1, \ldots, n, x_i \in [0,1], \sum_{i=1}^n x_i = 1\}$ est un polyèdre. Autrement dit, $S_n = \text{co}(\{e_1, \ldots, e_n\})$ où $(e_1, \ldots, e_n)$ sont des vecteurs unitaires indépendants.*

Exemple 1.1.13 (Simplexe standard et probabilité)**.** *Considérons une variable aléatoire discrète X à valeurs dans $\{x_1, \ldots, x_m\}$ avec les probabilités $p_1, \ldots, p_m$. Le vecteur de probabilités $p = (p_1, \ldots, p_m)$ appartient au simplexe standard S_m. S_m décrit l'ensemble des lois de probabilités discrètes à m valeurs, un point est une loi particulière.*

Exemple 1.1.14 (Simplexe (conique) de m points)**.** *Pour m points de R^n (linéairement indépendants)[*], le simplexe de m points est le polytope de ces m points. Pour m points de R^n (affinement indépendants)[†], le simplexe conique de m points est l'enveloppe conique de ces m points.*

Proposition 1.1.12 (Représentation unique)**.** *Soient $x_1, \ldots, x_m \in \mathbb{R}^n$.*

- *lorsque ces points sont linéairement indépendants, on note $P = \text{co}(\{x_1, \ldots, x_m\})$ le simplexe correspondant. Pour chaque point de P, il existe une unique représentation de x comme combinaison convexe des m points.*
- *lorsque ces points sont affinement indépendants, on note $P = \text{cone}(\{x_1, \ldots, x_m\})$ le simplexe conique correspondant. Pour chaque point de P, il existe une unique représentation de x comme combinaison positive des m points.*

1.1.5 Algorithme de calcul d'enveloppe convexe

On considère un ensemble M de points du plan de cardinal m dont on veut calculer l'enveloppe convexe $\text{co}(M)$. L'algorithme de Graham-Scan comporte deux étapes :

1. trier angulairement les points de M autour d'un point arbitraire $a \in M$. On obtient un cycle $L = \{s_1, \ldots, s_p\}$ de sommets. Si a se trouve sur le bord de l'enveloppe convexe, alors on insère a dans le cycle L.

   NB : $p = m$ ou $m - 1$.
2. tourner autour du cycle L en supprimant les points qui ne peuvent pas être des points extrémaux : notons les points examinés s_{i-1}, s_i, s_{i+1}.

   Pour $i = 1, \ldots, p$, on teste s_i
   - si s_i n'est pas extrêmal par rapport à s_{i-1}, s_{i+1}, alors on supprime s_i de L et on réitère avec $(s_{i-2}, s_{i-1}, s_{i+1})$.
   - sinon on continue avec (s_i, s_{i+1}, s_{i+2}).

   NB : par convention, $(s_0, s_1, s_2) = (s_p, s_1, s_2)$ et $(s_{p-1}, s_p, s_{p+1}) = (s_{p-1}, s_p, s_1)$.

Le première étape a une complexité de $m \log(m)$ tandis que la seconde étape ne possède au plus $2m$ itérations. La complexité totale est $m \log(m)$.

La seconde étape de cet algorithme se base sur la constatation suivante :

*. $x_1, \ldots, x_m$ sont linéairement indépendants ssi $\sum_i \lambda_i x_i = 0$ entraine $\forall i = 1, \ldots, m, \lambda_i = 0$.

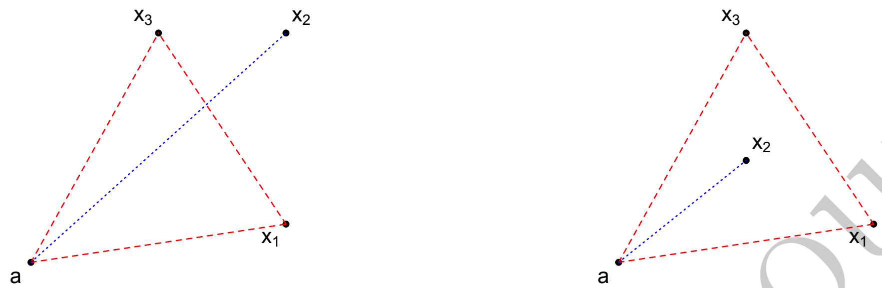
†. $x_1, \ldots, x_m$ sont affinement indépendants ssi $\sum_i \lambda_i x_i = 0$ et $\sum_i \lambda_i = 0$ entraine $\forall i = 1, \ldots, m, \lambda_i = 0$.

Polycopié de cours

— un point x ne peut pas être extrémal s'il est situé dans le triangle dont les sommets sont le point central a et les deux points voisins de x (le précédent $\underline{x}$ et le suivant $\overline{x}$ dans l'ordre trigonométrique autour de a)

— ou autrement dit si le triplet $(\underline{x}, x, \overline{x})$ ne respecte pas l'ordre trigonométrique (voir figure 1.8a).

Réciproquement, si l'orientation de tous les triplets consécutifs du cycle L est l'orientation trigonométrique, alors ce cycle donne le bord de l'enveloppe convexe de M.



(a) Un point x_2 étant extrémal

(b) Un point x_2 ne pouvant pas être extrémal

Un exemple complet d'exécution de l'algorithme est donné en figure 1.9.

1.2 Application à la programmation linéaire

1.2.1 Construction de polyèdres

Traitons le cas d'une seule inégalité $u^T x \leq b$ ou $u^T x \geq b$. La droite frontière a toujours pour vecteur orthogonal u. La zone des points possibles est à l'oppose du vecteur pour $u^T x \leq b$ et du coté du vecteur pour $u^T x \geq b$.

Exemple 1.2.1. *Pour $n = 2$, $A = (1\ 1) \in \mathbb{R}^{1\times 2}$, $b = 3$, on obtient pour les contraintes*

$$\begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} 3 \end{pmatrix} \Leftrightarrow \{\ x_1 + x_2 \leq 3,$$

Il est facile de voir que S est un ensemble convexe. Voir figure 1.10c.

Exemple 1.2.2. *Considérons les contraintes suivantes*

$$\begin{pmatrix} 0 & 1 \\ 4 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} 2.5 \\ 10 \\ -3 \end{pmatrix} \Leftrightarrow \begin{cases} x_2 \leq 2.5, \\ 4x_1 - x_2 \leq 10, \\ x_1 + x_2 \geq 3 \end{cases}$$

Déterminons les frontières de cet ensemble

$$\begin{cases} x_2 = 2.5, \\ 4x_1 - x_2 = 10, \\ x_1 + x_2 = 3 \end{cases} \Leftrightarrow \begin{cases} x_2 = 2.5, \\ x_2 = 4x_1 - 10, \\ x_2 = 3 - x_1 \end{cases}$$

Ce sont trois droites qui se coupent en $\begin{pmatrix} 12.5/4 \\ 2.5 \end{pmatrix}, \begin{pmatrix} 0.5 \\ 2.5 \end{pmatrix}, \begin{pmatrix} 13/5 \\ 2/5 \end{pmatrix}$. *Chaque droite a pour vecteur orthogonal les lignes de la matrice A, voir les vecteurs u, v et w sur la figure 1.10f. Il est facile de voir que S est un ensemble convexe.*

Polycopié de cours

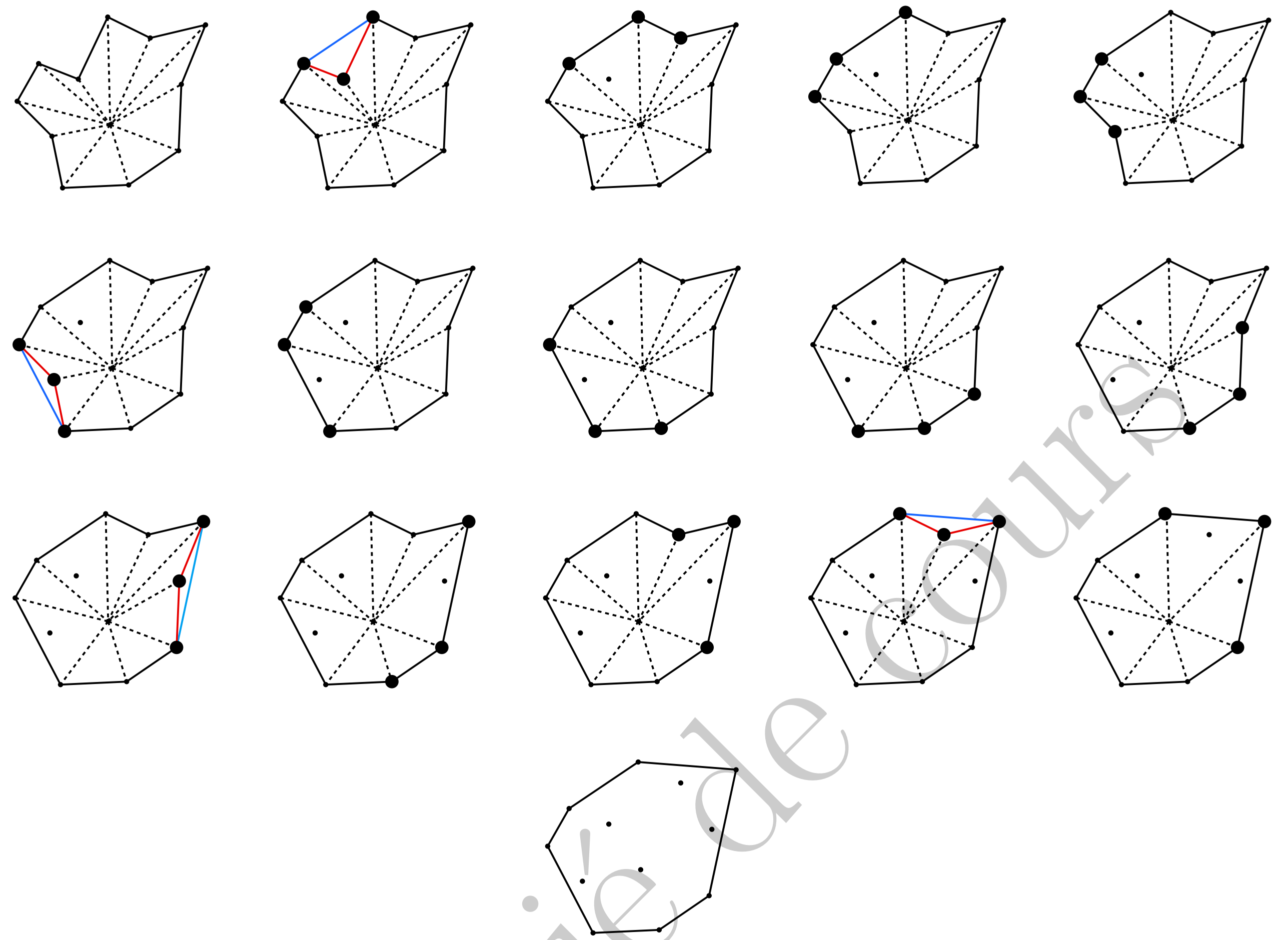


FIGURE 1.9 – Algorithme de Graham-Scan dans le plan

Exemple 1.2.3. *Pour $n = 2$, $A = (1\ 1) \in \mathbb{R}^{1\times 2}$, $b = 3$, on obtient pour les contraintes*

$$(1 \quad 1)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (3), \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \geq (0), \Leftrightarrow \begin{cases} x_1 + x_2 = 3, \\ x_1 \geq 0, \\ x_2 \geq 0 \end{cases} \Leftrightarrow \begin{cases} x_2 = 3 - x_1, \\ x_1 \geq 0, \\ 3 - x_1 \geq 0 \end{cases} \Leftrightarrow S = \{(x_1, 3 - x_1), x_1 \in [0, 3]\}.$$

Il est facile de voir que S est un ensemble convexe. Voir figure 1.10e. Autrement dit, l'ensemble est défini par

$$A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq b, \text{ où } A = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, b = \begin{pmatrix} 3 \\ 3 \end{pmatrix}.$$

Les vecteurs u et v sont donnés par les lignes de la matrice A.

Polycopié de cours

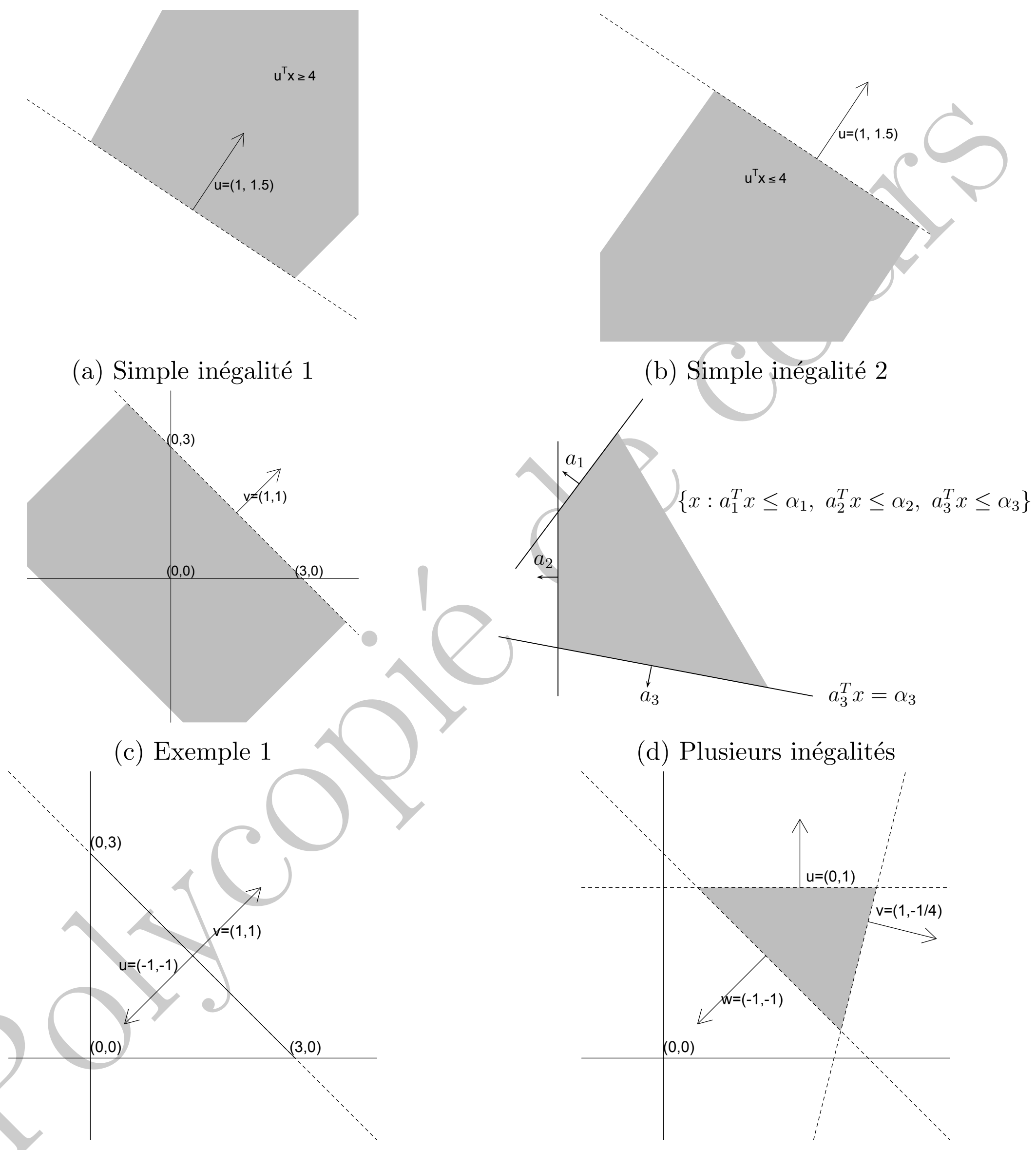


FIGURE 1.10 – polyèdres du plan

Polycopié de cours

1.2.2 Théorème de Caratheodory

Théorème 1.2.1 (Caratheodory pour les enveloppes convexes)**.** *Soit $S \subset \mathbb{R}^n$. Tout point $x \in \mathrm{co}(S)$ peut s'écrire comme une combinaison convexe de m points affinement indépendants dans S tels que $m \leq n+1$, voir figure 1.11.*

Théorème 1.2.2 (Caratheodory pour les cones)**.** *Soit $S \subset \mathbb{R}^n$. Tout point $x \in \mathrm{cone}(S)$ peut s'écrire comme une combinaison positive de m points linéairement indépendants dans S tels que $m \leq n$.*

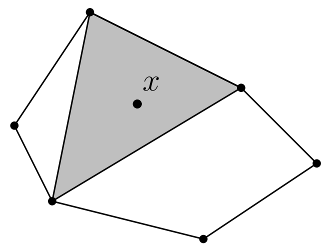


FIGURE 1.11 – Un point x comme une combinaison convexe

1.2.3 Programmes linéaires

Un programme linéaire consiste à résoudre $\max_{x\in P} c^T x$ où P est un polyèdre. Nous allons traiter deux classes de polyèdres.

Exemple 1.2.4. *Soit $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ et $A \in \mathbb{R}^{m\times n}$. Un programme linéaire est un problème d'optimisation du type*

$$\max_{x\in\mathbb{R}^n} c^T x, \text{ t.q. } Ax = b, x \geq 0.$$

Cela revient à maximiser $\sum_{i=1}^n x_i c_i$ sur un ensemble convexe. Remarquons que les contraintes $Ax = b, x \geq 0$ définissent bien un polyèdre

$$\begin{pmatrix} A \\ -A \\ -I \end{pmatrix} x \leq \begin{pmatrix} b \\ -b \\ 0 \end{pmatrix}.$$

Notons $A_{.,1}, \ldots, A_{.,n}$ les colonnes de A. Supposons que $m \leq n$. On a

$$P \neq \emptyset \Leftrightarrow \exists x \in \mathbb{R}^n, Ax = b, \forall x_i \geq 0 \Leftrightarrow \sum_{i=1}^n A_{.,i} x_i = b, \forall x_i \geq 0 \Leftrightarrow b \in \mathrm{cone}(A_{.,1}, \ldots, A_{.,n})$$

où $A_{.,i} \in \mathbb{R}^m$ sont les colonnes de A. Autrement P non vide équivaut à b est une combinaison positive d'éléments de $\mathbb{R}^m$.

Or d'après le théorème de Caratheodory pour les cônes, il existe t colonnes linéairement indépendantes représentant b comme une combinaison positive et $t \leq m \leq n$: $b = \sum_{i=1}^t x_i A^\star_{.,i}$. C'est à dire, il existe $x \in \mathbb{R}^n$ tel que $n-t$ composantes sont nulles et vérifiant $Ax = b$. Les composantes non-nulles correspondent aux indices des colonnes linéairement indépendantes. Si un point x quelconque du polyèdre s'écrit de cette façon, alors un point solution aura aussi $n-t$ composantes nulles et t composantes non nulles.

Exemple 1.2.5. *Soit $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ et $A \in \mathbb{R}^{m\times n}$. Un programme linéaire est un problème d'optimisation du type*

$$\max_{x\in\mathbb{R}^n} c^T x, \text{ t.q. } Ax \leq b.$$

Polycopié de cours

Si le polyèdre n'est pas borné, le programme d'optimisation n'a pas forcément de solutions. S'il est borné, par le théorème de Motskin, le polyèdre peut se réécrire comme un polytope. Le théorème de Caratheodory sur les enveloppes convexes s'applique : il existe t points affinement indépendants pour représenter un point du polytope avec $t \leq n+1$.

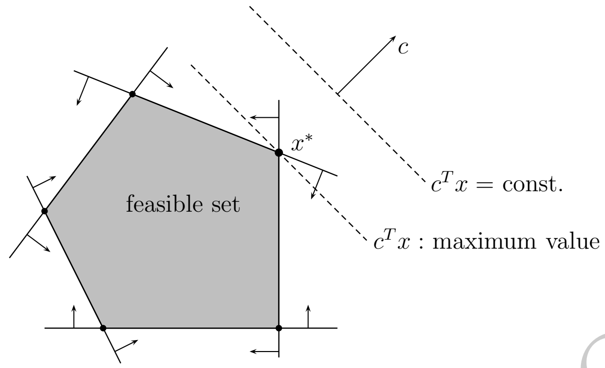


FIGURE 1.12 – Résolution graphique d'un problème linéaire

Proposition 1.2.3. *Pour un polyèdre borné $P \subset \mathbb{R}^n$ dont on note s_j les sommets, le programme linéaire $\max_{x \in P} c^T x$ a pour solution l'enveloppe convexe $\mathrm{co}(\{s_j, j \in J\})$, où l'ensemble J désigne les indices réalisant le maximum $J = \{j \in \mathbb{N}, c^T s_j = v\}$ et $v = \max_{j \in \mathbb{N}} c^T s_j$, voir figure 1.12.*

Démonstration. Si le polyèdre P est borné, on peut montrer par le théorème général des polyèdres de Motzkin que P peut se réécrire comme un polytope de $(s_j)_j$ les sommets. Par le théorème de Caratheodory, seul au plus t points $p_1, \dots, p_t$ de P permettent de représenter un élément du polyèdre avec $t \leq n+1$. Considérons donc le programme linéaire

$$\max_{x \in P} c^T x, \text{t.q. } P = \mathrm{co}(\{p_1, \dots, p_t\}) \subset \mathbb{R}^n.$$

Comme P est un polytope, n'importe quel point $x \in P$ s'écrit comme une combinaison convexe des éléments p_i

$$x = \sum_{i=1}^{t} \lambda_i p_i, \text{ où } \sum_{i=1}^{t} \lambda_i = 1, \lambda_i \geq 0.$$

Notons que la décomposition n'est pas forcément unique, voir figure 1.11. Donc pour tout $x \in P$, on a

$$c^T x = c^T \left(\sum_{i=1}^{t} \lambda_i p_i \right) = \sum_{i=1}^{t} \lambda_i c^T p_i \leq \sum_{i=1}^{t} \lambda_i \max_{i=1,\dots,t} c^T p_i = \max_{i=1,\dots,t} c^T p_i \sum_{i=1}^{t} \lambda_i = \max_{i=1,\dots,t} c^T p_i = v.$$

On a donc une borne pour l'ensemble des solutions optimales. La borne est atteinte pour l'ensemble J des j tels que $c^T p_j = v$, i.e. pour des x du type $\sum_{j \in J} \lambda_j p_j$ avec $\lambda_j > 0$. Autrement dit, les solutions appartiennent à $\mathrm{co}(\{p_j, j \in J\})$, voir figure 1.12. La solution est atteinte sur un des points extrêmaux. □

Cette méthode trouve son intérêt si P est un polyèdre fermé s'exprimant comme un polytope, où le nombre de points extrêmaux est faible. Cette méthode graphique n'est pas adaptée pour la résolution générique des programmes linéaires : la méthode du simplexe est l'alternative fiable.

Polycopié de cours

1.3 Théorème de séparation

1.3.1 Quelques définitions

Définition 1.3.1 (Hyperplan)**.** *L'ensemble $H \subset \mathbb{R}^n$ est un hyperplan de l'espace vectoriel $(\mathbb{R}^n, +, .)$ s'il est de la forme $H = \{x, a^T x = \alpha\}$ pour $\alpha \in \mathbb{R}$ et $a \neq 0 \in \mathbb{R}^n$. Le vecteur a est appelé le vecteur normal de H. L'hyperplan sera noté $H_{a,\alpha}$.*

Définition 1.3.2 (Demi-espace)**.** *Par construction, un hyperplan $H_{a,\alpha}$ sépare l'espace en deux. On définit les demi-espaces suivants*

$$H^+_{a,\alpha} = \{x, a^T x \geq \alpha\}, H^-_{a,\alpha} = \{x, a^T x \leq \alpha\}.$$

Définition 1.3.3 (Hyperplan supportant)**.** *Un hyperplan H supporte un ensemble $S \subset \mathbb{R}^n$ si on a*
- *— $H^+ \subset S$ ou $H^- \subset S$,*
- *— $H \cap S \neq \emptyset$.*

De plus, H est dit supporté S à x lorsque $x \in H \cap S$.

Définition 1.3.4 (Hyperplan séparateur)**.** *Un hyperplan H sépare deux ensembles $S, T \subset \mathbb{R}^n$ lorsque $S \subset H^-$ et $T \subset H^+$ ou vice versa. On dit aussi que S et T sont proprement séparés par H. Voir figure 1.13.*

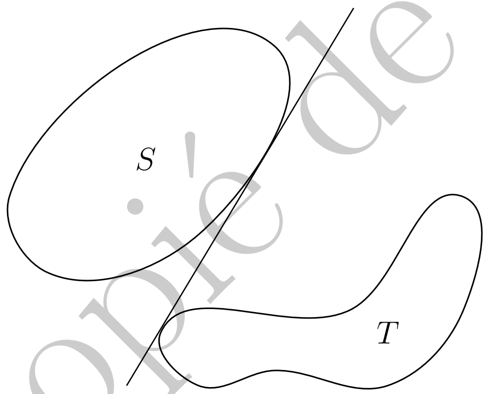


Figure 1.13 – Un hyperplan séparateur

Exemple 1.3.5. *Dans $\mathbb{R}$, $S =]-\infty, 1]$ et $T = [1, +\infty[$ sont séparés par $H_{(1),1}$.*

Dans $\mathbb{R}^2$,
- *— $S = \{x \in \mathbb{R}^2, x_1 \geq -1, x_2 = 0\}$ et $T = [0,1] \times [3,5]$ sont séparés par $H_{(-1,1),1}$.*
- *— $S = \{x \in \mathbb{R}^2, x \leq 1\}$ et $T = \{x \in \mathbb{R}^2, x \geq 1\}$ sont séparés par $H_{(1,0),0}$.*
- *— $S = B_{(1,1),1}$ et $T = B_{(-1,1),1/2}$ sont séparés par $H_{(1,0),0}$.*

Définition 1.3.6 (Séparation forte)**.** *Deux ensembles $S, T \subset \mathbb{R}^n$ sont fortement séparés lorsqu'il existe deux hyperplans parallèles distincts séparants S et T. C'est à dire il existe $a \in \mathbb{R}^n$ tel que pour $\alpha \neq \beta$, $H_{a,\alpha}$ et $H_{a,\beta}$ sont des hyperplans séparateurs.*

Proposition 1.3.1 (Caractérisation des hyperplans séparateurs)**.** *Deux ensembles $S, T \subset \mathbb{R}^n$ sont proprement séparés lorsque*

$$\exists a \in \mathbb{R}^n, \sup_{x \in S} a^T x \leq \inf_{y \in T} a^T y \text{ et } \inf_{x \in S} a^T x < \sup_{y \in T} a^T y.$$

Deux ensembles $S, T \subset \mathbb{R}^n$ sont fortement séparés lorsque

$$\exists a \in \mathbb{R}^n, \sup_{x \in S} a^T x < \inf_{y \in T} a^T y.$$

Polycopié de cours

1.3.2 Théorème de séparation

Définition 1.3.7 (intérieur relatif)**.** *L'intérieur relatif d'un ensemble $X \subset \mathbb{R}^n$ est l'ensemble des points x tels que*

$$\exists r > 0, B(x,r) \cap \mathit{Aff}(X) \subset X,$$

où $\mathit{Aff}(X) = \{\sum_{i=1}^m \lambda_i x_i, \forall i = 1, \ldots, m, x_i \in X, \sum_{i=1}^m \lambda_i = 1\}$ et $B(x,r)$ est la boule centrée en x et de rayon r. Notons que $ri(X) \subset X \subset cl(X) \subset \mathit{Aff}(X)$.

Théorème 1.3.2 (Théorème de séparation)**.** *Deux ensembles convexes non-vides $S, T \subset \mathbb{R}^n$ peuvent proprement séparés ssi leur intérieurs relatifs sont disjoints, i.e.*

$$ri(S) \cap ri(T) = \emptyset.$$

Polycopié de cours

Chapitre 2

Fonctions convexes univariées (4,5 séances)

Polycopié de cours

2.1 Fonctions convexes

Définition 2.1.1 (Fonction convexe)**.** *Une fonction $f : \mathbb{R} \mapsto \mathbb{R}$ est convexe si pour tout $\lambda \in]0,1[$ et pour tout $x, y \in \mathbb{R}$,*

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

Si l'inégalité est stricte, on parle de fonction strictement convexe.

Exemple 2.1.2 (Application linéaire)**.** *Soit $f(x) = ax$ pour $a \in \mathbb{R}$. Il est facile de voir*

$$f(\lambda x + (1-\lambda)y) = \lambda ax + (1-\lambda)ay = \lambda f(x) + (1-\lambda)f(y)$$

pour tout $x, y \in \mathbb{R}$ et pour tout $\lambda \in]0,1[$

Exemple 2.1.3 (Application constante puis affine)**.** *Soit $f(x) = a(x-b)1\!\!1_{x \geq b}$ pour $a \geq 0$ et $b \in \mathbb{R}$. Pour tout $x \leq y \in \mathbb{R}$ et pour tout $\lambda \in]0,1[$, on a*

$$f(\lambda x + (1-\lambda)y) = a(\lambda x + (1-\lambda)y - b)1\!\!1_{\lambda x + (1-\lambda)y \ \geq b}.$$

- *Si $x \leq y \leq b$, alors $f(x) = f(y) = 0 = f(\lambda x + (1-\lambda)y)$.*
- *Si $x \leq b < y$, alors $f(x) = 0$, $f(y) = a(y-b)$ et $f(\lambda x + (1-\lambda)y) = 0$ ou $a(\lambda x + (1-\lambda)y - b)$ suivant si $\lambda x + (1-\lambda)y \leq b$ ou $\lambda x + (1-\lambda)y > b$. Dans le premier cas, l'inégalité est vérifiée. Dans le second cas,*

$$a\lambda x + a(1-\lambda)y - ab = a\lambda \underbrace{(x-b)}_{\leq 0} + a(1-\lambda)(y-b) \leq (1-\lambda)a(y-b)$$

- *si $b < x \leq y$, alors on retrouve le cas linéaire $f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y)$.*

Exemple 2.1.4 (Application quadratique)**.** *Considérons $f(x) = x^2$. Pour tout $x \leq y \in \mathbb{R}$ et pour tout $\lambda \in]0,1[$, on a*

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= (\lambda x + (1-\lambda)y)^2 = \lambda^2 x^2 + (1-\lambda)^2 y^2 + 2\lambda(1-\lambda)xy \\ &= \lambda x^2 + (1-\lambda)y^2 + x^2(\lambda^2 - \lambda) + y^2(\lambda^2 - \lambda) + 2\lambda(1-\lambda)xy \\ &= \lambda f(x) + (1-\lambda)f(y) + \lambda \underbrace{(\lambda - 1)}_{\leq 0}(x-y)^2 \leq \lambda f(x) + (1-\lambda)f(y) \end{aligned}$$

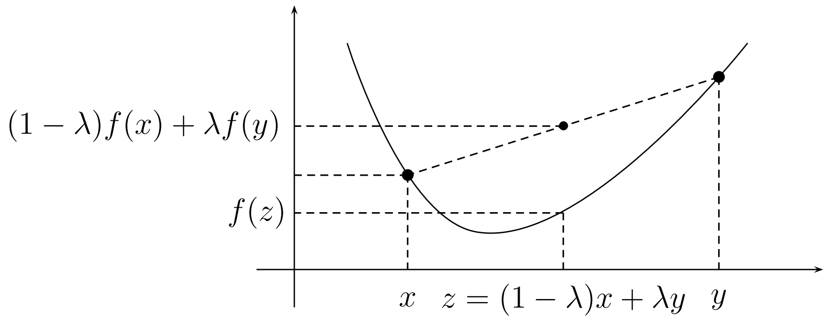


FIGURE 2.1 – Point de vue géométrique de la propriété de convexité

Proposition 2.1.1. *Soient $f : \mathbb{R} \mapsto \mathbb{R}$ et $x_1 < x_2 < x_3$. Notons les points du plan $P_i = (x_i, f(x_i))$ pour $i = 1, 2, 3$. La pente entre P_i et P_j est définie par*

$$D(i,j) = \frac{f(x_i) - f(x_j)}{x_i - x_j}.$$

L'inégalité des trois pentes est

$$P_2 \text{ est en dessous de la ligne } [P_1, P_3] \Leftrightarrow D(1,2) \leq D(1,3) \Leftrightarrow D(1,3) \leq D(2,3).$$

Démonstration. Le graph de la fonction f définis par les points $(x, f(x))_{x\in\mathbb{R}}$ révèle des propriétés de celle-ci. Montrons que

$$P_2 \text{ en dessous ligne } [P_1, P_3] \Leftrightarrow D(1,2) \leq D(1,3) \Leftrightarrow D(1,3) \leq D(2,3).$$

(1) $\Rightarrow$ (2) On a pour tout $\lambda \in [0,1]$,

$$\begin{aligned} f(x_2) &\leq \lambda f(x_1) + (1-\lambda)f(x_3) \Leftrightarrow f(x_2) - f(x_1) \leq (1-\lambda)(f(x_3) - f(x_1)) \\ &\Leftrightarrow \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq (1-\lambda)\frac{f(x_3) - f(x_1)}{x_2 - x_1} \end{aligned}$$

Comme $0 \leq 1 - \lambda \leq 1$, $x_2 - x_1 < x_3 - x_1$, on choisit λ tel que

$$\frac{x_2 - x_1}{1 - \lambda} = x_3 - x_1$$

Ainsi $D(1,2) \leq D(1,3)$.

(2) $\Rightarrow$ (3) On a

$$\begin{aligned} \frac{f(x_2) - f(x_1)}{x_2 - x_1} &\leq \frac{f(x_3) - f(x_1)}{x_3 - x_1} \\ \Leftrightarrow (f(x_2) - f(x_1))(x_3 - x_1) &\leq (f(x_3) - f(x_1))(x_2 - x_1) \\ \Leftrightarrow (f(x_2) - f(x_1))(x_3 - x_1) &\leq (f(x_3) - f(x_1))(x_2 - x_3 + x_3 - x_1) \\ \Leftrightarrow (f(x_2) - f(x_3))(x_3 - x_1) &\leq (f(x_3) - f(x_1))(x_2 - x_3) \\ \Leftrightarrow (f(x_2) - f(x_3))(x_3 - x_1) &\leq -(f(x_3) - f(x_1))(x_3 - x_2) \\ \Leftrightarrow \frac{f(x_2) - f(x_3)}{x_3 - x_2} &\leq -\frac{f(x_3) - f(x_1)}{x_3 - x_1} \\ \Leftrightarrow \frac{f(x_3) - f(x_2)}{x_3 - x_2} &\geq \frac{f(x_3) - f(x_1)}{x_3 - x_1} \end{aligned}$$

Donc $D(1,3) \leq D(2,3)$.

Polycopié de cours

(3) $\Rightarrow$ (1) On a

$$\begin{aligned}
\frac{f(x_3)-f(x_1)}{x_3-x_1} \quad \leq \quad & \frac{f(x_3)-f(x_2)}{x_3-x_2} \Leftrightarrow (f(x_3)-f(x_1))(x_3-x_2) \leq (f(x_3)-f(x_2))(x_3-x_1) \\
\Leftrightarrow \quad & f(x_2)(x_3-x_1) \leq f(x_3)(x_2-x_1)+f(x_1)(x_3-x_2) \\
\Leftrightarrow \quad & f(x_2) \leq f(x_3)\underbrace{\frac{x_2-x_1}{x_3-x_1}}_{1-p(x_2)}+f(x_1)\underbrace{\frac{x_3-x_2}{x_3-x_1}}_{p(x_2)} \\
\Leftrightarrow \quad & f(x_2) \leq p(x_2)f(x_1)+(1-p(x_2))f(x_3)
\end{aligned}$$

Autrement pour tout x_2, $D(1,3) \leq D(2,3)$ équivaut à $f(x_2) \leq p(x_2)f(x_1)+(1-p(x_2))f(x_3)$ où $p(x_2) \in [0,1]$.

□

Proposition 2.1.2. *$f : \mathbb{R} \mapsto \mathbb{R}$ est convexe ssi pour tout $x_0 \in \mathbb{R}$, le coefficient de pente*

$$D : x \mapsto \frac{f(x)-f(x_0)}{x-x_0}$$

est croissante sur $\mathbb{R} \setminus \{x_0\}$. En particulier, on a l'inégalité des trois pentes

$$a < b < c \Rightarrow \frac{f(a)-f(b)}{a-b} \leq \frac{f(a)-f(c)}{a-c} \leq \frac{f(b)-f(c)}{b-c}.$$

Démonstration. Par la remarque précédente, pour tout x_2, P_2 est en dessous de la ligne $[P_1, P_3]$ ($x_1 < x_2 < x_3$) équivaut à $D(1,3) \leq D(2,3)$ pour tout $x_1 < x_2 < x_3$. On prend $x_0 = x_3$. □

Définition 2.1.5 (Fonction convexe sur I)**.** *Pour un intervalle I, une fonction $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ est convexe si pour tout $\lambda \in]0,1[$ et pour tout $x, y \in I$,*

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

Si l'inégalité est stricte, on parle de fonction strictement convexe.

2.2 Inégalités pour les fonctions convexes

2.2.1 Inégalités

Théorème 2.2.1 (Inégalités de Jensen)**.** *Soit $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ une fonction convexe. Pour tout $x_1 < \cdots < x_n$ et $\lambda_1, \ldots, \lambda_n \in [0,1]$ tel que $\sum_{i=1}^n \lambda_i = 1$, alors*

$$f\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i f(x_i).$$

Si f est concave, alors l'inégalité est inversée.

Démonstration. Similaire à la proposition 1.1.6. □

Polycopié de cours

2.2.2 Application en probabilité et statistique

Exemple 2.2.1. *Si les x_i représentent les valeurs possibles d'une variable aléatoire discrète X avec les probabilités λ_i, alors on a l'inégalité suivante sur l'espérance*

$$f(E(X)) \leq E(f(X)).$$

moyennes arithmétique, géométrique et harmonique

Exemple 2.2.2 (Moyennes)**.** *Soit $x_1, \ldots, x_n$ un ensemble de n observations positives. On définit les moyennes arithmétiques, géométriques et harmoniques par*

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad g_{x,n} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}, \quad h_{x,n} = \frac{n}{\sum\limits_{i=1}^{n} \frac{1}{x_i}}.$$

En utilisant l'inégalité de Jensen, on obtient $g_{x,n} \leq \bar{x}_n$ et $h_{x,n} \leq \bar{x}_n$. En effet,

$$f(x) = -\log(x), \forall i = 1, \ldots, n, \lambda_i = 1/n.$$

Comme f est convexe,

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i) \Leftrightarrow -\log\left(\sum_{i=1}^{n} \frac{1}{n} x_i\right) \leq \sum_{i=1}^{n} (-\frac{1}{n}) \log(x_i) \Leftrightarrow \log\left(\sum_{i=1}^{n} \frac{1}{n} x_i\right) \geq \frac{1}{n}\sum_{i=1}^{n} \log(x_i)$$

En prenant l'exponentielle de chaque coté, on a

$$\sum_{i=1}^{n} \frac{1}{n} x_i \geq \exp(\frac{1}{n}\sum_{i=1}^{n} \log(x_i)) \Leftrightarrow \bar{x}_n \geq \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

Choisissons $g(x) = 1/x$ qui est aussi convexe.

$$g\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i g(x_i) \Leftrightarrow 1/\left(\sum_{i=1}^{n} \frac{1}{n} x_i\right) \leq \sum_{i=1}^{n} (\frac{1}{n}) 1/(x_i) \Leftrightarrow 1/\left(\sum_{i=1}^{n} \frac{1}{n} x_i\right) \leq \frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}$$

En prenant l'inverse, on obtient

$$\bar{x}_n \geq \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}} = h_{x,n}.$$

2.3 Lien avec les ensembles convexes

Définition 2.3.1 (Epigraphe)**.** *Pour une fonction $f : I \subset \mathbb{R} \mapsto \mathbb{R}$, l'épigraphe se définit comme* $\text{epi}(f) = \{(x, y) \in I \times \mathbb{R}, y \geq f(x)\}$.

Théorème 2.3.1. *Pour une fonction $f : I \subset \mathbb{R} \mapsto \mathbb{R}$, f est convexe $\Leftrightarrow$ $\text{epi}(f)$ est un ensemble convexe de $\mathbb{R}^2$.*

Démonstration. $\Rightarrow$ Soient $z_1 = (x_1, y_1), z_2 = (x_2, y_2) \in \text{epi}(f)$ et $\lambda \in [0, 1]$. Par définition,

$$y_1 \geq f(x_1), y_2 \geq f(x_2) \quad \Rightarrow \quad \lambda y_1 \geq \lambda f(x_1), (1-\lambda) y_2 \geq (1-\lambda) f(x_2).$$

Ainsi $\lambda y_1 + (1-\lambda) y_2 \geq \lambda f(x_1) + (1-\lambda) f(x_2)$. Or par convexité de f $\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(\lambda x_1 + (1-\lambda) x_2)$. Donc

$$\lambda y_1 + (1-\lambda) y_2 \geq f(\lambda x_1 + (1-\lambda) x_2),$$

i.e. $\lambda z_1 + (1-\lambda) z_2 \in \text{epi}(f)$.

Polycopié de cours

$\Leftarrow$ Soient $x_1, x_2 \in I$ et $\lambda \in [0,1]$. On pose $x_\lambda = \lambda x_1 + (1-\lambda)x_2$. Il est naturel que $(x_i, f(x_i)) \in \text{epi}(f)$. Par convexité de l'épigraphe,

$$\lambda(x_1, f(x_1)) + (1-\lambda)(x_2, f(x_2)) \in \text{epi}(f),$$

i.e. $(x_\lambda, \lambda f(x_1) + (1-\lambda)f(x_2)) \in \text{epi}(f)$. Donc $f(x_\lambda) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$.

$\square$

2.4 Cas des fonctions dérivables

2.4.1 Rappel sur la dérivée

Définition 2.4.1 (Dérivée à gauche/droite)**.** *Une fonction $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ est dérivable à gauche en x_0 lorsque la limite existe*

$$\lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0} = f'_-(x_0).$$

f est dérivable à droite en x_0 lorsque la limite existe

$$\lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0} = f'_+(x_0).$$

Si la limite existe en tout point $x_0 \in I$, alors f est dérivable à gauche ou à droite sur I.

Définition 2.4.2 (Dérivable)**.** *Une fonction $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ est dérivable en x_0 lorsque $f'_-(x_0)$ et $f'_+(x_0)$ existent et $f'_-(x_0) = f'_+(x_0) = f'(x_0)$. Si la limite existe en tout point $x_0 \in I$, alors f est dérivable sur I.*

Exemple 2.4.3. *Les polynômes, les fonctions trigonométrique, les fonctions logarithme et exponentielle sont des fonctions dérivables.*

Définition 2.4.4 (C^k)**.** *Une fonction $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ est de classe C^k si la fonction f est k dérivable sur I et $f^{(k)}$ est continue sur I.*

2.4.2 Caractérisation

Proposition 2.4.1 (Caractérisation)**.** *Soit $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ une fonction convexe.*
- *Si f est différentiable sur I, alors f convexe sur $I \Leftrightarrow f'$ est croissante sur I.*
- *Si f est deux fois différentiable sur I, alors f convexe sur $I \Leftrightarrow f''$ est positif ou nul sur I.*

Démonstration. Considérons f une fonction convexe et différentiable sur I. Il est évident que f convexe $\Rightarrow f'$ croissant. Puisque par la propriété 2.1.2, pour tout $x \leq y$ et $x_0 \in \mathbb{R}$,

$$\frac{f(x) - f(x_0)}{x - x_0} \leq \frac{f(y) - f(x_0)}{y - x_0}.$$

En prenant la limite lorsque $x_0 \to x$, on obtient $f'(x) \leq \frac{f(y)-f(x)}{y-x}$ et lorsque $x_0 \to y$, on obtient $\frac{f(y)-f(x)}{y-x} \leq f'(y)$. D'où $f'(x) \leq \frac{f(y)-f(x)}{y-x} \leq f'(y)$.

Montrons que f' croissant $\Rightarrow f$ convexe. Soit $x < y < z$. Par le théorème des accroissements finis, il existe $y_0 \in]x, y[$ tel que

$$f'(y_0) = \frac{f(x) - f(y)}{x - y}.$$

Polycopié de cours

De même il existe $z_0 \in]y, z[$ tel que
$$f'(z_0) = \frac{f(y) - f(z)}{y - z}.$$
Comme f' est croissante et $y_0 < y < z_0$, on en déduit $f'(y_0) \leq f'(z_0)$ équivaut à
$$\frac{f(x) - f(y)}{x - y} \leq \frac{f(y) - f(z)}{y - z} = \frac{f(z) - f(y)}{z - y}.$$
On conclut par la propriété 2.1.2. Pour f deux fois différentiable c'est immédiat. $\square$

Exemple 2.4.5. *Considérons pour $a \geq 0$ et $b \in \mathbb{R}$ la fonction*
$$f(x) = a(x - b)^2 1\!\!1_{[b,+\infty[}(x).$$
Vérifions la décidabilité en b.
$$\lim_{x \to b^-} \frac{f(x) - f(b)}{x - b} = \lim_{x \to b^-} \frac{0}{x - b} = 0$$
$$\lim_{x \to b^+} \frac{f(x) - f(b)}{x - b} \lim_{x \to b^+} \frac{a(x - b)^2}{x - b} \lim_{x \to b^+} a(x - b) = 0.$$
Donc $f'_-(b) = f'_+(b) = 0$. La fonction f est bien dérivable en b. Autrement dit, f est dérivable sur $\mathbb{R}$. On a
$$f'(x) = 2(x - b) 1\!\!1_{[b,+\infty[}(x).$$
Comme c'est une fonction croissante sur $\mathbb{R}$, alors f est convexe sur $\mathbb{R}$.

Exemple 2.4.6. *Montrer que les fonctions définies sur $\mathbb{R}_+$ sont convexes : $f(x) = x^2$, $f(x) = |x|$, $f(x) = x^p$ pour $p \leq 1$, $f(x) = e^x$ et $f(x) = -\log(x)$.*

Proposition 2.4.2 (Dérivabilité)**.** *Soit $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ une fonction convexe. f possède une dérivée à gauche et à droit en tout point de l'intérieur de I. De plus pour $x < y \in I$,*
$$f'_-(x) \leq f'_+(x) \leq \frac{f(y) - f(x)}{y - x} \leq f'_-(y) \leq f'_+(y)$$
En particulier f'_- et f'_+ sont croissantes.

Démonstration. Soient z un point intérieur de I et $x < z < y$. Par la propriété 2.1.2, la fonction $x \mapsto \frac{f(x)-f(z)}{x-z}$ est croissante et borné par $\frac{f(y)-f(z)}{y-z}$. Ceci entraine $f'_-(z)$ existe et telle que
$$f'_-(z) \leq \frac{f(y) - f(z)}{y - z}.$$
En faisant tendre $y \to z^+$, alors on déduit $f'_-(z) \leq f'_+(z)$. Ainsi
$$f'_-(x) \leq f'_+(x), \quad f'_-(y) \leq f'_+(y).$$
Par l'inégalité des pentes,
$$x < z < y \Rightarrow \frac{f(x) - f(z)}{x - z} \leq \frac{f(x) - f(y)}{x - y} \leq \frac{f(z) - f(y)}{z - y}.$$
Combiné avec
$$\lim_{z \to x^+} \frac{f(z) - f(x)}{z - x} = f'_+(x), \quad \lim_{z \to y^-} \frac{f(z) - f(y)}{z - y} = f'_-(y)$$
donne le résultat. $\square$

Polycopié de cours

Proposition 2.4.3 (Continuité)**.** *Soit $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ une fonction convexe. Pour tout $a < b \in I$, il existe $M > 0$ tel que*

$$\forall x, y \in [a,b], |f(x) - f(y)| \leq M|x - y|.$$

Autrement dit f est localement Lipschitz continue (donc continue) sur l'intérieur de I.

Démonstration. Soient $a < b \in I$ et $x, y \in [a,b]$. Pour $x < y$, on a $a < x < y < b$

$$f'_+(a) \leq \frac{f(y) - f(x)}{y - x} \leq f'_-(b) \Rightarrow \left|\frac{f(y) - f(x)}{y - x}\right| \leq \max(f'_-(b), -f'_+(a)) = M < \infty$$

M est fini car les dérivées à gauche et à droite existent. La continuité est immédiate. □

Proposition 2.4.4 (Ensemble de non-dérivabilité)**.** *Soit $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ une fonction convexe. Notons $Z \subset I$ l'ensemble des points où f n'est pas dérivable. Z est un ensemble dénombrable.*

Démonstration. Par la propriété 2.4.2, on a pour tout $z \in I$, $f'_-(z) \leq f'_+(z)$. En particulier pour tout $z \in Z$ pour lequel on a $f'_-(z) < f'_+(z)$ puisque non différentiable. On définit une fonction r par

$$\begin{array}{ll} Z \mapsto & \mathbb{Q} \\ z \mapsto & r(z) \text{ t.q. } f'_-(z) < r(z) < f'_+(z). \end{array}$$

Pour $z < z'$, on a par 2.4.2

$$f'_-(z) < r(z) < f'_+(z) \leq \frac{f(z) - f(z')}{z - z'} \leq f'_-(z') < r(z') < f'_+(z') \Rightarrow r(z) < r(z').$$

Comme f est une fonction strictement croissante sur Z, elle est injective avec $\mathbb{Q}$. $\mathbb{Q}$ étant dénombrable alors Z est au plus dénombrable. C'est équivalent à regarder les points de discontinuités de la fonction croissante f'_-. □

Exemple 2.4.7 (Exemple tordu)**.** *Considérons une fonction $\varphi : \mathbb{N}^\star \mapsto \mathbb{Q} \cap]0,1[$ bijective. Définissons*

$$s : \begin{array}{ccc}]0,1[\times \mathbb{N}^\star & \mapsto & \mathbb{R} \\ (x,n) & & s(x,n) \end{array}, \text{ où } s(x,n) = \begin{cases} 1/2^n & si\ \varphi(n) < x \\ 0 & si\ \varphi(n) \geq x \end{cases}.$$

On définit $f :]0,1[\mapsto]0,1[$ par

$$f(x) = \sum_{n=1}^{+\infty} s(x,n).$$

Donc f est strictement croissante et discontinue en tout point rationnel. Enfin la fonction convexe est définie par intégration $F :]0,1[\mapsto]0,1[$ par $F(x) = \int_0^x f(t)dt$. Par construction elle est discontinue et convexe par croissance de sa fonction "dérivée".

Proposition 2.4.5 (Hypothèse affaiblie)**.** *Soit $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ une fonction continue. Si f vérifie*

$$\forall x, y \in \mathbb{R}, f\left(\frac{x+y}{2}\right) \leq \frac{f(x) + f(y)}{2},$$

alors f est convexe.

Démonstration. Montrons par récurrence que pour $n \geq 1$, pour tout $x, y \in \mathbb{R}$ et $p \in \{0, 1, 2, \ldots, 2^p\}$, on a

$$f\left(\frac{p}{2^n}x + \frac{2^n - p}{2^n}y\right) \leq \frac{p}{2^n}f(x) + \frac{2^n - p}{2^n}f(y).$$

Polycopié de cours

— pour $n = 1$, la relation $f\left(\frac{p}{2}x + \frac{2-p}{2}y\right) \leq \frac{p}{2}f(x) + \frac{2-p}{2}f(y)$ découle de l'hypothèse ($p = 1$) ; $f(x) \leq f(x)$ pour $p = 0$; $f(y) \leq f(y)$ pour $p = 2$.

— supposons la relation au rang n. Soit $p \leq 2^n$. On décompose le paquet

$$\begin{aligned} f\left(\frac{p}{2^{n+1}}x + \frac{2^{n+1}-p}{2^{n+1}}y\right) &= f\left(\frac{1}{2}\left(\frac{p}{2^n}x + \frac{2^n-p}{2^n}y\right) + \frac{1}{2}\frac{2^n y}{2^n}\right) \\ &\leq \frac{1}{2}f\left(\frac{p}{2^n}x + \frac{2^n-p}{2^n}y\right) + \frac{1}{2}f(y) \\ &\leq \frac{1}{2}\left(\frac{p}{2^n}f(x) + \frac{2^n-p}{2^n}f(y)\right) + \frac{1}{2}f(y) \\ &\leq \frac{p}{2^{n+1}}f(x) + \frac{2^{n+1}-p}{2^{n+1}}f(y). \end{aligned}$$

Maintenant si $\lambda \in [0,1]$, il existe une suite $(\lambda_n)_n$ de la forme $p/2^n$ qui converge vers λ par densité de $\mathbb{Q}$ dans $\mathbb{R}$. En passant à la limite l'inégalité (car f est continue)

$$f\left(\frac{p}{2^n}x + \frac{2^n-p}{2^n}y\right) \leq \frac{p}{2^n}f(x) + \frac{2^n-p}{2^n}f(y),$$

on obtient la définition de la convexité

□

Proposition 2.4.6 (Caractérisation)**.** *Soit $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ une fonction différentiable. On a les équivalences suivantes*

1. *f est convexe sur I.*
2. *$\forall x, y \in I, f(x) \geq f(y) + f'(y)(x-y)$.*
3. *$\forall x, y \in I, (f'(x) - f'(y))(x-y) \geq 0$.*

Démonstration. (1) ⇒ (2) Soient $x, y \in I$ et $\lambda \in [0,1]$. On a

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

Donc

$$\begin{aligned} &\frac{f(y+\lambda(x-y)) - f(y)}{\lambda} \leq f(x) - f(y) \\ &\Rightarrow \frac{f(y+\lambda(x-y)) - f(y)}{\lambda} - f'(y)(x-y) \leq f(x) - f(y) - f'(y)(x-y) \\ &\Rightarrow \frac{f(y+\delta) - f(y)}{\delta/(x-y)} - f'(y)(x-y) \leq f(x) - f(y) - f'(y)(x-y) \\ &\Rightarrow \lim_{\delta \to 0} \frac{f(y+\delta) - f(y)}{\delta}(x-y) - f'(y)(x-y) = 0 \leq f(x) - f(y) - f'(y)(x-y) \end{aligned}$$

(2) ⇒ (3) On a donc $x, y \in I$

$$f(x) \geq f(y) + f'(y)(x-y) \text{ et } f(y) \geq f(x) + f'(x)(y-x),$$

par symétrie. En les additionant, on trouve

$$\begin{aligned} f(y) + f(x) &\geq f(x) + f(y) + f'(x)(y-x) + f'(y)(x-y) \\ \Leftrightarrow 0 &\geq (f'(y) - f'(x))(x-y) \\ \Leftrightarrow (f'(x) - f'(y))(x-y) &\geq 0 \end{aligned}$$

Polycopié de cours

(3) $\Rightarrow$ (1) Soient $x, y \in I$. Sans perte de généralités, $x \leq y$. On pose $g(t) = f(x + t(y - x))$ pour $t \in [0, 1]$. Soient $t_1 \leq t_2 \in [0, 1]$. On a

$$g'(t_2) - g'(t_1) = \left(f'(x + t_2(y - x)) - f'(x + t_1(y - x))\right)(y - x)$$

En posant $z_i = x + t_i(y - x)$, on a $z_1 - z_2 = (t_1 - t_2)(y - x)$ et

$$\begin{aligned}(t_2 - t_1)(g'(t_2) - g'(t_1)) &= \left(f'(x + t_2(y - x)) - f'(x + t_1(y - x))\right)(y - x)(t_2 - t_1)\\ &= (f'(z_2) - f'(z_1))(z_2 - z_1) \geq 0 \text{ par hypothèse.}\end{aligned}$$

Comme $t_2 - t_1 \geq 0$, on en déduit que $g'(t_2) - g'(t_1) \geq 0$. Donc g' est croissante, ainsi g est convexe par la propriété 2.4.1. Or $f(z) = g((z - x)/(y - x))$ pour $z \in [x, y]$. En posant $h : z \mapsto \frac{z-x}{y-x}$, on a $f = g \circ h$. Autrement dit, $f' = g' \circ h \times h'$ avec g' croissante, h croissante et h' constante positive donc f' est croissante, i.e. f convexe.

□

2.5 Opérations préservant la convexité

Proposition 2.5.1. *Soient $f, g : I \subset \mathbb{R} \mapsto \mathbb{R}$ deux fonctions convexes.*

1. *λf est convexe pour $\lambda \geq 0$.*
2. *$f + g$ est convexe sur I.*
3. *$\max(f, g)$ est convexe (valable aussi pour un nombre fini).*

Démonstration.

1. $\text{epi}(\lambda f) = \{(x, y), y \geq \lambda f(x)\}$. Soient $z_1 = (x_1, y_1), z_2 = (x_2, y_2) \in \text{epi}(\lambda f)$ et $\mu \in [0, 1]$.

$$\begin{aligned} & y_i \geq \lambda f(x_i)\\ \Rightarrow\ & \mu y_1 + (1 - \mu)y_2 \geq \mu\lambda f(x_1) + (1 - \mu)\lambda f(x_2)\\ \Leftrightarrow\ & \mu y_1 + (1 - \mu)y_2 \geq \lambda(\mu f(x_1) + (1 - \mu)f(x_2)) \geq \lambda f(\mu x_1 + (1 - \mu)x_2)\end{aligned}$$

   Donc $\mu z_1 + (1 - \mu)z_2 \in \text{epi}(\lambda f)$.

2. $\text{epi}(f + g) = \{(x, y), y \geq f(x) + g(x)\}$. Soient $z_1 = (x_1, y_1), z_2 = (x_2, y_2) \in \text{epi}(f + g)$ et $\mu \in [0, 1]$.

$$\begin{aligned} & y_i \geq f(x_i) + g(x_i)\\ \Rightarrow\ & \mu y_1 + (1 - \mu)y_2 \geq \mu(f(x_1) + g(x_1)) + (1 - \mu)(f(x_2) + g(x_2))\\ \Leftrightarrow\ & \mu y_1 + (1 - \mu)y_2 \geq \mu f(x_1) + (1 - \mu)f(x_2) + \mu g(x_1) + (1 - \mu)g(x_2)\\ \Leftrightarrow\ & \mu y_1 + (1 - \mu)y_2 \geq f(\mu x_1 + (1 - \mu)x_2) + g(\mu x_1 + (1 - \mu)x_2)\end{aligned}$$

   Donc $\mu z_1 + (1 - \mu)z_2 \in \text{epi}(f + g)$.

3.

$$\begin{aligned}\text{epi}(\max(f, g)) &= \{(x, y), y \geq \max(f(x), g(x))\} = \{(x, y), y \geq f(x), y \geq g(x)\}\\ &= \{(x, y), y \geq f(x)\} \cap \{(x, y), y \geq g(x)\} = \text{epi}(f) \cap \text{epi}(g)\end{aligned}$$

   Comme l'intersection de deux ensembles convexes est convexe, il est naturel que $\max(f, g)$ soit convexe.

□

Polycopié de cours

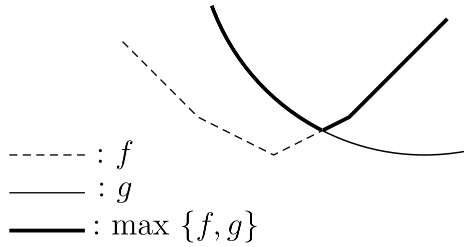


FIGURE 2.2 – Opération maximum de deux fonctions

Proposition 2.5.2. *Soient $I, J \subset \mathbb{R}$ deux intervalles de $\mathbb{R}$. Pour $f : J \mapsto \mathbb{R}$ et $g : I \mapsto J$ deux fonctions réelles, on définit la composée $f \circ g$ de f par g comme $f \circ g(x) = f(g(x))$.*
- *Si f et g sont convexes et f croissante alors $f \circ g$ est convexe.*
- *Si f et g sont différentiables sur I et J respectivement, alors $f \circ g$ convexe $\Leftrightarrow g' \times f' \circ g$ croissante sur I.*
- *Si f et g sont deux fois différentiables sur I et J respectivement, alors $f \circ g$ convexe $\Leftrightarrow g''(x) \times f' \circ g(x) + (g'(x))^2 \times f'' \circ g(x) \geq 0$ pour tout $x \in I$.*

En particulier, si g est affine, il suffit que f soit convexe pour garantir $f \circ g$ soit convexe.

Démonstration. Soient $x, y \in J$ et $\lambda \in [0, 1]$. Si $f : I \mapsto \mathbb{R}$ est croissante et $g : J \mapsto I$ est convexe. (les deux continues)

1. g convexe et $x, y \in J$
$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$$
2. f croissante et f convexe et $g(x), g(y) \in I$
$$f(g(\lambda x + (1-\lambda)y)) \leq f(\lambda g(x) + (1-\lambda)f(y)) \leq \lambda f(g(x)) + (1-\lambda)f(g(y))$$
3. autrement dit
$$f(g(\lambda x + (1-\lambda)y)) \leq \lambda f(g(x)) + (1-\lambda)f(g(y))$$
donc $f \circ g$ convexe.

On a $(f \circ g)'(x) = g'(x) \times f' \circ g(x)$ et $(f \circ g)''(x) = g''(x) \times f' \circ g(x) + (g'(x))^2 \times f'' \circ g(x)$. Par la proposition 2.4.1, c'est immédiat. □

Exemple 2.5.1. *Par exemple $g(x) = ax + b$ et $f(x) = x^2$, on a que $f \circ g(x) = (ax+b)^2$ est convexe sur $\mathbb{R}$.*

2.6 Variations sur la convexité

2.6.1 Fonctions concaves

Définition 2.6.1 (Fonction concave sur I)**.** *Une fonction $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ est concave si $-f$ est convexe. Autrement dit pour tout $\lambda \in]0, 1[$ et pour tout $x, y \in I$,*
$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y).$$
Si l'inégalité est stricte, on parle de fonction strictement concave.

Remarque 2.6.1. *Les fonctions $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ convexes et concaves sont les fonctions affines, i.e. $f(x) = ax + b$.*

Polycopié de cours

Exemple 2.6.2. *Avec des arguments du même type qu'au début du chapitre, on montre que* $f(x) = -x^2$ *est concave sur* $\mathbb{R}$.

Définition 2.6.3 (Hypographe)**.** *Pour une fonction* $f : I \subset \mathbb{R} \mapsto \mathbb{R}$*, l'hypographe se définit comme* $\mathrm{hyp}(f) = \{(x, y) \in I \times \mathbb{R}, y \leq f(x)\}$.

Théorème 2.6.2. *Pour une fonction* $f : I \subset \mathbb{R} \mapsto \mathbb{R}$*,* f *est concave* $\Leftrightarrow$ $\mathrm{hyp}(f)$ *est un ensemble convexe de* $\mathbb{R}^2$.

Démonstration. Immédiat □

2.6.2 Fonctions log-convexes et log-concaves

Définition 2.6.4 (Fonction log-concave/log-convexe sur I)**.** *Une fonction* $f : I \subset \mathbb{R} \mapsto \mathbb{R}_+$ *est log-convexe (log-concave) si* $\log(f)$ *est convexe (resp. si* $\log(f)$ *est concave). Autrement dit pour tout* $\lambda \in]0, 1[$ *et pour tout* $x, y \in I$,

$$\log(f(\lambda x + (1-\lambda)y)) \leq (\geq)\lambda \log(f(x)) + (1-\lambda)\log(f(y)) \Leftrightarrow f(\lambda x + (1-\lambda)y) \leq (\geq) f(x)^{\lambda} f(y)^{1-\lambda}.$$

Proposition 2.6.3. *Une fonction* $f : I \subset \mathbb{R} \mapsto \mathbb{R}_+$ *log-convexe est convexe sur* I*. Mais, une fonction* $f : I \subset \mathbb{R} \mapsto \mathbb{R}_+$ *log-concave n'est pas forcément concave sur* I.

Démonstration. Pour f log-convexe, on a

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

par concavité du *log*. Donc

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

Donc f convexe. □

Exemple 2.6.5 (densité de probabilité)**.** *Quelques exemples de probabilité :*
— *Loi exponentielle :* $f(x) = e^{-x}$ *est log-concave et log-convexe.*
— *Loi normale :* $f(x) = e^{-x^2/2}$ *est log-concave mais pas concave.*
— *Loi logistique :* $f(x) = 1/(1 + e^x)$ *est log-concave.*

2.7 Extrema des fonctions convexes univariées

2.7.1 Théorèmes

Théorème 2.7.1. *Soit* $f : I \subset \mathbb{R} \mapsto \mathbb{R}$ *différentiable et convexe pour* I *un intervalle ouvert. On a les équivalence suivantes pour* $x^\star \in I$ *:*
(1) $x^\star$ *est un minimum local de* f*, i.e.* $\exists r > 0, \forall x \in]x^\star - r, x^\star + r[, f(x^\star) \leq f(x)$.
(2) $x^\star$ *est un minimum global de* f*, i.e.* $\forall x \in I, f(x^\star) \leq f(x)$.
(3) $f'(x^\star) = 0$.
Si f *n'est pas différentiable, alors (3) devient* $0 \in [f'_-(x^\star), f'_+(x^\star)]$ *(toujours défini).*

Démonstration. (1) $\Rightarrow$ (2) Supposons que $x^\star$ est un minimum local de f sur $J \subset I$. Par construction $x^\star$ ne peut être aux points extrêmes de J mais à l'intérieur. S'il existe $y^\star \in I \setminus J$ tel que $y^\star$ soit le minimum global de f. Ainsi

$$\min_J f(x) = f(x^\star), \quad \min_{I \setminus J} f(x) = f(y^\star), \quad f(y^\star) < f(x^\star).$$

Polycopié de cours

Autrement dit
$$\forall x \in J, f(x) \geq f(x^\star), \quad \forall x \in I \setminus J, f(x) \geq f(y^\star)$$

Soit $\lambda \in]0,1[$. Par la convexité de f, on a

$$f(\lambda y^\star + (1-\lambda)x^\star) \leq \lambda(f(y^\star) - f(x^\star)) + f(x^\star) < f(x^\star),$$

puisque $f(y^\star) - f(x^\star) < 0$ et $\lambda > 0$. Comme J est ouvert, il existe λ_0 petit tel que $\forall \lambda < \lambda_0$, $x_\lambda = \lambda y^\star + (1-\lambda)x^\star \in J$ avec $f(x_\lambda) < f(x^\star)$. C'est impossible car $x^\star$ est un minimum local sur J. Ainsi par contradiction $f(x^\star) \leq f(y^\star)$.

(2) $\Rightarrow$ (3) Supposons que $x^\star$ est un minimum global de f sur I. C'est une propriété d'analyse. On a

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

pour $h > 0$. En $x^\star$, on a par hypothèse pour tout x, $f(x) \geq f(x^\star)$. Donc

$$\frac{f(x^\star + h) - f(x^\star)}{h} \geq 0, \quad \frac{f(x^\star - h) - f(x^\star)}{-h} \geq 0.$$

$$\Leftrightarrow \frac{f(x^\star + h) - f(x^\star)}{h} \geq 0, \quad \frac{f(x^\star - h) - f(x^\star)}{h} \leq 0.$$

Donc
$$\lim_{h \to 0} \frac{f(x^\star + h) - f(x^\star)}{h} = f'(x^\star) \geq 0, \quad \lim_{h \to 0} \frac{f(x^\star - h) - f(x^\star)}{h} = f'(^\star) \leq 0.$$

(3) $\Rightarrow$ (1) Comme f est convexe pour tout $x \in I$

$$f(x) \geq f(x^\star) + f'(x^\star)(x - x^\star)$$

Par hypothèse $f'(x^\star) = 0$, donc $f(x) \geq f(x^\star)$.

□

Exemple 2.7.1. *Considérons $f(x) = x^2$ et $I = \mathbb{R}$. Il est facile de voir que le minimum (global) est atteint à $x^\star = 0$. Le fait que 0 soit un minimum local se traduit par $f(0) = \min_{x \in]-1,1[} f(x)$ mais aussi $f(0) = \min_{x \in]-2,2[} f(x)$...Le fait que 0 soit un minimum glob al se traduit par $f(0) = \min_{x \in \mathbb{R}} f(x)$.*

Notons que si on choisit $I =]0, 10[$ alors le problème $\min_I f(x)$ n'a pas de solution dans I.

2.7.2 Application économétrique

Considérons une séries de flux $F_0, \dots, F_T$ versés aux dates $t = 0, 1, 2, \dots, T$, par exemple un remboursement de prêt ou les intérêts d'un placement. La somme actualisée des flux est définie par

$$f(r) = \sum_{t=0}^{T} \frac{F_t}{(1+r)^t},$$

pour $r > 0$. En pratique, $\text{sign}(F_0) \neq \text{sign}(F_1) = \cdots = \text{sign}(F_T)$. S est une fonction C^∞ sur $\mathbb{R}_+$. De plus, S est aussi convexe si $\text{sign}(F_i) > 0$ pour $i \geq 1$ et concave si $\text{sign}(F_i) < 0$ pour $i \geq 1$ puisque

$$f'(r) = -\sum_{t=1}^{T} \frac{tF_t}{(1+r)^{t+1}}, f''(r) = \sum_{t=1}^{T} \frac{t(t+1)F_t}{(1+r)^{t+2}}.$$

Le taux actuariel ou taux de rendement interne est le point $\bar{r} > 0$ tel que $f(\bar{r}) = 0$. L'existence est garantit par le théorème des valeurs intermédiaires puisque $f(0)f(\infty) < 0$. La convexité garantit une convergence rapide du calcul de la racine.

Polycopié de cours

Considérons la méthode de la tangente, initialisée par deux points x_0 et x_1, et de récurrence

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} f(x_n).$$

Etudions l'écart entre un point et la solution $\bar{r}$

$$|x_{n+1} - \bar{r}| = \left| x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} f(x_n) - \bar{r} \right| = |x_n - \bar{r}| \left| 1 - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \frac{f(x_n) - f(\bar{r})}{x_n - \bar{r}} \right|$$

Par la proposition 2.4.2, on a

$$m \leq \min(f'_-(\bar{r}), f'_-(x_n)) \leq \frac{f(x_n) - f(\bar{r})}{x_n - \bar{r}} \leq \max(f'_+(\bar{r}), f'_+(x_n)) \leq M$$

On a

$$1 - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \frac{f(x_n) - f(\bar{r})}{x_n - \bar{r}} = \underbrace{\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}}_{A_n} \left(\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} - \frac{f(x_n) - f(\bar{r})}{x_n - \bar{r}} \right)$$

et

$$\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} - \frac{f(x_n) - f(\bar{r})}{x_n - \bar{r}} = (x_{n-1} - \bar{r}) \underbrace{\frac{\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} - \frac{f(x_n) - f(\bar{r})}{x_n - \bar{r}}}{x_{n-1} - \bar{r}}}_{B_n}$$

Par convexité on borne A_n et B_n pour aboutir à

$$|x_{n+1} - \bar{r}| = |x_n - \bar{r}| \, |x_{n-1} - \bar{r}| \left| \frac{A_n}{B_n} \right| \leq |x_n - \bar{r}| \, |x_{n-1} - \bar{r}| \, M$$

En notant $e_n = |x_n - \bar{r}|$, on a une suite par récurrence des erreurs

$$e_{n+1} \leq e_n e_{n-1} M \leq \cdots \leq (e_0)^{p^n}$$

où p vérifie l'équation $p^2 = 2p + 1$ c'est à dire $p = (1 + \sqrt{5})/2$.

Polycopié de cours

Chapitre 3

Fonctions convexes multivariées (3 séances)

3.1 Fonctions convexes

Définition 3.1.1 (Fonction convexe)**.** *Une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}$ est convexe si pour tout $\lambda \in]0, 1[$ et pour tout $x, y \in \mathbb{R}^n$,*

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

Si l'inégalité est stricte, on parle de fonction strictement convexe.

Exemple 3.1.2 (Application linéaire)**.** *Soit $f(x) = a^T x$ pour $a \in \mathbb{R}$. Il est facile de voir*

$$f(\lambda x + (1-\lambda)y) = \lambda a^T x + (1-\lambda)a^T y = \lambda f(x) + (1-\lambda)f(y)$$

pour tout $x, y \in \mathbb{R}$ et pour tout $\lambda \in]0, 1[$

Exemple 3.1.3 (Application quadratique)**.** *Considérons $f(x) = x^T Ax$ pour $A \in \mathbb{R}^{n\times n}$ symétrique semi-définie positive. Pour tout $x \leq y \in \mathbb{R}^n$ et pour tout $\lambda \in]0, 1[$, on a*

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= (\lambda x + (1-\lambda)y)^T A(\lambda x + (1-\lambda)y) = ||\lambda x + (1-\lambda)y||_A^2 \\ &\leq ||\lambda x||_A^2 + ||(1-\lambda)y||_A^2 = \lambda ||x||_A^2 + (1-\lambda)||y||_A^2 = \lambda f(x) + (1-\lambda)f(y) \end{aligned}$$

car A définit une norme sur R^n.

Définition 3.1.4 (Fonction convexe sur C)**.** *Une fonction $f : C \subset \mathbb{R} \mapsto \mathbb{R}$ pour C un ensemble convexe est convexe si pour tout $\lambda \in]0, 1[$ et pour tout $x, y \in C$,*

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

Si l'inégalité est stricte, on parle de fonction strictement convexe.

Définition 3.1.5 (Fonction concave sur C)**.** *Une fonction $f : C \subset \mathbb{R} \mapsto \mathbb{R}$ pour C un ensemble convexe est concave si pour tout $\lambda \in]0, 1[$ et pour tout $x, y \in C$,*

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y).$$

Si l'inégalité est stricte, on parle de fonction strictement concave.

Polycopié de cours

3.2 Liens avec les ensembles convexes

Définition 3.2.1 (Epigraphe)**.** *Pour une fonction $f : C \subset \mathbb{R}^n \mapsto \mathbb{R}$, l'épigraphe se définit comme* $\mathrm{epi}(f) = \{(x, y) \in \mathbb{R}^{n+1}, y \geq f(x)\}$.

Théorème 3.2.1. *Pour une fonction $f : C \subset \mathbb{R}^n \mapsto \mathbb{R}$, f est convexe $\Leftrightarrow$ $\mathrm{epi}(f)$ est un ensemble convexe de $\mathbb{R}^{n+1}$.*

Démonstration. Similaire au cas univarié. □

Définition 3.2.2 (Hypographe)**.** *Pour une fonction $f : I \subset \mathbb{R} \mapsto \mathbb{R}$, l'hypographe se définit comme* $\mathrm{hyp}(f) = \{(x, y) \in I \times \mathbb{R}, y \leq f(x)\}$.

Théorème 3.2.2. *Pour une fonction $f : I \subset \mathbb{R} \mapsto \mathbb{R}$, f est concave $\Leftrightarrow$ $\mathrm{hyp}(f)$ est un ensemble convexe de $\mathbb{R}^2$.*

Démonstration. Immédiat □

3.3 Opérations préservant la convexité

Proposition 3.3.1. *Soient $f, g : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ deux fonctions convexes.*

1. *λf est convexe pour $\lambda \geq 0$.*
2. *$f + g$ est convexe.*
3. *$\max(f, g)$ est convexe (valable aussi pour un nombre fini).*

Démonstration. Similaire au cas univarié. □

3.4 Inégalités pour les fonctions convexes

3.4.1 Inégalités

Théorème 3.4.1 (Inégalités de Jensen)**.** *Soit $f : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ une fonction convexe. Pour tout $x_1, \ldots, x_r \in C$ et $\lambda_1, \ldots, \lambda_r \in [0, 1]$ tel que $\sum_{i=1}^r \lambda_i = 1$, alors*

$$f\left(\sum_{i=1}^{r} \lambda_i x_i\right) \leq \sum_{i=1}^{r} \lambda_i f(x_i).$$

Si f est concave, alors l'inégalité est inversée.

Démonstration. Similaire à la proposition 1.1.6. □

Proposition 3.4.2 (Inégalités de Hölder et Minkowski)**.** *Soient $p, q > 0$ tels que $1/p + 1/q = 1$. Considérons $a_1, \ldots, a_n$ et $b_1, \ldots, b_n$ $2n$ réels strictement positifs.*

L'inégalité de Hölder est

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} a_i^p\right)^{1/p} \left(\sum_{i=1}^{n} b_i^q\right)^{1/q}.$$

L'inégalité Minkowski est

$$\left(\sum_{i=1}^{n} (a_i + b_i)^p\right)^{1/p} \leq \left(\sum_{i=1}^{n} a_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} b_i^p\right)^{1/p}.$$

Polycopié de cours

Démonstration. Démonstration sans utiliser les normes.

1. Comme log est concave
$$\log(x^p/p + y^q/q) \geq \log(x^p)/p + \log(y^q)/q.$$
En passant à l'exponentielle,
$$x^p/p + y^q/q \geq xy.$$
2. En prenant la relation en a_i et b_i on obtient
$$a_i b_i \leq a_i^p/p + b_i^q/q$$
qui en les sommant donne
$$\sum_{i=1}^n a_i b_i \leq \sum_{i=1}^n a_i^p/p + \sum_{i=1}^n b_i^q/q = 1/p + 1/q = 1$$
3. On cherche une transformation α_i, β_i des a_i, b_i tels que $\sum_{i=1}^n \alpha_i^p = 1$ et $\sum_{i=1}^n \beta_i^q = 1$. Prendre $\alpha_i = a_i/(\sum_{i=1}^n a_i^p)$ ne suffit pas. On pose
$$\alpha_i = \frac{a_i}{(\sum_{j=1}^n a_j^p)^{1/p}}, \beta_i = \frac{b_i}{(\sum_{j=1}^n b_j^q)^{1/q}}.$$
On a bien $\sum_{i=1}^n \alpha_i^p = 1$ et $\sum_{i=1}^n \beta_i^q = 1$. Donc d'après la question précédente,
$$\sum_{i=1}^n \alpha_i \beta_i = \sum_{i=1}^n \frac{a_i}{(\sum_{j=1}^n a_j^p)^{1/p}} \frac{b_i}{(\sum_{j=1}^n b_j^q)^{1/q}} \leq 1.$$
On obtient l'inégalité de Holder.
4. On décompose
$$(a_i + b_i)^p = (a_i + b_i)^{p-1} a_i + (a_i + b_i)^{p-1} b_i.$$
Ainsi
$$\sum_{i=1}^n (a_i + b_i)^p = \sum_{i=1}^n (a_i + b_i)^{p-1} a_i + \sum_{i=1}^n (a_i + b_i)^{p-1} b_i$$
$$\Rightarrow \sum_{i=1}^n (a_i + b_i)^p \leq \left(\sum_{i=1}^n (a_i + b_i)^{pq-q}\right)^{1/q} \left(\sum_{i=1}^n a_i^p\right)^{1/p} + \left(\sum_{i=1}^n (a_i + b_i)^{pq-q}\right)^{1/q} \left(\sum_{i=1}^n b_i^p\right)^{1/p}$$
Or $1/p + 1/q = 1$ équivaut à $pq - q = p$, donc
$$\Leftrightarrow \sum_{i=1}^n (a_i + b_i)^p \leq \left(\sum_{i=1}^n (a_i + b_i)^p\right)^{1-1/p} \left(\sum_{i=1}^n a_i^p\right)^{1/p} + \left(\sum_{i=1}^n (a_i + b_i)^p\right)^{1-1/p} \left(\sum_{i=1}^n b_i^p\right)^{1/p}$$
$$\Leftrightarrow \sum_{i=1}^n (a_i + b_i)^p \leq \left(\sum_{i=1}^n (a_i + b_i)^p\right)^{1-1/p} \left(\left(\sum_{i=1}^n a_i^p\right)^{1/p} + \left(\sum_{i=1}^n b_i^p\right)^{1/p}\right)$$
$$\Leftrightarrow \left(\sum_{i=1}^n (a_i + b_i)^p\right)^{1/p} \leq \left(\sum_{i=1}^n a_i^p\right)^{1/p} + \left(\sum_{i=1}^n b_i^p\right)^{1/p}$$

□

Polycopié de cours

3.4.2 Liens avec les normes et les distances

Exemple 3.4.1 (L_p normes)**.** *Pour $x \in \mathbb{R}^n$, la norme L_p est*

$$||x||_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}.$$

Notons que pour $p = \infty$, on a $||x||_\infty = \max_i |x_i|$. Pour tout $p \geq 1$, la fonction $||.||_p$ est une norme, i.e. positive, définie, positivement homogène et vérifiant l'inégalité triangulaire.

Pour tout $p \geq 1pas$, la fonction norme $||.||_p$ est une fonction convexe. En effet, pour tout $x, y \in \mathbb{R}^n$ et $\lambda \in [0,1]$,

$$||\lambda x + (1-\lambda)y||_p \leq ||\lambda x||_p + ||(1-\lambda)y||_p = \lambda||x||_p + (1-\lambda)||y||_p.$$

Autrement dit, l'inégalité de Hölder a pour cas particulier l'inégalité de Cauchy Schwarz

$$< a, b > \leq ||a||_p||b||_p.$$

L'inégalité de Minkowski se réécrit comme l'inégalité triangulaire

$$||a+b||_p \leq ||a||_p + ||b||_p.$$

Exemple 3.4.2 (Fonction support d'un programme linéaire)**.** *Soit f la fonction définie par $f(c) = \max\{c^T x, x \in S\}$ où S est un ensemble convexe non vide. Typiquement S est un polytope de $\mathbb{R}^n$. Soit $\lambda > 0$.*

$$f(\lambda c) = \max\{(\lambda c)^T x, x \in S\} = \max\{\lambda(c^T x), x \in S\} = \lambda \max\{c^T x, x \in S\} = \lambda f(c).$$

Comme S est positivement homogène, on le représente comme un polytope de ses points extrêmaux $S = \mathrm{co}(\{v_1, \dots, v_p\})$. La solution d'un problème linéaire est atteinte sur un des points extrêmaux, donc

$$f(c) = \max_{j=1,\dots,p} c^T v_j = \max_{j=1,\dots,p} f_j, \text{ où } f_j(c) = c^T v_j.$$

Comme f_j sont linéaires, elles sont aussi convexes. Donc $f = \max_j f_j$ est aussi convexe par 3.3.1.

3.5 Cas des fonctions différentiables

3.5.1 Dérivée directionnelle

Définition 3.5.1. *Pour une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}$, la dérivée directionnelle de f en x_0 selon la direction z est définie par*

$$f'(x_0; z) = \lim_{t \to 0} \frac{f(x_0 + tz) - f(x_0)}{t}.$$

Pour $z = e_j$ le $j$ème vecteur unitaire, $f'(x_0; e_j)$ est notée

$$\frac{\partial f}{\partial x_j}(x_0) = \lim_{t \to 0} \frac{f(x_0 + te_j) - f(x_0)}{t}.$$

Exemple 3.5.2. *Dans $\mathbb{R}^2$, considérons $f(x_1, x_2) = e^{x_1+x_2} x_1 x_2$.*

$$\begin{aligned}
f'(x; z) &= \lim_{t \to 0} \frac{e^{x_1+x_2+tz_1+tz_2}(x_1+tz_1)(x_2+tz_2) - e^{x_1+x_2}x_1x_2}{t} \\
&= \lim_{t \to 0} e^{x_1+x_2} \frac{(1 + tz_1 + tz_2 + o(t))(x_1+tz_1)(x_2+tz_2) - x_1x_2}{t} \\
&= \lim_{t \to 0} e^{x_1+x_2} \frac{(tz_1 + tz_2 + o(t))(x_1x_2 + tz_1x_2 + tz_2x_1 + o(t^2)) + tz_1x_2 + tz_2x_1 + o(t^2)}{t} \\
&= \lim_{t \to 0} e^{x_1+x_2} \left[(z_1 + z_2 + o(1))(x_1x_2 + tz_1x_2 + tz_2x_1 + o(t^2)) + z_1x_2 + z_2x_1 + o(t)\right] \\
&= e^{x_1+x_2}\left[(z_1+z_2)x_1x_2 + z_1x_2 + z_2x_1\right]
\end{aligned}$$

Polycopié de cours

Pour $z = (1,0)$ et $z = (0,1)$, on retrouve ce que l'on obtiendrait en dérivant naturellement

$$\frac{\partial f}{\partial x_1}(x) = e^{x_1+x_2}x_1x_2 + e^{x_1+x_2}x_2, \ \frac{\partial f}{\partial x_2}(x) = e^{x_1+x_2}x_1x_2 + e^{x_1+x_2}x_1.$$

Proposition 3.5.1. *Pour une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}^n$, on définit*

$$g_{x,z} : t \mapsto g_{x,z}(t) = f(x+tz),$$

pour x, z fixé. $g_{x,z}$ convexe sur $\mathbb{R}$ pour tout $x, z \in \mathbb{R}^n$ équivaut à f convexe sur R^n.

Démonstration. $\Leftarrow$ Par simple manipulation,

$$g(t_1\lambda + (1-\lambda)t_2) = f(x + (t_1\lambda + (1-\lambda)t_2)z) = f((x+t_1z)\lambda + (1-\lambda)(x+t_2z))$$

Par hypothèse $f((x+t_1z)\lambda + (1-\lambda)(x+t_2z)) \leq f(x+t_1z)\lambda + (1-\lambda)f(x+t_2z)$, donc

$$g(t_1\lambda + (1-\lambda)t_2) \leq g(t_1)\lambda + (1-\lambda)g(t_2)$$

$\Rightarrow$ On a par hypothèse

$$f(\lambda x + (1-\lambda)y) = f(y + \lambda(x-y)) = g(\lambda) = g(\lambda \times 1 + (1-\lambda) \times 0) \leq \lambda g(1) + (1-\lambda)g(0)$$

Ainsi

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(y+x-y) + (1-\lambda)f(y+0) = \lambda f(x) + (1-\lambda)f(y).$$

□

Exemple 3.5.3. *Dans $\mathbb{R}^2$, considérons $f(x_1,x_2) = e^{x_1+x_2}x_1x_2$. On définit*

$$g(t) = e^{x_1+x_2+t(z_1+z_2)}(x_1+tz_1)(x_2+tz_2)$$

donc la dérivée devient

$$g'(t) = (z_1+z_2)e^{x_1+x_2+t(z_1+z_2)}(x_1+tz_1)(x_2+tz_2) + e^{x_1+x_2+t(z_1+z_2)}(z_1+z_2).$$

Il est difficile de voir si cette fonction n'est pas croissante par rapport à t pour tout x, z.

$$\begin{array}{rcl} g''(t) & = & (z_1+z_2)^2e^{x_1+x_2+t(z_1+z_2)}(x_1+tz_1)(x_2+tz_2) + (z_1+z_2)e^{x_1+x_2+t(z_1+z_2)}(z_1+z_2) + e^{x_1+x_2+t(z_1+z_2)}(z_1+z \\ & = & (z_1+z_2)^2e^{x_1+x_2+t(z_1+z_2)}[(x_1+tz_1)(x_2+tz_2)+2] \end{array}$$

Pour $z_1 = -1, z_2 = 0$, $g''(t) = e^{x_1+x_2-t}[(x_1-t)x_2+2]$ n'est pas positive pour tout t et x.

Théorème 3.5.2. *Soit $f : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ une fonction convexe. f est continue sur C et admet des dérivées directionnelles.*

Définition 3.5.4 (Différentiable)**.** *Pour une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}$, f est différentiable si pour tout x, il existe un unique d_x tel que*

$$\lim_{||h|| \to 0} \frac{f(x+h) - f(x) - d_x^T h}{||h||} = 0.$$

Si f est différentiable, alors f admet une dérivée directionnelle pour $z \neq 0$.

Polycopié de cours

Définition 3.5.5 (Gradient)**.** *Pour une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}$ différentiable, le gradient de f en x est défini par*

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

$\nabla f(x)$ est le vecteur d_x dans la définition précédente.

Définition 3.5.6 (matrice Hessienne)**.** *Pour une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}$ deux fois différentiable, la matrice Hessienne de f en x est définie par*

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ & \ddots & \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}.$$

C'est une matrice symétrique.

3.5.2 Caractérisation par le gradient ou matrice Hessienne

Théorème 3.5.3. *Pour une fonction $f : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ différentiable, On a les équivalences suivantes*

1. *f est convexe sur C*
2. *pour tout $x, y \in C$, $f(x) \geq f(y) + \nabla f(y)^T (x - y)$.*
3. *pour tout $x, y \in C$, $(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$.*

Démonstration. 1. $\Rightarrow$ 2. Soient $x, y \in C$, $\lambda \in [0, 1]$.

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

donc

$$f(x + \lambda(y - x)) - f(y) \leq \lambda(f(x) - f(y)) \Rightarrow$$

$$\frac{f(x + \lambda(y - x)) - f(y)}{\lambda} + \nabla f(y)^T (x - y) \leq f(x) - f(y) + \lambda f(x) + \nabla f(y)^T (x - y).$$

Lorsque $\lambda \to 0^+$, on obtient

$$0 \leq f(x) - f(y) + \nabla f(y)^T (x - y).$$

2. $\Rightarrow$ 3. Par symétrie, on a

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

En utilisant l'hypothèse et en sommant, on trouve

$$f(x) + f(y) \geq f(x) + f(y) + \nabla f(x)^T (y - x) + \nabla f(y)^T (x - y) \Leftrightarrow (\nabla f(x) - \nabla f(y))^T (x - y) \geq 0.$$

3. $\Rightarrow$ 1. On pose $g : t \mapsto f(x + t(y - x))$ pour $t \in [0, 1]$. Donc $g(0) = f(x)$ et $g(1) = f(y)$. En dérivant on obtient pour $0 \leq t_1 < t_2 \leq 1$, par la règle de composition (par ex Wikipedia) $g'(t) = \nabla f(x + t(y - x))^T (y - x)$. Donc

$$\begin{aligned} g'(t_2) - g'(t_1) &= \nabla f(\underbrace{x + t_2(y - x)}_{x_2})^T (y - x) - \nabla f(\underbrace{x + t_1(y - x)}_{x_1})^T (y - x) \\ &= \frac{(\nabla f(x_2) - \nabla f(x_1))^T (x_2 - x_1)}{t_2 - t_1} \geq 0, \end{aligned}$$

puisque le numérateur est positif par (3) et le dénominateur par $t_1 < t_2$. Donc g' est croissante. Par la proposition 2.4.1, g est convexe et par la proposition 3.5.1, f est convexe.

□

Polycopié de cours

Remarque 3.5.4. *Comme dans le cas univarié, le théorème montre qu'une approximation de Taylor à l'ordre 1 sous-estime toujours la valeur de la fonction f.*

Théorème 3.5.5. *Pour une fonction $f : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^2, On a les équivalences suivantes*

1. *f est convexe sur C*
2. *pour tout $x \in C$, $\nabla^2 f(x)$ semi-définie positive, i.e. $\forall z \in \mathbb{R}^n, z^T \nabla^2 f(x) z \geq 0$.*

Démonstration. Par la proposition 3.5.1, on s'intéresse à la fonction $g_{x,z}$. Comme f est C^2, g l'est aussi et donc

$$g'(t) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x+tz) \times z_j, \quad g''(t) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x+tz) \times z_j \times z_i = z^T \nabla f(x+tz) z.$$

Par la proposition 2.4.1, g convexe équivaut à $g''(t) \geq 0$ pour $t \in \mathbb{R}$. Donc pour tout $x, z \in \mathbb{R}^n$,

$$\forall t, z^T \nabla f(x+tz) z \geq 0 \Rightarrow z^T \nabla f(x) z \geq 0.$$

On a procédé par équivalence. □

Exemple 3.5.7. *Soit $f : \mathbb{R}^n \mapsto \mathbb{R}$ définie par $f(x) = x^T A x + b^T x + c$, avec A symétrique semi-définie positive. Il est facile de voir que $\nabla f(x) = x^T A + Ax + b = 2Ax + b$ et $\nabla^2 f(x) = 2A$. C'est une matrice semi-définie positive.*

3.6 Optimisation convexe

Théorème 3.6.1. *Soit $f : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ une fonction différentiable et convexe sur C un ensemble ouvert et convexe. On a les équivalence suivantes pour $x^\star \in C$:*

- *$x^\star$ est un minimum local de f.*
- *$x^\star$ est un minimum global de f.*
- *$\nabla f(x^\star) = 0$.*

Exemple 3.6.1. *Considérons le cas simple d'une fonction quadratique pour $x \in \mathbb{R}^n$ (avec A symétrique)*

$$f(x) = x^T A x + b^T x + c.$$

La condition d'optimalité est $2Ax + b = 0$.

Théorème 3.6.2. *Soit $f : C \subset \mathbb{R}^n \mapsto \mathbb{R}$ une fonction C^1 et convexe sur C un ensemble convexe. $x^\star \in C$ est un minimum (global) de f sur C ssi*

$$\forall x \in C, \nabla f(x^\star)^T (x - x^\star) \geq 0.$$

Un tel point $x^\star$ est appelé un point stationnaire.

Démonstration. $\Rightarrow$ Supposons $\nabla f(x^\star)^T (x_0 - x^\star) < 0$ pour $x_0 \in C$. On applique le théorème des accroissements finis à la fonction $g(\epsilon) = f(x^\star + \epsilon(x_0 - x^\star))$. Pour tout $\epsilon > 0$, il existe $s \in [0,1]$

$$f(x^\star + \epsilon(x_0 - x^\star)) = f(x^\star) + \epsilon \nabla f(x^\star + s\epsilon(x_0 - x^\star))^T (x_0 - x^\star).$$

Comme $\nabla f(x^\star)^T (x_0 - x^\star) < 0$ et que le gradient est continu, pour ϵ suffisamment petit, $\nabla f(x^\star + s\epsilon(x_0 - x^\star)) < 0$. Autrement dit $f(x^\star + \epsilon(x_0 - x^\star)) < f(x^\star)$ et comme C est convexe, $x^\star + \epsilon(x_0 - x^\star) \in C$. C'est contradictoire.

Polycopié de cours

$\Leftarrow$ Par le théorème 3.5.3, on a pour tout $x \in C$,

$$f(x) \geq f(x^\star) + \nabla f(x^\star)^T (x - x^\star) \geq f(x^\star),$$

par hypothèse.

$\square$

Exemple 3.6.2. *Soit $f : C \mapsto \mathbb{R}$ où $C = \mathbb{R}^n_+$. Un point stationnaire de f vérifie pour tout $x \in \mathbb{R}^n_+$*

$$\nabla f(x^\star)^T (x - x^\star) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^\star)(x_i - x_i^\star) \geq 0$$

En particulier pour $x_i \in [0, x_i^\star]$. Si $x_i^\star > 0$, alors $\frac{\partial f}{\partial x_i}(x^\star) = 0$. Sinon $x_i^\star = 0$, on a $\frac{\partial f}{\partial x_i}(x^\star) \geq 0$.

3.6.1 Fonction de Lagrange et dualité de Lagrange

Considérons le problème d'optimisation suivant

$$\min_{x \in \mathbb{R}^n} f(x) \text{ tel que } g_l(x) \leq 0 \text{ pour } l = 1 \ldots, m. \qquad (P)$$

Définition 3.6.3 (Lagrangien)**.** *Pour le problème (P), le Lagrangien $L : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$ est défini par*

$$L(x, \lambda) = f(x) + \sum_{l=1}^{m} \lambda_i g_i(x).$$

Les multiplicateurs de Lagrange λ_i s'interprète comme la sensibilité de l'optimum à cette contrainte.

Proposition 3.6.3 (Condition d'optimalité KKT (premier ordre))**.** *Supposons que f soit C^1. Si $x^\star$ résout le problème d'optimisation (P), alors il existe $\lambda^\star \in \mathbb{R}^m$ tel que*

- *minimalité (premier ordre) : $\nabla_x L(x^\star, \lambda^\star) = 0$,*
- *condition primale : pour tout l, $g_l(x^\star) \leq 0$,*
- *condition duale : pour tout l, $\lambda_l^\star \geq 0$,*
- *complémentarité : pour tout l, $\lambda_l^\star g_l(x^\star) = 0$.*

Dans le cas d'une égalité pour g_l, la condition primale devient $g_l(x^\star) = 0$, la condition duale n'est plus nécessaire, tandis que la condition de complémentarité est toujours vérifiée.

Proposition 3.6.4 (Condition second ordre)**.** *Si les fonctions f et g_l sont convexes, alors la solution $x^\star$ des conditions du premier ordre est un minimum global du problème d'optimisation.*

Exemple 3.6.4. *Soit $f(x) = x_1^2 + 3x_2^2 - 2x_1x_2 - x_1$ sous contrainte $g_1(x) = x_1 + x_2 - 1$. Le lagrangien vaut*

$$L(x, \lambda_1) = x_1^2 + 3x_2^2 - 2x_1x_2 - x_1 + \lambda_1(x_1 + x_2 - 1).$$

La condition du premier ordre est

$$\begin{cases} 2x_1 - 2x_2 - 1 + \lambda_1 = 0 \\ 6x_2 - 2x_1 + \lambda_1 = 0 \end{cases} \Leftrightarrow \begin{cases} 2x_1 - 1 + \lambda_1 = 2x_2 \\ 6(x_1 - 1/2 + \lambda_1/2) + \lambda_1 = 2x_1 \end{cases} \Leftrightarrow \begin{cases} 2x_2 = 2x_1 - 1 + \lambda_1 \\ 4x_1 = -4\lambda_1 + 3 \end{cases}$$

La contrainte de complémentarité correspond à

$$\lambda_1(x_1 + x_2 - 1) = 0.$$

- *Si $\lambda_1 = 0$, alors on trouve*

$$x_1 = 3/4, x_2 = 1/4.$$

C'est une solution possible car $g_1(3/4, 1/4) = 0$.

- *Si $\lambda_1 > 0$ alors $x_1 + x_2 = 1$. Ainsi en utilisant $x_2 = 1 - x_1$, on trouve un système sans solution*

$$\begin{cases} 2 - 2x_1 = 2x_1 - 1 + \lambda_1 \\ 4x_1 = -4\lambda_1 + 3 \end{cases} \Leftrightarrow \begin{cases} 4x_1 = -4\lambda_1 + 3 \\ 3 - 4x_1 = \lambda_1 \end{cases} \Leftrightarrow \begin{cases} 4x_1 = -4\lambda_1 + 3 \\ 3 + 4\lambda_1 - 3 = \lambda_1 \end{cases} \Leftrightarrow \begin{cases} 4x_1 = -4\lambda_1 + 3 \\ 4 = 1 \end{cases}$$

C'est donc impossible. Ainsi l'unique solution est $(x_1^\star, x_2^\star, \lambda_1^\star) = (1/4, 3/4, 0)$.

Polycopié de cours

3.6.2 Problèmes linéaires

Considérons le problème linéaire

$$\min_{x\in\mathbb{R}^n} c^T x \text{ tel que } Ax = b, x \geq 0 \qquad (LP)$$

Autrement dit

$$\left\{\begin{array}{l} g_k(x) = -x_k, k = 1,\dots,n \\ h_k(x) = A_{k,.}x - b, k = 1,\dots,m \end{array}\right.$$

Le lagrangien s'écrit

$$L(x,\lambda,\mu) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \sum_{i=1}^{m} \mu_i (A_{i,.}x - b)$$

Les conditions du premier ordre s'écrivent

$$\left\{\begin{array}{l} \nabla_x L(x,\lambda) = 0 \Leftrightarrow c - \lambda + A^T\mu = 0 \\ g_k(x) \leq 0, h_k(x) = 0 \Leftrightarrow x \geq 0, \quad Ax = b, \\ \lambda_k \geq 0 \Leftrightarrow \lambda \geq 0 \\ \lambda_k g_k(x) = 0 \Leftrightarrow \lambda_k x_k = 0, k = 1,\dots,m \end{array}\right.$$

En éliminant λ, on obtient

$$\left\{\begin{array}{l} c + A^T\mu = \lambda \\ x \geq 0, \quad Ax = b, \\ c + A^T\mu \geq 0 \\ (c + A^T\mu)_k \times x_k = 0, k = 1,\dots,m \end{array}\right.$$

3.6.3 Problèmes quadratiques

Considérons le problème quadratique

$$\min_{x\in\mathbb{R}^n} 1/2x^T Dx - q^T x \text{ tel que } Ax = b \qquad (LP)$$

où A est semi-définie positive. Autrement dit

$$\left\{ h_k(x) = A_{k,.}x - b, k = 1,\dots,m \right.$$

Le lagrangien s'écrit

$$L(x,\lambda) = 1/2x^T Dx - q^T x + \sum_{i=1}^{m} \lambda_i (A_{i,.}x - b)$$

Les conditions du premier ordre s'écrivent

$$\left\{\begin{array}{l} \nabla_x L(x,\lambda) = 0 \Leftrightarrow Dx - q + A^T\lambda = 0 \\ h_k(x) = 0 \Leftrightarrow Ax = b, \\ \lambda_k \geq 0 \Leftrightarrow \lambda \geq 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} \begin{pmatrix} D & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} q \\ b \end{pmatrix} \\ \lambda \geq 0 \end{array}\right.$$

Un problème quadratique plus simple est le problème des moindres carrés, cf. le TD6.

Polycopié de cours