Complément d'agreg : Convexité et Optimisation

Baptiste Huguet

Table des matières

Chapitre 1

Convexité

1.1 Généralité

Soit E un espace vectoriel sur $\mathbb{R}$ ou $\mathbb{C}$.

Définition 1. *Soit $x, y \in E$. La droite passant par x et y est $\{\lambda x + (1-\lambda)y / \lambda \in \mathbb{R}\}$ et le segment entre x et y est $[x, y] = \{\lambda x + (1-\lambda)y / 0 \leq \lambda \leq 1\}$.*

Définition 2. *Un sous-ensemble $C \subset E$ est convexe si pour tout $x, u \in C$, $[x, y] \subset C$.*

Exemple 3. *Si E est un evn alors les boules (ouvertes et fermées) sont convexes. Dans $E = \mathbb{R}^d$ cela inclus aussi les pavés.*

Contre-exemple 4. *Dans un espace métrique, les boules ne sont pas nécessairement convexes. Par exemple dans le Fréchet $\mathcal{C}(\mathbb{R}, \mathbb{R})$ muni de*

$$d(f,g) = \sum_{n=1}^{\infty} 2^{-n} \frac{d_n(f,g)}{1 + d_n(f,g)}, \quad d_n(f,g) = \sup_{|x| \leq n} |f(x) - g(x)|$$

on définit $f(x) = (1 - |x|)_+$ et $g(x) = 100 f(x-2)$. On a $d(0,f) = 1/2$, $d(0,g) = 50/101$ mais $h = (f+g)/2 \notin B(0, 1/2)$.

La notion de convexité est stable par les opérations suivantes

Proposition 5. *(i) Soit I un ensemble et $(C_i)_{i \in I}$ une famille de sous-ensembles convexes de E, alors $\cap_{i \in I} C_i$ est convexe.*

(ii) (Somme de Minkowski) Si C_1 et C_2 convexes de E et $\alpha, \beta \in \mathbb{R}$, alors $\alpha C_1 + \beta C_2 \subset E$ est convexe.

(iii) Soit $C_1 \subset E$ et $C_2 \subset F$ convexes, alors $C_1 \times C_2 \subset E \times F$ est convexe.

Définition 6. *Un vecteur $y \in E$ est une combinaison convexe des vecteurs $x_1, \ldots x_n \in E$ s'il existe $\lambda_1, \ldots \lambda_n \in [0,1]$ tel que $\sum \lambda_i = 1$ et $y = \sum \lambda_i x_i$.*

Proposition 7. *Soit $C \subset E$. C est convexe ssi C est stable par combinaison convexe.*

Démonstration. $\Leftarrow$ Immédiat (le segment $[x, y]$ est l'ensemble des combinaisons convexe de x et y).

$\Rightarrow$ on procède par récurrence sur la taille $n \geq 2$ des combinaisons convexes. On suppose que C est convexe. Le cas $n = 2$ (initialisation) correspond à la définition de la convexité. Il est donc vérifié. On suppose que le cas n est vérifié. Soient $x_1, \ldots x_{n+1} \in C$, $\lambda_1, \ldots \lambda_{n+1} \in [0,1]$ tel que $\sum \lambda_i = 1$ et $y = \sum \lambda_i x_i$. On peut supposer que $\lambda_{n+1} \notin \{0, 1\}$. On a alors

$$y = \lambda_{n+1} X_{n+1} + (1 - \lambda_{n+1}) \sum_{i=1}^{n} \mu_i x_i,$$

avec $\mu_i = \lambda_i / (1 - \lambda_{n+1})$. En particulier, $\mu_i \geq 0$ et $\sum \mu_i = 1$ donc $\mu_i \in [0,1]$ et $z = \sum \mu_i x_i \in C$ par hypothèse de récurrence. D'où $u \in C$ par définition de la convexité. Ce qui achève la récurrence. $\square$

Définition 8. *Soit $A \subset E$. l'enveloppe convexe de A est le plus petit ensemble convexe* $\mathrm{conv}(A) \subset E$ *contenant A, i.e*

$$conv(A) = \bigcap_{A \subset C, C cvx} C.$$

Proposition 9. *Soit $A \subset E$, alors* $\mathrm{conv}(A)$ *est l'ensemble des combinaisons convexes de points de A.*

Démonstration. Notons C l'ensemble des combinaison convexe de A.

$C \subset \mathrm{conv}(A)$. $\mathrm{conv}(A)$ est convexe et contient A. D'après la Proposition 7, $\mathrm{conv}(A)$ contient les combinaisons convexes d'éléments de $\mathrm{conv}(A)$, donc en particulier de A.

$\mathrm{conv}(A) \subset C$. Il suffit de montrer que C est convexe. En effet, $A \subset C$ et $\mathrm{conv}(A)$ est le plus petit convexe contenant A. Soient $x, y \in C$, il existe $x_1, \ldots, x_n$ et $y_1, \ldots, y_p \in A$ ainsi que $\lambda_i, \mu_i \in [0, 1]$ tel que $\sum \lambda_i = \sum \mu_i = 1$ et $x = \sum \lambda_i x_i$, $y = \sum \mu_i y_i$. pour tout $\nu \in [0, 1]$, on a

$$\nu x + (1 - \nu) y = \sum_{i=1}^{n+p} \nu_i z_i,$$

avec $z_i = x_i$, $nu_i = \nu \lambda_i$ pour $1 \leq i \leq n$ et $z_i = y_{i-n}$, $\nu_i = (1 - \nu)\mu_{i-n}$ pour $n + 1 \leq i \leq n + p$. On a bien $z_i \in A$, $nu_i \geq 0$ et $\sum \nu_i = 1$. Ce qui montre que C est convexe. □

Application 10 (Théorème de Gauß-Lucas). *Soit $P \in \mathbb{C}[X]$ non constant. Alors les racines de P' sont dans l'enveloppe convexes des racines de P.*

Démonstration. $P(X) = \Pi_{i=1}^{d} (X - z_i)^{m_i}$. Considérons sa dérivée logarithmique $Q = P'/P = \sum_{i=1}^{d} m_i/(X - z_i)$. Soit z une racine de P' qui ne soit pas une racine de P. On a alors

$$0 = \sum \frac{m_i}{z - z_i} = \sum \frac{m_i}{|z - z_i|^2} (\bar{z} - \bar{z}_i).$$

Ainsi $z = \sum \alpha_1 z_i$ avec $\alpha_i = m_i/(|z - z_i|^2 \sum m_j/|z - z_j|^2) \geq 0$ et $\sum \alpha_i = 1$. □

Dans le cas où E est de dimension finie ($E = \mathbb{R}^d$), on peut borner la taille des combinaisons convexes nécessaires.

Theorem 11 (Carathéodory). *Si $E = \mathbb{R}^d$, $d \geq 1$ et $A \subset E$, alors* $\mathrm{conv}(A)$ *est l'ensemble des combinaisons convexes d'au plus $d + 1$ points de A.*

Contre-exemple 12. *Attention ! Le théorème NE DIT PAS qu'il suffit de $d + 1$ points pour engendrer* $\mathrm{conv}(A)$, *seulement que pour tout élément de* $\mathrm{conv}(A)$ *il existe $d + 1$ points dont il est une combinaison. Par exemple, trois points n'engendreront jamais* $\mathrm{conv}(\{(\pm 1, \pm 1)\})$ *dans $\mathbb{R}^2$.*

Démonstration. Notons C_{d+1} l'ensemble des combinaisons convexes d'au plus $d + 1$ éléments. On a de manière immédiate $C_{d+1} \subset \mathrm{conv}(A)$. Soit $x \in \mathrm{conv}(A)$. D'après la proposition précédente, $x = \sum_{i=1}^{n} \lambda_i x_i$. Si $n < d + 2$ alors $x \in C_{d+1}$, sinon, on montre que x est une combinaison convexe d'au plus $n - 1$ élément de A. En effet, la famille $\{x_2 - x_1, \ldots, x_n - x_1\} \subset \mathbb{R}^d$ est liée. On dispose donc de $\alpha_2, \ldots, \alpha_n \in \mathbb{R}$, non tous nuls, tels que

$$\sum_{i=2}^{n} \alpha_i (x_i - x_1) = 0.$$

On pose alors $\alpha_1 = -\sum_{i=2}^{n} \alpha_i$ de sorte que $\sum_{i=1}^{n} \alpha_i = 0$. On en déduit qu'il existe au moins un α_i strictement positif. On définit alors

$$\mu = \min_i \{\frac{\lambda_i}{\alpha_i} / \alpha_i > 0\}.$$

En particulier, pour tout i, on a $\mu_i = \lambda_i - \mu \alpha_i \geq 0$ et au moins l'un de ses coefficient est nul. De plus on a $\sum \mu_i = \sum \lambda_i - \mu \sum \alpha_i = 1$ et $x = \sum \mu_i x_i$. Ainsi, x est combinaison convexe d'au plus $n - 1$ point. On termine par récurrence. □

Une conséquence de ce théorème est la préservation de la compacité en dimension finie.

Corollaire 13. *Si $E = \mathbb{R}^d$ et $A \subset E$ compact, alors* $\operatorname{conv}(A)$ *est compact.*

Démonstration. Notons $\Delta = \{(\lambda_i) \in [0,1]^{d+1} / \sum_i \lambda_i = 1\}$. C'est un fermé borné de $\mathbb{R}^{d+1}$, donc compact. On considère alors l'application

$$\varphi : \begin{array}{ccc} \mathbb{R}^{d+1} \times (\mathbb{R}^d)^{d+1} & \to & \mathbb{R}^d \\ (\lambda_1, \dots, \lambda_{d+1}, x_1, \dots, x_{d+1}) & \mapsto & \sum_i \lambda_i x_i \end{array}.$$

Cette application est continue et vérifie $\varphi(\Delta \times A^{d+1}) = \operatorname{conv}(A)$. Donc $\operatorname{conv}(A)$ est compact. □

En dimension ∞, l'enveloppe convexe d'un compact n'est pas nécessairement compact, ni même fermée.

Contre-exemple 14. *Dans $E = \ell^2$, on pose $u_n = e_n/n$ où $e_n(k) = \delta_{nk}$. On pose $A = \{0\} \cup \{u_1, u_2, \dots\}$. A est compact mais la suite $x_n = \sum_{i=1}^n \frac{1}{2^i} u_i \in$* $\operatorname{conv}(A)$ *converge vers $x^* \notin$* $\operatorname{conv}(A)$.

En revanche, le caractère borné est systématiquement préservé.

Proposition 15. *Soit E un evn et $A \subset E$ borné. Alors* $\operatorname{conv}(A)$ *est bornée et de même diamètre que A.*

Démonstration. Par inclusion $A \subset \operatorname{conv}(A)$, on a $\delta(A) \leq \delta(\operatorname{conv}(A))$. Par ailleurs, $A \subset B(0,R)$ qui est convexe, donc $\operatorname{conv}(A) \subset B(0,R)$. Ainsi, $\operatorname{conv}(A)$ est bornée.

Soit $x \in \operatorname{conv}(A)$ et $y \in A$, on a $x = \sum_i \lambda_i x_i$ avec $x_i \in A$. On a alors

$$\|x - y\| \leq \sum \lambda_i \|x_i - y\| \leq \sum \lambda_i \delta(A) = \delta(A).$$

Soit $x, y \in \operatorname{conv}(A)$, alors d'après l'inégalité précédente, on a

$$\|x - y\| \leq \sum \lambda_i \|x_i - y\| \leq \sum \lambda_i \delta(A) = \delta(A).$$

On a donc montré que $\delta(\operatorname{conv}(A)) \leq \delta(A)$. □

De manière générale, et ce en toute dimension, le passage à l'enveloppe convexe ne préserve pas le caractère fermé.

Contre-exemple 16. *Dans $E = \mathbb{R}^2$, avec $A = (0,0) \cup \mathbb{R} \times \{1\}$. Alors* $\operatorname{conv}(A) = (0,0) \cup \mathbb{R} \times]0,1]$.

En revanche le caractère ouvert est préservé.

Proposition 17. *Soient un evn et $A \subset E$ ouvert, alors* $\operatorname{conv}(A)$ *est ouverte.*

Démonstration. Soit $x \in \operatorname{conv}(A)$, on dispose d'une décomposition $x = \sum_{i=1}^n \lambda_i x_i$. On peut supposer que $\lambda_1 \neq$. On définit alors l'application

$$f : z \in E \mapsto \lambda_1^{-1} \left(z - \sum_{i=2}^n \lambda_i x_i \right).$$

Cette application est continue et on a $f^{-1}(A)$ est un ouvert. Par ailleurs, $f(x) = x_1$ donc $x \in f^{-1}(A)$. Enfin, on a

$$f^{-1}(A) = \{z \in E / \exists y \in A, z = \lambda_1 y + \sum_{i=2}^n \lambda_i x_i\} \subset \operatorname{conv}(A).$$

Donc $f^{-1}(A)$ est un voisinage de x dans $\operatorname{conv}(A)$. □

Pour conclure, montrons deux résultats de stabilité lié à la topologie.

Proposition 18. *Soit $A \subset E$ un convexe, alors $\bar{A}$ est convexe. De plus si E est un evn, alors $\mathring{A}$ est convexe.*

Démonstration. $\bullet \bar{A}$. Soit $x, y \in \bar{A}$ et $\lambda \in [0,1]$, il existe $(x_n)_n, (y_n)_n$ dans A qui convergent vers x et y. On pose $z_n = \lambda x_n + (1-\lambda) y_n$. Par convexité de A, $z_n \in A$. De plus, z_n converge vers $z = \lambda x + (1-\lambda) y$. Donc $z \in \bar{A}$ qui est ainsi fermé.

$\bullet \mathring{A}$. Raisonnons par l'absurde. Supposons qu'il existe $x_0, y_0 \in \mathring{A}$ et $\lambda_0 \in [0,1]$ tels que $z_0 = \lambda_0 x_0 + (1-\lambda_0) y_0 \notin \mathring{A}$. Remarquons que $\lambda_0 \notin \{0,1\}$. Comme $y_0 \in \mathring{A}$ alors il existe $r > 0$ tel que $B(y_0, r) \subset A$. Alors l'application $f : z \in E \mapsto (z - \lambda_0 x_0)/(1-\lambda_0)$ est continue et vérifie $f(z_0) = y_0$. Ainsi, il existe $\delta > 0$ tel que $f(B(z_0, \delta)) \subset B(y_0, r) \subset A$. Mais comme $z_0 \notin \mathring{A}$ alors il existe $z_1 \in B(z_0, \delta) \cap A^c$. En notant $y_1 = f(z_1)$, on a $y_1 \in A$ et $z_1 \notin A$. Or $z_1 = \lambda_0 x_0 + (1-\lambda_0) y_1$. C'est absurde par convexité de A. □

1.2 Convexité dans les espaces de Hilbert

On suppose dans cette partie que $E = H$ unespae de Hilbert, muni du produit scalaire $\langle \cdot, \cdot \rangle$.

1.2.1 Théorème de projection

Theorem 19 (Projection sur un convexe fermé)**.** *Soit $C \subset H$ un convexe fermé non-vide. Alors pour tout $x \in H$ il existe un unique $y^* \in C$ tel que $\|x - y^*\| = \min_{y \in C} \|x - y\|$. Ce y^* est appelé projeté orthogonal de x sur C. Par unicité, cela définit un application de projection $\pi_C : x \in H \mapsto y^* \in C$. Enfin, $\pi_C(x)$ est caractérisée par les deux propriétés suivantes :*

(i) $\pi_C(x) \in C$

(ii) pour tout $y \in C$, on a $\langle x - \pi_C(x), y - \pi_C(x) \rangle \leq 0$.

La propriété (ii) peut être interprétée géométriquement en disant que, pour tout $y \in C$, l'angle entre $x - \pi_C(x)$ et $y - \pi_C(x)$ est obtus. En effet, cet angle est par définition

$$\theta = \arccos\left(\frac{\langle x - \pi_C(x), y - \pi_C(x) \rangle}{\|x - \pi_C(x)\| \|y - \pi_C(x)\|} \right) \in \left[\frac{\pi}{2}, \pi \right].$$

Démonstration. L'idée de cette démonstration est de résoudre un problème d'optimisation. Cela donne un avant-goût de de la suite du cours. Tout d'abord, si $x \in C$, on définit $\pi_C(x) = x$. Supposons $x \notin C$. On définit alors la fonction objectif suivante : $f : y \in H \mapsto \|x - y\| \in \mathbb{R}$. On souhaite minimiser f sur C. Pour commencer, f est continue (et même 1-Lipschitzienne), en utilisant de la seconde inégalité triangulaire. De plus, f est positive. Soit $(y_n)_n \in C^{\mathbb{N}}$ une suite minimisante, c'est à dire telle que $f(y_n) \to \inf_{y \in C} f(y) \geq 0$. Une telle suite existe par définition de l'infimum. Notons $\delta = \inf_{y \in C} f(y)$. On commence par remarquer que $\delta > 0$. En effet, $\delta = 0$ ssi $x \in \bar{C} = C$ ce qui a été exclu. Pour conclure à l'existence de y^* il suffit de montrer que la suite $(y_n)_n$ converge. Pour ce faire, on va montrer qu'elle est de Cauchy. Le sous-ensemble C étant fermé dans l'espace complet H, alors il est lui-même complet et cela suffira à montrer la convergence. Pour montrer que la suite est de Cauchy, on utilise l'identité du parallélogramme[1]. Pour tout $p, q \in \mathbb{N}$, on a

$$\|(x - y_p) + (x - y_q)\|^2 + \|y_p - y_q\|^2 = 2\|x - y_p\|^2 + 2\|x - y^q\|^2.$$

Or par convexité, $(y_p + y_q)/2 \in C$ et ainsi $\|(x - y_p) + (x - y_q)\|^2 \geq 4\delta^2$. On a donc

$$0 \leq \|y_p - y_q\|^2 \leq 2\left(\|x - y_p\|^2 + \|x - y_q\| - 2\delta^2 \right) \to 0.$$

Ce qui montre que la suite $(y_n)_n$ est de Cauchy et donc converge vers $y^* \in C$. Par continuité de f, on a $f(y^*) = \delta = \min_{y \in C} f(y)$. Ce qui prouve l'existence de y^*.

Pour montre l'unicité d'un tel y^*, supposons qu'il existe un autre $z^* \in C$ tel que $\|x - z^*\| = \delta$. La suite $(z_n)_n$ qui alterne entre y^* etz* est une suite minimisante. Elle est donc convergente d'après ce qui précède. Ainsi, $z^* = y^*$.

1. cette identité est équivalente au fait d'être issu d'un produit scalaire

Montrons à présent que la projection satisfait les points (i) et (ii). Par construction, (i) est clair. Soit $y \in C$, on a

$$\begin{aligned}\|x-\pi_C(x)\|^2 &\leq \|x-y\|^2 \\ &= \|(x-\pi_C(x))-(y-\pi_C(x))\|^2 \\ &= \|(x-\pi_C(x))\|^2+\|(y-\pi_C(x))\|^2-2\langle x-\pi_C(x), y-\pi_C(x)\rangle\end{aligned}$$

Ainsi, pour tout $y \in C$, on a $\|(y-\pi_C(x))\|^2 \geq 2\langle x-\pi_C(x), y-\pi_C(x)\rangle$. Fixons $y_0 \in C$. L'inégalité est vraie pour $y_\lambda = \lambda y_0 + (1-\lambda)\pi_C(x)$, pour tout $\lambda \in [0,1]$. Ainsi,

$$\lambda^2\|(y_0-\pi_C(x))\|^2 \geq 2\lambda\langle x-\pi_C(x), y_0-\pi_C(x)\rangle.$$

En simplifiant par λ et faisant tendre vers 0, on montre que π_C satisfait (ii).

Pour conclure, montrons que $\pi_C(x)$ est l'unique vecteur satisfaisant les deux conditions. Soit $z \in C$ tel que pour tout $y \in C$ on ait $\langle y-z, x-z\rangle \leq 0$. On a alors,

$$\begin{aligned}\|y-x\|^2 = \|(y-z)-(x-z)\|^2 &= \|y-z\|^2+\|x-z\|^2-2\langle y-z, x-z\rangle \\ &\geq \|y-z\|^2+\|x-z\|^2 \geq \|x-z\|^2\end{aligned}$$

Ceci montre que z vérifie $\|x-z\| = \min_{y\in C}\|x-y\|$ et par unicité de ce minimum, $z = \pi_C(x)$. □

Corollaire 20. *Soit $C \subset H$ un convexe fermé non vide. Alors l'application π_C est 1-lipschitzienne.*

Démonstration. Soient $x, y \in H$, on a

$$\begin{aligned}\|\pi_C(x)-\pi_C(y)\|^2 =& \langle \pi_C(x)-\pi_C(y), \pi_C(x)-\pi_C(y)\rangle \\ =& \underbrace{\langle \pi_C(x)-\pi_C(y), y-\pi_C(y)\rangle}_{\leq 0} + \langle \pi_C(x)-\pi_C(y), x-y\rangle \\ &+ \underbrace{\langle \pi_C(x)-\pi_C(y), \pi_C(x)-x\rangle}_{\leq 0} \\ \leq& \|\pi_C(x)-\pi_C(y)\|\|x-y\|\end{aligned}$$

D'où le résultat. □

1.2.2 Séparation et Hahn-Banach

On va utiliser le théorème de projection pour obtenir des résultats de séparation par de hyperplans dans les Hilbert et une preuve simple du théorème de Hahn-Banach. Notons que ce théorème s'énonce de manière générale dans des espace vectoriel topologique[2] mais se démonstration demande alors l'utilisation de l'axiome du choix (Cf [3])

Définition 21. *(i) Un hyperplan de H est $H(v,\alpha) = \{x \in H / \langle v, x\rangle = \alpha\}$, pour $v \in H$ et $\alpha \in \mathbb{R}$.*

(ii) L'hyperplan $H(v,\alpha)$ divise H en deux régions :

$$H_+(v,\alpha) = \{x \in H / \langle v, x\rangle \geq \alpha\} \quad et \quad H_-(v,\alpha) = \{x \in H / \langle v, x\rangle \leq \alpha\}.$$

(iii) Soient $A, B \subset H$. On dit que$H(v,\alpha)$ sépare A et B si $A \subset H_\pm(v,\alpha)$ et $B \subset H_\mp(v,\alpha)$.

(iv) On dit que$H(v,\alpha)$ sépare strictement A et B si

$$\mathbb{A} \subset H_\pm(v,\alpha) \setminus H(v,\alpha) \quad et \quad B \subset H_\mp(v,\alpha) \setminus H(v,\alpha).$$

(v) On dit que $H(v,\alpha)$ est un hyperplan d'appui pour A, s'il existe $x \in \bar{A} \cap H(v,\alpha)$ et $A \subset H_\pm(v,\alpha)$.

2. et pas dans des Banach!!!

La définition classique d'un hyperplan, comme noyau d'une forme linaire continue, correspond au cas $\alpha = 0$ (d'après le théorème de représentation de Riesz). La définition présente permet d'élargir à des hyperplans affines. Le vecteur v est appelé vecteur normal à l'hyperplan $H(v, \alpha)$ car pour tout $x, y \in H(v, \alpha)$, on a $\langle v, x - y \rangle = 0$.

Theorem 22 (Séparation). *Soit $A \subsetneq H$ un sous-ensemble strict de H, convexe, fermé, non-vide. Alors :*

(i) Pour tout $x \notin A$, il existe un hyperplan qui sépare strictement $\{x\}$ et A.

(ii) A admet des hyperplans d'appui.

Démonstration. (i) On pose $v = x - \pi_A(x)$ et $\delta = \|v\| > 0$. On a

$$\begin{aligned}\langle v, x \rangle =& \langle x - \pi_A(x), x - \pi_A(x) \rangle + \langle x - \pi_A(x), \pi_A(x) \rangle \\ =& \delta^2 + \langle x - \pi_A(x), \pi_A(x) \rangle \\ >& \langle x - \pi_A(x), \pi_A(x) \rangle + \frac{1}{2}\delta^2\end{aligned}$$

Ainsi, en posant, $\alpha = \langle x - \pi_A(x), \pi_A(x) \rangle + \frac{1}{2}\delta^2$, on a $\{x\} \subset H_+(v, \alpha) \setminus H(v, \alpha)$. Par ailleurs, pour tout $y \in A$, on a

$$\begin{aligned}\langle v, y \rangle =& \langle y - \pi_A(x), x - \pi_A(x) \rangle + \langle \pi_A(x), x - \pi(A)(x) \rangle \\ \leq& \alpha - \frac{1}{2}\delta^2\end{aligned}$$

D'où $A \subset H_-(v, \alpha) \setminus H(v, \alpha)$.

(ii). On choisit un $x \notin A$ arbitraire et on pose $v = x - \pi_A(x)$ et $\alpha = \langle x - \pi_A(x), \pi_A(x) \rangle$. D'après les calculs précédents, cela convient. □

Theorem 23 (Hahn-Banach géométrique). *Si $A \subset H$ est un convexe fermé non-vide et $B \subset H$ est un convexe compact non-vide, disjoint de A, alors il existe un hyperplan qui sépare strictement A et B.*

Démonstration. On définit $f : x \in B \mapsto \|x - \pi_A(x)\|$. Cette fonction est continue et par compacité de B, il existe $x^* \in B$ qui réalise son minimum. On pose $v = x^* - \pi_A(x^*)$, $\delta = \|v\|$ et $\alpha = \langle x - \pi_A(x), \pi_A(x) \rangle$. D'après le théorème précédent, $H(v, \alpha)$ sépare strictement $\{x^*\}$ et A. Il reste à montrer que $B \subset H_+(v, \alpha) \setminus H(v, \alpha)$. On pose $y^* = \pi_A(x^*)$. y^* admet une projection sur B et on a

$$\begin{aligned}\|\pi_B(y^*) - y^*\| =& \|\pi_B(y^*) - \pi_A(x^*)\| \\ &= \min_{z \in B} \|z - \pi_A(x^*)\| \leq \|x^* - \pi_A(x^*)\| \\ &= \min_{x \in B} \|x - \pi_A(x)\| \leq \|\pi_B(y^*) - \pi_A(\pi_B(y^*))\| \\ &= \min_{y \in A} \|\pi_B(y^*) - y\| \leq \|\pi_B(y^*) - y^*\|\end{aligned}$$

On a donc que des égalités et par unicité du minimum dans la projection, on a $x^* = \pi_B(y^*)$. Ainsi, on a $v = \pi_B(y^*) - y^*$ et pour tout $x \in B$

$$\langle x, v \rangle = -\langle x, y^* - \pi_B(y^*) \rangle \geq \alpha + \frac{1}{2}\delta^2.$$

□

Application 24 (Enveloppe convexe fermée [2]). *L'enveloppe convexe ne préserve pas le caractère fermé. On définit donc l'enveloppe convexe fermée de $A \subset H$, $\overline{\text{conv}}(A)$ comme étant le plus petit convexe fermé contenant A :*

$$\overline{\text{conv}}(A) = \bigcap_{C \text{ cvx fermé}, A \subset C} C.$$

A partir du théorème de Hahn-Banach, on montre que

$$\overline{\text{conv}}(A) = \bigcap_{(v,\alpha), A \subset H(v,\alpha)} H(v\alpha).$$

En particulier, on peut utiliser ce résultat pour déterminer l'enveloppe convexe de $\mathcal{O}_n(\mathbb{R})$.

1.2.3 Sous-espace orthogonal

Le théorème de projection permet aussi de démontrer la décomposition d'un espace de Hibert en somme orthogonale.

Définition 25. *Si $F \subset H$ sev, son orthogonal est défini par*

$$F^\perp = \{x \in H / \forall y \in H, \langle x, y \rangle = 0\}.$$

Proposition 26. *(i) $F^\perp$ est fermé*

(ii) $F^\perp = (\overline{F})^\perp$

(iii) $\overline{F} \subset (F^\perp)^\perp$

Démonstration. (i) Si $(x_n)_n$ est une suite de $F^\perp$ qui converge vers $x \in H$ alors pour tout $y \in F$, on a $\langle x, y \rangle = \lim \langle x_n, y \rangle = 0$ par continuité du produit scalaire. Donc $x \in F^\perp$ qui est fermé.

(ii) Soit $x \in \overline{F^\perp}$, alors x est limite d'une suite $(x_n)_n$ de $F^\perp$. Soit $y \in \overline{F}$, y est limite d'une suite $(y_n)_n$ de F. Par continuité, on a $\langle x, y \rangle = \lim \langle x_p, y_q \rangle = 0$. Donc $F^\perp = \overline{F^\perp} \subset \overline{F}^\perp$. De plus, puisque $F \subset \overline{F}$ alors $\overline{F}^\perp \subset F^\perp$. D'où l'égalité.

(iii) Soit $x \in F$ alors pour tout $y \in F^\perp$, on a $\langle x, y \rangle = 0$. Donc $x \in (F^\perp)^\perp$. De plus, $(F^\perp)^\perp$ est fermé, donc $\overline{F} \subset (F^\perp)^\perp$. □

Corollaire 27. *Soit $F \subset H$ un sev fermé, alors π_F est linéaire, continue, surjective. De plus, on a la décomposition $H = F \oplus F^\perp$. En particulier, $F = (F^\perp)^\perp$.*

Démonstration. π_F est surjective par construction et continue d'après un résultat précédant. Pour tout $x \in H$, on a la décomposition $x = \pi_F(x) + (x - \pi_F(x))$. Il faut donc vérifier que $(x - \pi_F(x)) \in F^\perp$ pour avoir $H = F + F^\perp$. Pour tout $z \in F$, on a $\langle z - \pi_F(x), x - \pi_F(x) \rangle \leq 0$. En appliquant à λz et en divisant par λ, on obtient

$$0 \geq \langle z - \frac{1}{\lambda}\pi_F(x), x - \pi_F(x) \rangle, \forall \lambda > 0$$
$$0 \leq \langle z - \frac{1}{\lambda}\pi_F(x), x - \pi_F(x) \rangle, \forall \lambda < 0$$

En faisant tendre $|\lambda|$ vers $+\infty$, on en déduit que $\langle z, x - \pi_F(x) \rangle = 0$. Donc $(x - \pi_F(x)) \in F^\perp$. Supposons qu'il y ait une seconde décomposition $x = x_1 + x_2$ avec $x_1 \in F$ et $x_2 \in F^\perp$. Alors on a $x_1 - \pi_F(x) = (x - \pi_F(x)) - x_2$. Ainsi, $x_1 - \pi_F(x) \in F \cap F^\perp = \{0\}$. On en déduit l'unicité de la décomposition. En particulier, cela implique que $F = (F^\perp)^\perp$.

La linéarité de π_F découle cette unique décomposition. En effet, pour tout $x, y \in H$ et $\alpha, \beta \in \mathbb{C}$, on a

$$\begin{aligned} \alpha x + \beta y &= \underbrace{\pi_F(\alpha x) + \pi_F(\beta y)}_{\in F} + \underbrace{\alpha x + \beta y - (\pi_F(\alpha x) + \pi_F(\beta y))}_{\in F^\perp} \\ &= \underbrace{\pi_F(\alpha x + \beta y)}_{\in F} + \underbrace{\alpha x + \beta y - \pi_F(\alpha x + \beta y)}_{\in F^\perp} \end{aligned}$$

□

Application 28 (Densité)**.** *Ce corollaire est souvent utiliser pour montrer qu'une famille orthonormée est une base hilbertienne (i.e est totale). En effet, pour un sev F quelconque, on a montrer que* $(F^\perp)^\perp = \overline{F}$. *Ainsi* $F \subset H$ *est dense ssi* $F^\perp = \{0\}$. *on retrouvera ce critère dans l'étude des séries de Fourier, des polynômes orthogonaux, l'espace de Bergmann etc.*

Application 29 (Théorème ergodique de Von Neumann, [2], Exo 3.6)**.** *Soit* $T \in \mathcal{L}(H)$, $\|T\| \leq 1$. *Alors pour tout* $x \in H$, *on a*

$$\frac{1}{n}\sum_{n=0}^{n-1} T^n(x) \longrightarrow \pi_{\ker(T-\mathrm{id}}(x).$$

Application 30 (Espérance conditionnelle, [2], App 3.23)**.** *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité et* $\mathcal{G}$ *une sous-tribu de* $\mathcal{F}$. *L'espérance conditionnelle par rapport à* $\mathcal{G}$ *se définit sur* $\mathbb{L}^2(\mathcal{F})$ *comme la projection sur le sous-espace* $\mathbb{L}^2(\mathcal{G})$.

1.2.4 Semi-cône et lemme de Farkas

La décomposition en sous-espace orthogonaux se généralise à des sous-ensemble F qui ont une structure plus faible que celle de sev : les semi-cônes.

Définition 31. *(i) Un sous ensemble* $K \subset H$ *est un semi-cône si pour tout* $x \in K$ *et tout* $\lambda \geq 0$ *on a* $\lambda x \in K$.

(ii) Pour un sous-ensemble $K \subset H$, *on définit son semi-cône polaire* K° *par*

$$K^\circ = \{x \in H / \forall y \in K,\ \langle x, y\rangle \leq 0\}.$$

Remarquons que si K est un semi-cône, alors $0 \in K$. Dans le cas "plat" des sev, on a $F^\circ = F^\perp$.

Lemme 32. *Si K est un semi-cône convexe fermé alors* $K^\circ = \{x \in H / \pi_K(x) = 0\}$.

Démonstration. Soit $x \in K^\circ$ alors pour tout $y \in K$, on a $\langle x, y\rangle \leq 0$. Or $0 \in K$ et $\langle x - 0, y - 0\rangle \leq 0$. Donc par caractérisation de la projection, $\pi_K(x) = 0$. Réciproquement, si $\pi_K(x) = 0$ alors pour tout $y \in K$, on a $0 \geq \langle x - 0, y - 0\rangle$. Donc $x \in K^\circ$. □

Le théorème suivant donne une décomposition de H en somme de semi-cônes, similaire à la décomposition orthogonale 27.

Theorem 33 (Moreau, [2])**.** *Soit K un semi-cône convexe fermé non-vide. Alors pour tout* $x \in H$, *le projeté* $\pi_K(x)$ *admet caractérisation suivante*

(i) $\pi_K(x) \in K$

(ii) $x - \pi_K(x) \in K^\circ$ *et* $\langle x - \pi_K(x), \pi_K(x)\rangle = 0$.

De plus, on a l'équivalence entre les propriétés suivantes

(a) x admet une décomposition $x = x_K + x_\circ$ *avec* $x_K \in K$, $x_\circ \in K^\circ$ *et* $\langle x_K, x_\circ\rangle = 0$

(b) $x_k = \pi_K(x)$ *et* $x_\circ = \pi_{K^\circ}(x)$.

Corollaire 34. *Soit K un semi-cône connexe non-vide. Alors* $\overline{K} = (K^\circ)^\circ$.

Application 35 (Lemme de Farkas)**.** *Soient* $u_1, \ldots, u_n \in \mathbb{R}^d$ *et* $v \in \mathbb{R}^d$. *Alors on a équivalence entre*

(i) $\{x \in \mathbb{R}^d / \langle x, u_i\rangle \leq 0 \forall i\} \subset \{x \in \mathbb{R}^d / \langle x, v\rangle \leq 0\}$

(ii) $v \in \{\sum_i \lambda_i u_i / \lambda_i \geq 0\}$.

Démonstration. Soit $K = \{\sum_i \lambda_i u_i / \lambda_i \geq 0\}$. C'est un cône convexe fermé[3] non-vide. On sait alors que $K = (K^\circ)^\circ$. Ainsi, $v \in K$ ssi pour tout $x \in K^\circ$, $\langle v, x\rangle \leq 0$, c'est à dire, pour tout x tel que pour tout $\lambda_i \geq 0$, $\langle x, \sum \lambda_i u_i\rangle \leq 0$, $\langle x, v\rangle \leq 0$. C'est équivalent à ce que pour tout x tel que $\langle x, u_i\rangle \leq 0$ pour tout i, on a $\langle x, v\rangle \leq 0$. □

La traduction dans le cadre plat du lemme de Farkas se lit :

$$\bigcap u_i^\perp \subset v^\perp \Leftrightarrow v \in Vectu_{i_i}.$$

3. Évident si les u_i sont libres. Montrer que l'on peut s'y ramener sinon

1.3 Convexité et convergence faible

Définition 36. *Une suite $(x_n)_n$ dans H converge faiblement vers x^* si pour tout $y \in H$ on a $\langle x_n, y\rangle \to \langle x^*, y\rangle$. On note $x_n \rightharpoonup x^*$.*

Proposition 37. *(i) Si $x_n \to x^*$ alors $x_n \rightharpoonup x^*$, i.e si A est faiblement fermé, alors A est fortement fermé.*

(ii) Si $x_n \rightharpoonup x^*$, alors $\|x^*\| \leq \liminf \|x_n\|$

(iii) Si $x_n \rightharpoonup x^*$, alors $x_n \to x^*$ ssi $\|x_n\| \to \|x^*\|$

(iv) Si $(x_n)_n$ est bornée alors elle admet une sous-suite faiblement convergente.

Theorem 38. *Soit $K \subset H$ convexe non-vide. Alors K est faiblement fermé ssi K fortement fermé.*

Démonstration. Soit K fermé et supposons par l'absurde qu'il n'est pas faiblement fermé. On dispose donc d'une suite $(x_n)_n \in K$ tel que $x_n \rightharpoonup x^*$ et $x^* \notin K$. Puisque K est convexe fermé non-vide, on a $\delta = \|x^* - \pi_K(x^*)\| > 0$. Par caractérisation du projeté, pour tout $y \in K$, on a $\langle y - \pi_K(x^*), x^* - \pi_K(x^*)\rangle \leq 0$. En particulier, pour $y = x_n$, on a

$$0 \leq \langle x_n - \pi_K(x^*), x^* - \pi_K(x^*)\rangle \to \langle x^* - \pi_K(x^*), x^* - \pi_K(x^*)\rangle = \delta^2 > 0.$$

C'est absurde. $\square$

Chapitre 2

Fonctions convexes

La seconde face de la convexité concerne les fonctions à valeurs réelles. La notion de convexité pour une fonction permet d'améliorer significativement les résultats classiques d'optimisation. Elle sera cruciale dans la section suivante.

2.1 Définition et premières propriétés

Définition 39. *Soit $K \subset E$ convexe et $f : K \to \mathbb{R}$.*

(i) f est convexe si pour tout $x, y \in K$ et tout $\lambda \in [0,1]$, on a

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

(ii) f est strictement convexe si pour tout $x \neq y \in K$ et tout $\lambda \in]0,1[$, on a

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

(iii) Si de plus E est un espace de Hilbert, f est α-convexe pour $\alpha > 0$ si pour tout $x, y \in K$ et tout $\lambda \in [0,1]$, on a

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{\alpha}{2}\lambda(1-\lambda)\|x-y\|^2.$$

(iv) f est concave (resp. strictemnt concave, resp. α-concave) si $-f$ est convexe (resp.. . .).

On a clairement l'implication α-convexe $\Rightarrow$ strictement convexe $\Rightarrow$ convexe.

Exemple 40. *(a) Toute forme linéaire est convexe*

(b) $x \in \mathbb{R} \mapsto e^x$ est str convexe. $x \in \mathbb{R} \mapsto x^2$ est 2-convexe.

Proposition 41. *Soit $K \subset E$ convexe et $f : K \to \mathbb{R}$.*

(i) f est convexe ssi pour tout $x, y \in K, \varphi : t \in [0,1] \mapsto f(tx + (1-t)y)$ est convexe.

(ii) (Jensen) f est convexe ssi pour tout $n \geq 2$ et tout combinaison convexe de K $\sum_{i=1}^n \lambda_i x_i$, on a $f(\sum_{i=1}^n \lambda_i x_i) \leq \sum_{i=1}^n \lambda_i f(x_i)$.

(iii) Si $K = \mathrm{conv}(\{x_1, \ldots, x_n\}$ et f convexe, alors $\max_{x \in K} f(x) \leq \max f(x_i)$.

Remarquons que la propriété (ii) est une version discrète et finie de l'inégalité de Jensen.

Application 42 (Inégalité de convexité). *La convexité est un outil très puissant pour obtenir des inégalités. Voici quelques exemples.*

(i) (Inégalité arithmético-géométrique) Pour tout $x_1, \ldots, x_n \geq 0$, on a

$$\left(\prod_{i=1}^n x_i\right)^{1/n} \leq \frac{1}{n}\sum_{i=1}^n x_i.$$

(ii) *(Inégalité de Young) Pour tout $x, y \geq 0$ et $1/p + 1/q = 1$, on a*

$$xy \leq \frac{1}{p}x + \frac{1}{q}y.$$

(iii) *(Inégalité de Hölder) Pour $x_1, \ldots, x_n, y_1, \ldots, y_n$ et $1/p + 1/q = 1$, on a*

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum x_i^p\right)^{1/p} \left(\sum y_i^q\right)^{1/q}.$$

(iv) *Pour tout $x, y > 1$, $\log\left(\frac{x+y}{2}\right) \geq \sqrt{\log(x)\log(y)}$.*

(v) *(Inégalité de kantorovitch) Pour tout $x \in \mathbb{R}^d$, on a*

$$\|x\|^4 \leq \langle Ax, x\rangle\langle A^{-1}x, x\rangle \leq \frac{1}{4}\left(\kappa(A)^{1/2} + \kappa(A)^{-1/2}\right)^2 \|x\|^4.$$

On dispose des propriétés de stabilité suivantes.

Proposition 43. *Soient $f, g : K \to \mathbb{R}$ convexes, $f : \mathbb{R} \to \mathbb{R}$ croissante et $\alpha, \beta \geq 0$.*

(i) *$\alpha f + \beta g$ est convexe.*

(ii) *$h \circ f$ est convexe.*

Contre-exemple 44. *La convexité n'est pas préservée par composition : $x \mapsto -x$ est convexe et f et $-f$ ne sont pas simultanément convexes (à moins d'être affines).*

Proposition 45. *Soit $(f_i)_{i\in I}$ une famille de fonctions convexes sur K, telle que la fonction $g : x \in K \mapsto \sup_i f_i(x)$ soit bien définie*[1]*. Alors elle est convexe.*

2.2 Régularité des fonctions convexes

Commençons par remarquer qu'en dimension infinie la convexité ne garantie aucune régularité. En effet, n'importe quelle forme linéaire est convexe mais certaine ne sont pas continue. On se placera donc en dimension finie, sur $E = \mathbb{R}^n$. On commence par traiter le cas de la dimension 1.

Définition 46. *Soit $I \subset \mathbb{R}$ et $f : I \to \mathbb{R}$. Pour tout $x_0 \in I$, on définit la fonction pente*

$$p_{x_0} : y \in I \setminus \{x_0\} \mapsto \frac{f(y) - f(x_0)}{y - x_0}.$$

Theorem 47. *$f : I \to \mathbb{R}$ est convexe ssi pour tout $x_0 \in I$ la fonction pente p_{x_0} est croissante.*

Corollaire 48. *Soit $f : I \to \mathbb{R}$ convexe. Alors f admet en tout point $x \in \mathring{I}$ des dérivées à gauche et à droite satisfaisant $f'_g(x) \leq f'_d(x)$. En particulier, f est continue sur $\mathring{I}$.*

Application 49 (Inégalité de Jensen)**.** *Soit $\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, X une variable aléatoire réelle et $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction convexe telles que $\varphi(X)\mathbb{L}^1$. Alors*

$$\varphi\left(\mathbb{E}[X]\right) \leq \mathbb{E}[\varphi(X)].$$

Si de plus φ est strictement convexe alors il y a égalité ssi X est constante ps.

Démonstration. Soit $\ell \in [\varphi'_g(\mathbb{E}[X]), \varphi'_d(\mathbb{E}[X])]$, alors on $\varphi(X) \geq \varphi(\mathbb{E}[X]) + \ell(X - \mathbb{E}[X])$ p.s. En passant à l'espérance on obtient le résultat. □

On s'intéresse maintenant au cas de la dimension $n \geq 2$.

1. majoration uniforme par exemple

Theorem 50. *Soit $K \subset \mathbb{R}^d$ un ouvert convexe, et $f : K \to \mathbb{R}$ convexe. Alors f est continue.*

Démonstration. Soit $x_0 \in K$. puis que K est ouvert, il existe $r > 0$ tel que le cube $C = x_0 + [-r, r]^d \subset K$. On note $V = \{x_0 \pm e_i / 1 \leq i \leq d\}$ l'ensemble des sommets du cube et $\alpha = \max_{x \in V} f(x)$. On a $C = \mathrm{conv}(V)$ et f convexe donc $\max_{x \in C} f(x) \leq \alpha$. Soit $y \in C$ distinct de x_0 et on définit $y(t) = ty + (1-t)x_0$ pour tout $t \in \mathbb{R}$. Alors $t^* = r / \|y - x_0\|_\infty$ vérifie $y(\pm t^*) \in \partial C$. On exprime alors y et x_0 comme combinaison convexe de x_0 et $y(t^*)$ et y $y(-t^*)$ respectivement. Par convexité, on a

$$f(y) \leq \frac{1}{t^*} f(y(t^*)) + (1 - \frac{1}{t^*}) f(x_0), \quad f(x_0) \leq \frac{1}{1+t^*} f(y(-t^*)) + (1 - \frac{1}{1+t^*}) f(x_0).$$

On en déduit que

$$|f(y) - f(x_0)| \leq f(y) \leq \frac{1}{t^*} \left(|f(x_0) + \max |f(y(\pm t^*))| \right) \leq \frac{|f(x_0)| + \alpha}{r} \|y - x_0\|_\infty.$$

Ce qui montre la continuité en x_0. □

La démonstration précédente suggère un résultat plus fort (Cf [6]).

Theorem 51. *Soit $f : K \subset \mathbb{R}^d \mapsto \mathbb{R}$ convexe. Alors est est lipschitzienne sur tous les compacts de K. En particulier, f est dérivable presque partout.*

On peut même montre que la constante de Lipschitz ne dépend que du diamètre du compact et d'une borne de f sur ce compact : $L \leq 2M/\delta$. La deuxième partie du théorème est une conséquence du théorème de Rademacher.

Application 52 (Théorème "à la Dini" convexe)**.** *Soit une suite de fonctions $f_n : K \subset \mathbb{R}^n \to \mathbb{R}$ convexes qui converge simplement vers $f : K \to \mathbb{R}$. Alors f est convexe et la convergence est uniforme sur tout compact de K.*

2.3 Caractérisation de la convexité

On commence par une caractérisation général et plutôt abstraite.

Définition 53. *Soit $K \subset E$ et $f : K \to \mathbb{R}$. l'épigraphe de f est l'ensemble*

$$\mathrm{epi}(f) = \{(x, \alpha) \in K \times \mathbb{R} / \alpha \leq f(x)\}.$$

Proposition 54. *$f : K \to \mathbb{R}$ est convexe ssi $\mathrm{epi}(f)$ est convexe.*

Démonstration. $\Rightarrow$ On suppose f convexe. Soient $(x, \alpha), (y, \beta) \in \mathrm{epi}(f)$ et soit $\lambda \in [0, 1]$. Par convexité de f, on a

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

$\Leftarrow$ On suppose $\mathrm{epi}(f)$ convexe. Soit $x, y \in K$ et $\lambda \in [0, 1]$. Alors, $(x, f(x)$ et $(y, f(y) \in \mathrm{epi}(f)$ et par convexité $\lambda(x, f(x)) + (1-\lambda)(y, f(y))$ aussi. Donc f est convexe. □

On va maintenant donner des critères de convexité selon la régularité de la fonction. En dimension 1, en affinant le corollaire 48 sur la croissance des pentes, on a la caractérisation suivante.

Theorem 55. *Soit $f : I \subset \mathbb{R} \to \mathbb{R}$ de classe $\mathcal{C}^1$. Alors*

(i) f est convexe ssi f' est croissante.

(ii) f est strictement convexe ssi f' est strictement croissante.

(iii) f est α-convexe ssi pour tout $x \leq y$, $f'(x) \leq f'(y) - \alpha(y - x)$.

En dimension supérieure, la notion de "dérivée" croissante n'a pas de sens. Cependant, en dimension 1, f' croissante est équivalent à la propriété : pour tout $x, y \in I$, $(f'(x) - f'(y))(x - y) \geq 0$. Cette formulation a un sens en dimension supérieure.

Theorem 56. *Soit $f : K \subset \mathbb{R}^d \to \mathbb{R}$ différentiable. Alors*

(i) f est convexe ssi pour tout $x, y \in K$, $\langle df(x) - df(y), x - y\rangle \geq 0$.

(ii) f est strictement convexe ssi pour tout $x \neq y \in K$, $\langle df(x) - df(y), x - y\rangle > 0$.

(iii) f est α-convexe ssi pour tout $x, y \in K$, $\langle df(x) - df(y), x - y\rangle \geq \alpha\|x - y\|^2$.

La caractérisation de dimension 1 "f est au-dessus de ses tangentes" se généralise aussi en dimension supérieure.

Theorem 57. *Soit $f : K \subset \mathbb{R}^d \to \mathbb{R}$ différentiable. Alors*

(i) f est convexe ssi pour tout $x, y \in K$, $f(y) - f(x) \geq \langle df(x), y - x\rangle$.

(ii) f est strictement convexe ssi pour tout $x \neq y \in K$, $f(y) - f(x) > \langle df(x), y - x\rangle$.

(iii) f est α-convexe ssi pour tout $x, y \in K$, $f(y) - f(x) \geq \langle df(x), y - x\rangle + \frac{\alpha}{2}\|x - y\|^2$.

Démonstration. $(i) \Rightarrow$ On suppose f convexe. Soit $x, y \in K$ et $\lambda \in [0, 1]$. On pose $z = \lambda x + (1 - \lambda)y = x + (1 - \lambda)(y - x)$. Par convexité, on a

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

Par ailleurs, comme f est $\mathcal{C}^1$, on a

$$\frac{f(x + (1 - \lambda)(y - x)) - f(x)}{1 - \lambda} \to \langle df(x), y - x\rangle.$$

D'où le résultat.

$(i) \Leftarrow$ On applique l'inégalité entre x et z et entre y et z :

$$\begin{cases} f(x) - f(z) \geq \langle df(z), x - z\rangle \\ f(y) - f(z) \geq \langle df(z), y - z\rangle \end{cases}$$

En combinant, on obtient $\lambda f(x) + (1 - \lambda)f(y) - f(z) \geq 0$.

(iii) et $(ii) \Leftarrow$ Même démonstration. En revanche, pour $(ii) \Rightarrow$, il faut faire attention à le pas passer à la limite pour conserver l'inégalité stricte. On Suppose f strictement convexe. En utilisant le point (i) entre x et z, on a

$$f(y) - f(x) > \frac{1}{1 - \lambda}(f(z) - f(x)) \geq \langle df(x), \frac{z - x}{1 - \lambda}\rangle = \langle df(x), y - x\rangle.$$

□

En supposant une régularité $\mathcal{C}^2$, on obtient un caractérisation d'ordre 2 sur la hessienne. On note $\mathcal{S}_d^+(\mathbb{R})$ l'ensemble des matrices réelle symétriques positives et $\mathcal{S}_n^{++}(\mathbb{R})$ celui des matrice symétriques définies positive.

Proposition 58. *Soit $f : K \subset \mathbb{R}^d \to \mathbb{R}$ deux-fois différentiable. Alors*

(i) f est convexe ssi pour tout $x \in K$, $\mathrm{Hess}(f)(x) \in \mathcal{S}_d^+(\mathbb{R})$.

(ii) Si $\mathrm{Hess}(f)(x) \in \mathcal{S}_d^{++}(\mathbb{R})$ pour tout $x \in K$, alors f est strictement convexe.

(iii) f est α-convexe ssi pour tout $x \in K$, $\mathrm{Hess}(f)(x) \geq \alpha\,\mathrm{id}$ i.e pour tout $h \in \mathbb{R}^d$

$$\langle \mathrm{Hess}(f)(x)h, h\rangle \geq \alpha\|h\|^2.$$

Contre-exemple 59. *La fonction $x \in \mathbb{R} \mapsto x^4$ montre que le point (ii) n'est pas une équivalence.*

Démonstration. On va détailler seulement le (i). Les autres se démontre de manière similaire. $\Rightarrow$ On écrit la formule de Taylor-Young

$$f(y) - f(x) - \langle df(x), y - x\rangle = \frac{1}{2}\langle \mathrm{Hess}(f)(x)(y - x), y - x\rangle + o\left(\|y - x\|^2\right).$$

D'après la caractérisation d'ordre 1 de la convexité, le membre de gauche est positif. En particulier, pour tout $h \in \mathbb{R}^d$, pour tout $\lambda > 0$, en appliquant à $y = x + \lambda h / \|h\|$, on a

$$\langle \mathrm{Hess}(f)(x)h, h\rangle = \lim_{\lambda \to 0} \langle \mathrm{Hess}(f)(x)h, h\rangle + o(1) \geq 0.$$

$\Leftarrow$ On applique la formule de Taylor avec reste exacte : pour tout $x, y \in K$, il existe $z \in [x, y]$ tel que

$$f(y) - f(x) - \langle df(x), y - x\rangle = \frac{1}{2} \underbrace{\langle \mathrm{Hess}(f)(z)(y-x), y - x\rangle}_{\geq 0}.$$

D'après la caractérisation d'ordre 1, f est convexe. $\square$

Exemple 60 ($\star$ "La" fonctionnelle quadratique $\star$)**.** *Soit $A \in \mathcal{S}_d(\mathbb{R})$ et $b \in \mathbb{R}^d$. On définit la fonctionnelle quadratique*

$$f : x \in \mathbb{R}^d \mapsto \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle.$$

Alors f est convexe si et seulement si $A \in \mathcal{S}_d^+(\mathbb{R})$ et f est α-convexe si et seulement si le spectre de A est inclus dans $[\alpha, +\infty[$.

Les caractérisations 57 et 58 restent vraies dans un espace vectoriel normé de dimension quelconque (Cf [5]).

Chapitre 3

Optimisation

Le but de cette partie est d'étudier des problèmes de la forme : trouver $x^* \in K$ tel que

$$f(x^*) = \inf_{x \in K} f(x),$$

où $K \subset E$ est un sous-ensemble d'un evn E et $f : K \to \mathbb{R}$.

3.1 Définition

Définition 61. *On dit que $x^* \in K$ est un minimum (resp. maximum) global de f sur K si pour tout $x \in K$, on a $f(x^*) \leq f(x)$ (resp. $f(x^*) \geq f(x)$).*

(i) Maximiser f est équivalent à minimiser $-f$

(ii) Les extrema peuvent ne pas exister. $x \in \mathbb{R} \mapsto e^x$.

(iii) Un extremum peut ne pas être unique. $x \in \mathbb{R} \mapsto (x^2 - 1)^2$.

Définition 62. *On dit que x^* est un minimum local de f sur K si c'est un minimum global sur un voisinage de lui-même dans K i.e il existe $r > 0$ tel que pour tout $x \in K \cap B(x^*, r)$, $f(x^*) \leq f(x)$.*

3.2 Résultats généraux d'existence et unicité

On commence par quelques résultats de base.

Proposition 63. *Si $K \subset E$ est compact non-vide et* $: K \to \mathbb{R}$ *est continue alors f admet un minimum et un maximum global sur K.*

La démonstration repose sur la même stratégie que le Théorème de projection : prendre une suite minimisante et obtenir sa convergence par un argument idoine (structure hilbertienne pour le théorème de projection et compacité ici).

Démonstration. Soit $(x_n)_n$ une suite minimisante. Par compacité, elle admet au moins une valeur d'adhérence $x^* \in K$. Par continuité de f, on a

$$f(x^*) = \lim_k f(x_{n_k}) = \inf_{x \in K} f(x).$$

$\square$

En dimension finie, la caractérisation des compactes permet d'obtenir l'existence de minimum sous des hypothèses moins contraignantes.

Corollaire 64. *On suppose $E = \mathbb{R}^d$, $K \subset E$ fermé non-vide et $f : K \to \mathbb{R}$ continue. S'il existe $\bar{x} \in K$ tel que le sous-niveau $\{x \in K / f(x) \leq f(\bar{x})\}$ est borné, alors f admet un minimum global sur K.*

Démonstration. Par continuité, le sous-niveau, noté $K_{\bar{x}}$, est fermé. Par hypothèse, il est borné donc compact. D'après la proposition précédent, f admet un minimum global sur $K_{\bar{x}}$, x^*. De plus, si $x \in K \setminus K_{\bar{x}}$, alors on a $f(x) \geq f(\bar{x}) \geq f(x^*)$. Donc x^* est un minimum global sur K. □

Pour montre l'existence d'un maximum, il suffit d'exhiber un sur-niveau $\{x \in K / f(x) \geq f(\bar{x})\}$ borné.

Application 65 (Pavé de volume maximal)**.** *On veut montrer qu'il existe un pavé de $\mathbb{R}^3$ de surface $S > 0$ fixé, de volume maximal. Cela revient à maximiser la fonction $f : x \in (\mathbb{R}_+)^3 \mapsto x_1 x_2 x_3$ sur l'ensemble*

$$K = \{x \in (\mathbb{R}_+)^3 / 2(x_1 x_2 + x_1 x_3 + x_2 x_3) = S\}.$$

Fixons $\bar{x} \in K$ tel que $f(\bar{x}) > 0$ (il en existe). Soit $x \in K$ tel que $f(x) \geq f(\bar{x})$, en particulier, $x_i > 0$ pour tout i. De plus pour tout $i \neq j$, $x_i x_j \leq S/2$ donc $f(\bar{x}) \leq f(x) \leq S^2/(4x_i)$. Donc $x_i \leq S^2/(4f(\bar{x})) := R$. Ainsi, le sur-niveau $\{x \in K / f(x) \geq f(\bar{x})\}$ est inclus dans $[0, R]^3$. Donc f admet un maximum global sur K.

Il existe un critère simple garantissant l'existence de sous-niveau borné : la coercivité.

Définition 66. *Soit $K \subset E$, $f : K \to \mathbb{R}$ est coercive si pour tout $M > 0$, il existe $R > 0$ tel que $x \in K \cap B(0, R)^c \Rightarrow f(x) > M$.*

Autrement dit, f explose.

Corollaire 67. *Soit $E = \mathbb{R}^d$, $K \subset E$ fermé non-vide et $f : K \to R$ continue coercive. Alors f admet un minimum global sur K.*

Démonstration. Soit $\bar{x} \in K$. En prenant $M = f(\bar{x})$, il existe $R > 0$ tel que le sous-niveau associé à $\bar{x}$ soit inclus dans $B(0, R)$. On peut alors conclure avec la proposition précédente. □

3.3 Optimisation et convexité

La convexité permet d'obtenir des résultats d'unicité et d'existence, y compris en dimension infinie.

Proposition 68 (Passage du local au global)**.** *Soit E un evn, $K \subset E$ convexe et $f : K \to \mathbb{R}$ convexe. Si x^* est un minimum local de f sur K alors, c'est un minimum global.*

Démonstration. Par définition, il existe $r > 0$ tel que pour tout $z \in K \cap B(x^*, r)$ $f(x^*) \leq f(z)$. Soit $y \in K$ alors pour tout $\lambda \in [0, 1]$, $z_\lambda = \lambda y + (1 - \lambda)x^* \in K$ et pour λ suffisamment petit, $z_\lambda \in B(x^*, r)$. Par convexité, on a

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

Ce qui implique $f(x^*) \leq f(y)$ pour tout $y \in K$. □

Proposition 69. *Soit E un evn, $K \subset E$ convexe et $f : K \to \mathbb{R}$ strictement convexe. Alors f admet au plus un minimum.*

Démonstration. Si $x^* \neq y^*$ sont deux minima globaux sur K alors par stricte convexité, on a

$$f\left(\frac{x^* + y^*}{2}\right) < \frac{1}{2} f(x^*) + \frac{1}{2} f(y^*) = \min_{x \in K} f(x).$$

Ce qui est absurde □

Contre-exemple 70. *Ce résultat ne dit pas que la stricte convexité, seule, implique l'existence d'un minimum ($x \in \mathbb{R} \mapsto e^x$).*

En dimension fini, l'α-convexité assure l'existence d'un minimum.

Proposition 71. *Soit $E = \mathbb{R}^d$, $K \subset E$ convexe fermé non-vide et $f : K \to R$ différentiable α-convexe. Alors f admet un unique minimum global sur K.*

Démonstration. Soit $\bar{x} \in K$ quelconque. Alors pour tout $x \in K$, on a

$$\begin{aligned} f(x) \geq & f(\bar{x}) + \langle df(\bar{x}), x - x\bar{x} \rangle + \frac{\alpha}{2}\|x - \bar{x}\|^2 \\ \geq & f(\bar{x}) - \|df(\bar{x})\| \|x - \bar{x}\| + + \frac{\alpha}{2}\|x - \bar{x}\|^2 \\ \longrightarrow & + \infty \end{aligned}$$

Donc f est coercive. f admet donc un minimum, qui est unique par stricte convexité. □

En dimension infinie, une manière (classique) d'obtenir l'existence de minimiseur (entre autre) est de remplacer la convergence forte par de la convergence faible, c'est à dire, ajouter des compactes. Dans ce cas, on peut aussi relaxer l'hypothèse de continuité.

Définition 72. *Soit E un evn et $f : E \to \mathbb{R}$. f est semi-continue inférieurement (sci) si pour tout $x \in E$ et $\varepsilon > 0$ il existe un voisinage $\mathcal{B}_x$ de x tel que pour tout $y \in \mathcal{B}_x$, $f(y) \geq f(x) - \varepsilon$.*

On a la caractérisation suivante de la sci.

Proposition 73 ([3])**.** *f sci $\Leftrightarrow$ $\mathrm{epi}(f)$ fermé $\Leftrightarrow$ pour tout $x_n \to x$, $f(x) \leq \liminf_n f(x_n)$.*

Proposition 74. *Soit E un espace de Hilbert et $K \subset E$ convexe fermé non-vide. Soit $f : K \to \mathbb{R}$ convexe sci coercive. Alors f admet un minimum global sur K.*

Démonstration. En utilisant le coercivité, on peut se ramener à chercher un minimum sur $K_R = K \cap B(0, R)$. Soit $(x_n)_n$ une suite minimisante dans K_R. Cette suite est bornée donc on peut en extraire une sous-suite faiblement convergente $x_{n_k} \rightharpoonup x^*$. Par ailleurs, K_R est fermé (fortment) et convexe, il est donc faiblement fermé. Ainsi $x^* \in K_R$. Enfin, f est sci et convexe, donc $\mathrm{epi}(f)$ est faiblement fermé. Ainsi, on a

$$f(x^*) \leq \liminf f(x_{n_k}) = \inf_{x \in K_R} f(x).$$

Donc x^* est un minimiseur de f sur K_R et donc sur K. □

Il existe de nombreuses variantes de ce théorème, avec un jeu d'hypothèse changeant, souvent plus restrictif (notamment en développement). Il est bon de savoir lesquelles peuvent être affaiblies et comment.

Corollaire 75 (Stampacchia)**.** *Soit E un espace de Hilbert, $K \subset E$ convexe fermé non-vide, $a(\cdot, \cdot)$ une forme bilinéaire continue coercive*[1] *et φ un forme linéaire continue. Alors la fonctionnelle $f : x \in K \mapsto \frac{1}{2}a(x, x) - \varphi(x)$ admet un unique minimum sur K.*

Démonstration. La fonctionnelle f est continue, donc sci et coercive (au sens de la définition par hypothèse 66). Enfin, pour tout $x \neq y \in K$ et $\lambda \in]0, 1[$, on a

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) = & \frac{\lambda^2}{2}a(x, x) + \frac{(1 - \lambda)^2}{2}a(y, y) + \lambda(1 - \lambda)a(x, y) - \lambda\varphi(x) - (1 - \lambda)\varphi(y) \\ = & \lambda f(x) + (1 - \lambda)f(y) - \frac{\lambda(1 - \lambda)}{2}a(x - y, x - y)) \\ > & \lambda f(x) + (1 - \lambda)f(y) \end{aligned}$$

Donc f est strictement convexe. On conclut à l'existence par la proposition précédent et à l'unicité par stricte convexité. □

1. au sens des formes bilinéaire, i.e $a(x, x) \geq \alpha\|x\|^2$ pour $\alpha > 0$

Application 76 (Problème avec obstacle ([7] Exo. 6 p.324)). *Soit $f \in \mathbb{L}^2(\Omega)$ et $\chi \in \mathcal{C}^0(\Omega)$. On pose $K = \{u \in H_0^1(\Omega) / u \geq \chi \, pp\}$, sous-ensemble convexe de $H_0^1(\Omega)$ et la fonctionnelle $J(v) = \frac{1}{2}\|v\|_{H_0^1}^2 + \langle f, v \rangle$. Alors J admet un unique minimum $u \in K$. De plus, si ce minimum est continu alors il vérifie*

$$\begin{cases} u \in K \\ -\Delta u + f \geq 0 \quad sur \quad \Omega \\ -\Delta u + f = 0 \quad sur \quad \{u > \chi\} \end{cases}$$

Réciproquement, sous des hypothèses supplémentaires de régularité, une solution de ce système est un minimum de J.

3.4 Optimisation sans-contrainte - caractérisation

Dans cette section, on étudie les caractérisation d'un minimum local lorsque K est un ouvert (voire E tout entier). Ces caractérisations se déclinent selon la régularité de la fonction objectif f.

Définition 77. *Soit $\Omega \subset E$ un ouvert et $f : \Omega \to \mathbb{R}$ différentiable sur Ω. $x^* \in \Omega$ est un point critique de f si $df(x^*) = 0$.*

L'équation $df(x) = 0$ est appelé équation d'Euler.

Proposition 78. *Soit $\Omega \subset E$ un ouvert et $f : \Omega \to \mathbb{R}$ différentiable sur Ω. Si $x^* \in \Omega$ est un extremum local de f sur Ω alors c'est un point critique de f.*

Contre-exemple 79. *Cette condition n'ai pas suffisante. $x \in \mathbb{R} \mapsto x^3$ admet un point critique en 0 mais pas d'extremum.*

En revanche, en dimension finie, avec de la convexité, on obtient une CNS.

Proposition 80. *Soit $\Omega \subset \mathbb{R}^d$ un ouvert convexe et $f : \Omega \to \mathbb{R}$ convexe et différentiable sur Ω. Alors tout point critique de f est un minimum global.*

Démonstration. Soit x^* un point critique. D'après la caractérisation de la convexité 57, on a pour tout $x \in \Omega$

$$f(x) - f(x^*) \geq \langle df(x^*), x - x^* \rangle = 0.$$

Donc x^* est un minimum global. □

Application 81. *On considère la fonctionnelle quadratique $f(x) = \frac{1}{2}\langle Ax, x \rangle - \langle b, x \rangle$ avec $A \in \mathcal{S}_d^{++}(\mathbb{R})$. On a vu que f admet un unique minimum global $x^* \in \mathbb{R}^d$. Ce minimum vérifie $Ax^* = b$. En pratique, pour résoudre numériquement un système linéaire, on cherchera à minimiser f.*

Application 82 (Théorème de Rolle). *Soit $f : \mathbb{R}^d \to \mathbb{R}$ différentiable On suppose que f est constante sur $\mathbb{S}^{d-1}$. Alors il existe $x^* \in \mathbb{B}(0, 1)$ tel que $df(x^*) = 0$.*

Cette condition d'ordre 1 ne permet pas de distinguer les minima des maxima. Pour cela, il faut faire appel à une information d'ordre supérieur, i.e pousser le DL à l'ordre 2.

Proposition 83. *Soit $\Omega \subset E$ un ouvert et $f : \Omega \to \mathbb{R}$ différentiable sur Ω et deux fois différentiable en $x^* \in \Omega$. Si $x*$ est un minimum local de f alors $df(x^*) = 0$ et $d^2f(x^*) \in \mathcal{S}_d^+(\mathbb{R})$*

Pour un maximum, la hesienne est négative. Cette condition est seulement nécessaire comme le montre le contre-exemple $x \mapsto x^3$.

Application 84 (Principe du maximum). *Si $f \in \mathcal{C}^2(\overline{B}(0, 1))$ est harmonique sur $B(0, 1)$ alors pour tout $x \in B(0.1)$*

$$\min_{\|y\|=1} f(y) \leq f(x) \leq \max_{\|y\|=1} f(y).$$

En dimension finie, il existe une condition d'ordre 2 suffisante.

Proposition 85. *Soit $\Omega \subset E$ un ouvert et $f : \Omega \to \mathbb{R}$ différentiable sur Ω et deux fois différentiable en $x^* \in \Omega$. Si $df(x^*) = 0$ et $d^2f(x^*) \in \mathcal{S}_d^{++}(\mathbb{R})$ alors x^* est un minimum local de f.*

Contre-exemple 86 ([1])**.** *L'application*

$$f : x \in \ell^2 \mapsto \sum_{n=1}^{+\infty} \frac{x_n^2}{n} - x_n^3$$

est de classe $\mathcal{C}^\infty$ et vérifie $df(0) = 0$ et $\mathrm{Hess}(f)(0)(h,h) > 0$ pour tout $f \in \ell^2 \setminus \{0\}$. Cependant, 0 n'est pas un minimum local de f.

Dans le cas Hilbert, si $\mathrm{Hess}(f)(x^*)$ est positive, inversible d'inverse compact alors la proposition précédente reste vraie.

3.5 Optimisation sous-contrainte

Dans cette section, on étudie des problème d'optimisation sur un sous-ensemble $K \subset \Omega$ fermé. Dans ce cas, l'égalité de Euler et les caractérisation précédente ne permettent plus de traiter les $x \in \partial K$.

Définition 87. *Soit $x^* \in K$. Un vecteur $v \in E$ est une direction admissible en x^* s'il existe une suite $\varepsilon_k \to 0^+$ et $v_k \to v$ tels que $x^* + \varepsilon_k v_k \in K$. On note $K(x^*)$ l'ensemble des direction admissible en $x^* \in K$.*

Exemple 88. *(i) Si $K \subset E$ est ouvert alors $K(x^*) = E$ pour tout $x^* \in K$.*

(ii) Si $K \subset E$ est convexe. $K(x^*) = \overline{\{\lambda(y - x^*)/\lambda > 0, y \in K\}}$ est un semi-cône fermé.

Dans le cadre de l'optimisation sous contrainte, il existe une version affaiblie du critère d'ordre 1.

Theorem 89 (Inéquation d'Euler)**.** *Si $f : K \subset \Omega \to \mathbb{R}$ est différentiable en x^* et x^* est un minimum local de f sur K alors*

$$\langle df(x^*), h\rangle \geq 0, \forall h \in K(x^*).$$

Ce critère est assez peu utilisable en soit. Cependant, il peut être décliner, selon la forme de la contrainte K, en des critères spécifiques plus maniables : le théorème des extrema liés ou théorème de Karush-Kuhn-Tucker.

3.5.1 Contraintes de type égalité

Dans le cas où K est une sous-variété de R^d, on a une caractérisation plus précise de l'ensemble des directions admissibles.

Définition 90. *Soit $p \geq 1$ et $g : \Omega \subset R^d \to \mathbb{R}^p$ de classe $\mathcal{C}^1$ sur Ω. Si $K = g^{-1}(0)$ alors, le problème d'optimisation ${}_m in_K f$ est un problème de minimisation sous contraintes de type égalité. De plus, on dit que les contraintes sont qualifiées en $x \in K$ si la famille $\{\nabla g_1(x), \ldots, \nabla g_p(x)\}$ est libre.*

On peut remarquer que si g est un submersion (dans ce cas K est une sous-variété) alors les contraintes sont qualifiées en tout point de K. Par ailleurs, si $\Omega = \mathbb{R}^d$ alors K est fermé (on peut donc utiliser les résultats d'existence précédent...).

Le lemme suivant décrit l'ensemble des directions admissibles pour des contraintes de type égalité. C'est l'argument clef du théorème des extrema liés et ne peut pas être trivialisé.

Lemme 91. *Si les contraintes sont qualifiées en $x^* \in K$ alors on a*

$$K(x^*) = \mathrm{Vect}\{\nabla g_1(x^*), \ldots, \nabla g_p(x^*)\}^\perp.$$

Démonstration. • L'inclusion direct est triviale (et ne nécessite pas l'hypothèse de qualification). En effet, soit $h \in K(x^*)$, on dispose de suite h_n et ε_n par définition. Puisque x^* et $x^* + \varepsilon_n h_n$ sont dans K, alors on a $g_i(x^* + \varepsilon_n h_n) - g_i(x^*) = 0$. En passant à la limite, on a $\langle \nabla g_i(x^*), h \rangle = 0$.

• L'inclusion réciproque est plus élaborée. Soit $h \in \mathbb{R}^n$ tels que pour tout i, $\langle \nabla g_i(x^*), h \rangle = 0$. La contrainte étant qualifiée en x^*, la matrice $J_g(x^*)$ est de rang plein. On pose $J_g(x^*) = (J_1 \| J_2)$ avec $J_1 \in \mathcal{M}_{p,n-p}(\mathbb{R})$ et $J_2 \in \mathcal{M}_p(\mathbb{R})$. Quitte à permuter les coordonnées, on peut supposer que J_2 est inversible et on note alors $x^* = (x_1^*, x_2^*) \in \mathbb{R}^{n-p} \times \mathbb{R}^p$. D'après le théorème des fonctions implicites, il existe un voisinage $U \subset \mathbb{R}^{n-p}$ de x_1^*, un voisinage $V \subset \mathbb{R}^p$ de x_2^* et un $\mathcal{C}^1$-difféomorphisme $\varphi : U \to V$ tel que $\varphi(x_1^*) = x_2^*$ et pour tout $(x_1, x_2) \in U \times V$

$$g(x_1, x_2) = 0) \Leftrightarrow x_2 = \varphi(x_1).$$

On a de plus $J_\varphi(x_1^*) = -J_2^{-1} J_1$.

A partir d'un certain rang, on a $x_1 + h_1/k \in U$ donc $g(x_1 + h_1/k, \varphi(x_1 + h_1/k)) = 0$. Or d'après la formule de Taylor-Young, on a

$$\begin{aligned}\varphi(x_1 + h_1/k) &= \varphi(x_1^*) + \frac{1}{k} J_\varphi(x_1^*) h_1 + \frac{1}{k}\eta(1/k) \\ &= x_2^* - \frac{1}{k} J_2^{-1} J_1 h_1 + \frac{1}{k}\eta(1/k)\end{aligned}$$

où $\eta(t) \to 0$ quand $t \to 0$. Par hypothèse, on a $J_g(x^*)h = 0$, ce qui se traduit par $J_1 h_1 + J_2 h_2 = 0$. On a donc

$$\varphi(x_1 + h_1/k) = x_2^* + \frac{1}{k}\left(h_2 + \eta(1/k)\right).$$

On a donc montré que

$$(x_1^* + \frac{h_1}{k}, \varphi(x_1^* + \frac{h_1}{k})) = (x_1^*, x_2^*) + \frac{1}{k}\left(h_1, h_2 + \eta(1/k)\right) \in K$$

Ainsi, en posant $\varepsilon_k = 1/k > 0$ et $h_k = (h_1, h_2 + \eta(1/k)) \in \mathbb{R}^n$ on a montré que $h \in K(x^*)$. □

Du lemme précédent et de l'inégalité d'Euler, découle le théorème des extrema liés.

Theorem 92 (extrema liés). *Si f admet un extremum local en $x^* \in K$ et si les contraintes sont qualifiées en x^* alors il existe $(\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$ tels que*

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) = 0. \tag{3.1}$$

Démonstration. D'après l'inégalité d'Euler, pour tout $h \in K(x^*)$, on a $\langle \nabla f(x^*), h \rangle \geq 0$. En particulier, comme $K(x^*)$ est un espace vectoriel, $\langle \nabla f(x^*), h \rangle = 0$. Cela signifie que $\nabla f(x^*) \in K(x^*)^\perp = \mathrm{Vect}\{\nabla g_1(x^*), \dots, \nabla g_p(x^*)\}$. □

Les coefficient λ_i sont appelé multiplicateur de Lagrange. L'équation (3.1) est appelée équation d'Euler-Lagrange. Elle peut être traduite comme l'annulation du gradient d'une fonction auxiliaire, le lagrangien, dépendant et de la fonction objectif et des contrainte

$$\mathcal{L} : (x, \lambda) \in \Omega \times \mathbb{R}^p \mapsto f(x) + \sum_{i=1}^{p} \lambda_i g_i(x).$$

On est ramené à rechercher les points critiques de $\mathcal{L}$, sans contraintes. La Hessienne de $\mathcal{L}$ selon la coordonnée x permet aussi d'avoir une information sur la nature de l'extremum. Pour finir, ce théorème est adaptable dans un espace de Banach.

Application 93. *(i) Inégalité d'Hadamard. Pour $v_1, \dots, v_d \in \mathbb{R}^d$, on a*

$$|\det(v_i, \dots, v_d)| \leq \prod_i |v_i|.$$

(ii) Théorème du min-max de Courant-Fischer. Pour tout $A \in \mathcal{S}_d(\mathbb{R})$

$$\min_{\|x\|=1} \langle Ax, x\rangle = \lambda_1, \quad \max_{\|x\|=1} \langle Ax, x\rangle = \lambda_d.$$

(iii) Inégalité de Minkowski. On montre que les solutions de

$$\max_{\|x\|_p=\alpha, \|y\|_p=\beta} \|x+y\|_p^p$$

sont atteintes sur des vecteurs x^*, y^* colinéaires.

3.5.2 Contraintes de type inégalités

Le travail sur les semi-cônes et leurs cônes polaires associés va permettre de généraliser le théorème des extrema liés au cas des contraintes de type inégalité. On suppose désormais que le domaine admissible a la forme suivante

$$K = g^{-1}(]-\infty, 0]^p) = \{x \in \Omega / g_1(x) \leq 0, \ldots g_p(x) \leq 0\}.$$

Définition 94. *L'ensemble des contraintes saturées en $x^* \in K$ est $\mathcal{I}(x^*) = \{i \in [\![1,p]\!] / g_i(x) = 0\}$. Les contraintes sont qualifiées en x^* s'il existe une direction entrante $v \in \mathbb{R}^d$ en x^*, telle que pour tout $i \in \mathcal{I}(x^*)$, $\langle \nabla g_i(x^*), v\rangle < 0$.*

En particulier, si la famille $\{\nabla g_i(x^*), \ldots, \nabla g_p(x^*)\}$ est libre alors les contraintes sont qualifiées en x^*. Cependant, la vrai condition équivalente, c'est la positive liberté.

Contrairement au cas des contraintes de type égalité la caractérisation des directions admissibles n'utilise pas de théorème complexe mais découle de la définition de contraintes qualifiées.

Lemme 95. *Si les contraintes sont qualifiées en $x^* \in K$ alors*

$$K(x^*) = \{h \in \mathbb{R}^d / \langle \nabla g_i(x^*), h\rangle \leq 0 \quad \forall i \in \mathcal{I}(x^*)\}.$$

L'ingrédient pointu (qui remplace le théorème des fonctions implicites) est le lemme de Farkas. On a le résultat suivant.

Theorem 96 (Karush-Kuhn-Tucker). *Si $x^* \in K$ est un minimum local de f sur K et si les contraintes sont qualifiées en x^*, alors il existe des multiplicateurs de Lagrange $\lambda_1, \ldots, \lambda_p \geq 0$ tels que*

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) = 0. \\ \lambda_i g_i(x^*) = 0, \forall i \end{cases}$$

Démonstration. D'après l'inégalité d'Euler, on a $\forall h \in K(x^*)$, $\langle \nabla f(x^*), h\rangle \geq 0$. Or d'après le lemme précédent, on a

$$\{h \in \mathbb{R}^d / \langle \nabla g_i(x^*), h\rangle \leq 0 \quad \forall i \in \mathcal{I}(x^*)\} \subset \{h \in \mathbb{R}^d / \langle -\nabla f(x^*), h\rangle \leq 0\}.$$

Ainsi, le lemme de Farkas conclu que $-\nabla f(x^*)$ est une combinaison linéaire à coefficients positifs des $\nabla g_i(x^*)$ pour $i \in \mathcal{I}(x^*)$. □

Chapitre 4

Méthodes numériques pour l'optimisation

On considère le problème d'optimisation (en dimension finie) $\min_K f$, avec $f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$, Ω ouvert et $K \subset \Omega$. On suppose que se problème admet au moins une solution x^* et on cherche des valeurs approches de x^*. Une manière concrète de le faire est de procéder itérativement : on construit un suite $(x_k)_k \in K^{\mathbb{N}}$ telle que $x_k \to x^*$. Plus précisément, on cherche x_{k+1} comme une correction de x_k de la forme

$$x_{k+1} = x_k - \rho_k v_k,$$

où $v_k \in \mathbb{R}^d$ est la k-ième direction de descente et $\rho_k > 0$ le pas dans la direction v_k. Un premier problème consiste à trouver des suites v_k et ρ_k tel que $x_k \to x^*$. Un second problème concret est de se fixer une condition d'arrêt pour cesser les calcul dès que x_k est suffisamment proche de x^*. n'ayant pas accès à x^*, on a en pratique recours à l'un des deux critère suivants :

— on s'arrête dès que $\|x_{k+1} - x_k\| < \varepsilon$: l'incrément entre deux itération n'est plus significatif;

— ou dès que $\|\nabla f(x_k)\| < \varepsilon$: x_k est alors presque un point critique.

Cependant, aucun de ces critère ne garantit que x_k est proche de la cible x^*.

4.1 Méthode de relaxation

Une première méthode, naïve, consiste à explorer successivement toutes les directions : $v_k = e_{j_k}$ où $j_k = 1 + (k \mod d)$. A chaque étape, on va ensuite chercher à minimiser f le long de cette direction. On définit ainsi

$$\rho_k = \operatorname{argmin}_{t\in\mathbb{R}} f(x_k - tv_k),$$

sous réserve que ce minimum existe. Remarquons que, étant défini comme un argmin, ρ_k satisfait

$$\langle \nabla f(x_k - \rho_k v_k), v_k \rangle = 0. \tag{4.1}$$

Par construction, la méthode de relaxation est une méthode de descente, c'est à dire que f est décroissante le long de la suite $(x_k)_k$. On peut alors espérer que cette suite soit bien une suite minimisante. C'est le cas sous de bonnes hypothèses.

Theorem 97. *Supposons que f est de classe $\mathcal{C}^1$ et α-convexe. Alors elle admet un unique minimum x^* et pour toute condition initiale $x_0 \in \mathbb{R}^d$, la méthode de relaxation converge vers ce minimum.*

Application 98 (Méthode de Gauss-Seidel)**.** *Dans le cas de la fonctionnelle $f(x) = \langle Ax, x\rangle - \langle b, x\rangle$, les hypothèse du théorème sont satisfaites si et seulement si $A \in \mathcal{S}_d^{++}(\mathbb{R})$. Dans ce cas particulier, le pas ρ_k est explicite*

$$\rho_k = \frac{\langle Ax_k, v_k\rangle - \langle b, v_k\rangle}{\langle Av_k, v_k\rangle}.$$

Autrement dit, si $v_k = i_0$, alors $x_{k+1}^{(i)} = x_k^{(i)}$ pour tout $i \neq i_0$ et $x_{k+1}^{(i_0)}$ vérifie

$$\sum_j a_{i_0 j} x_{k+1}^{(j)} = b_{i_0}.$$

C'est la méthode de Gauss-Seidel ! Le vitesse de convergence de cette méthode est géométrique de raison $\rho((L+D)^{-1}U)$ avec $A = L + D + U$ la décomposition en partie triangulaire supérieure, inférieure et diagonale t ρ le rayon spectral.

4.2 Méthodes de gradient

Un des inconvénient de la méthode précédente est de devoir à chaque étape, résoudre un problème d'optimisation (unidimensionnel, certes). De manière général, on ne pourra obtenir qu'une approximation du pas ρ_k. De plus, le cette de recherche de ρ_k peut être couteuse. L'idée des méthodes de gradient consiste à choisir une direction de descente plus fine : la direction de plus forte pente. D'après la formule de Taylor-Young, on a

$$f(x_{k+1}) = f(x_k) - \rho_k \langle \nabla f(x_k), v_k \rangle + o(rho_k).$$

On cherche alors une direction v_k de sorte que $\langle \nabla f(x_k), v_k \rangle$ soit strictement positive et le plus grand possible. D'après l'inégalité de Cauchy-Schwarz, l' meilleur choix est $v_k = \nabla f(x_k)$.

4.2.1 Méthode de gradient à pas fixe

La méthode de gradient à pas fixe est la méthode naïve. Elle consiste à itérer avec la direction optimale avec un pas $\rho > 0$ indépendant de k :

$$x_{k+1} = x_k - \rho \nabla f(x_k).$$

Theorem 99. *Supposons que f est de classe $\mathcal{C}^2$ et α-convexe. Alors il existe $\rho^* > 0$ tel que pour tout $0 < \rho < \rho^*$ et pour toute condition initiale $x_0 \in \mathbb{R}^d$, la méthode de gradient à pas fixe converge vers l'unique minimum global de f.*

Dans ce cas la convergence est géométrique de raison dépendant de ρ, ce qui incite à choisir ρ le plus grand possible. Le problème c'est que l'on ne connaît pas ρ^*. De plus, cette méthode n'est pas une méthode de descente.

4.2.2 Méthode de gradient à pas optimal

Afin de palier à ce dernier problème, la méthode de gradient à pas optimale garde la direction optimal $v_k = \nabla f(x_k)$ tout en recherchant le pas de descente optimal

$$\rho_k = \operatorname{argmin}_{t \in \mathbb{R}} f(x_k - t v_k).$$

En particulier, la relation (4.1) se réécrit : $\langle v_{k+1}, v_k \rangle = 0$: deux direction de descentes successive sont orthogonales.

Theorem 100. *Supposons que f est de classe $\mathcal{C}^1$ et α-convexe. Alors pour toute condition initiale $x_0 \in \mathbb{R}^d$, la méthode de gradient à pas optimal converge vers l'unique minimum de f.*

On peut affaiblir l'hypothèse de forte convexité, qui sert essentiellement à garantir l'existence d'un unique minimum global. Si f admet un minimum local x^* et qu'elle est α-convexe sur une boule $B(x^*, r)$ (par exemple si $df(x^*) = 0$ et $d^2 f(x^*) \in \mathcal{S}_d^{++}(R)$) alors les méthodes précédentes convergeront vers x^* pour toute condition initiale dans $B(x^*, r)$.

Pour la fonctionnelle quadratique, le pas est toujours explicite et vaut

$$\rho_k = \frac{|v_k|^2}{\langle A v_k, v_k \rangle}.$$

On peut même calculer la vitesse de convergence de manière explicite.

Theorem 101. *Pour tout $x_0 \in \mathbb{R}^d$ et $k \geq 1$*

$$\|x_k - x^*\| \leq \sqrt{\kappa(A)} \left(\frac{\kappa(A)-1}{\kappa(A)+1} \right)^k \|x_0 - x^*\|,$$

où $\kappa(A)$ désigne le conditionnement de A.

La démonstration de ce théorème repose sur une inégalité de convexité.

Lemme 102 (Inégalité de Kantorovitch)**.** *Pour tout $x \in \mathbb{R}^d$, on a*

$$\|x\|^4 \leq \langle Ax, x\rangle \langle A^{-1}x, x\rangle \leq \frac{1}{4}\left(\kappa(A)^{1/2} + \kappa(A)^{-1/2}\right)^2 \|x\|^4.$$

Remarquons qu'une estimation grossière, sans utiliser la convexité, donnerait

$$\kappa(A)^{-1}\|x\|^4 \leq \langle Ax, x\rangle \langle A^{-1}x, x\rangle \leq \kappa(A)\|x\|^4.$$

Ce résultat suggère que lorsque la matrice est mal conditionnée, l'algorithme ne converge pas rapidement. On peut s'en convaincre en considérant la fonction $f(x,y) = x^2 + 100y^2$, associé à une matrice de conditionnement 100. Dans ce cas, les lignes de niveau sont très aplaties et la direction du gradient peut être quasi-orthogonal à la direction $x^* - x_k$. L'algorithme avance alors par très petit pas dans des directions peu avantageuses.

4.2.3 Méthode du gradient à pas conjugué

Le choix du gradient comme direction de descente n'est un choix optimal que localement. Le cas de la fonction f précédente le montre clairement. Une manière de d'améliorer la performance de l'algorithme consiste à utiliser une direction optimal plus global, en gardant en mémoire les directions utilisées précédemment. La descente de gradient à pas conjuguée procède de cette manière. À l'étape k, on choisit simultanément la direction et le pas $w_k = \rho_k v_k$ de sorte que

$$f(x_{k+1}) = \min_{w \in F_k} f(x_k + w), \quad \text{où} \quad F_k = \mathrm{Vect}\{\nabla f(x_0), \ldots, \nabla f(x_k)\}.$$

L'inégalité de Euler et le fait que F_k soit un s.e.v, impliquent que $\langle \nabla f(x_{k+1}), w\rangle = 0$ pour tout $w \in F_k$. En particulier, la famille $(\nabla f(x_k))$ est une famille orthogonale. Cela signifie qu'à l'étape k, on optimise sur un espace de dimension $k+1$ et que $x_{d-1} = x^*$. Cette méthode est exacte. En revanche chaque itération demande de résoudre un problème d'optimisation plus complexe.

Dans le cas particulier de la fonctionnelle quadratique $f(x) = \langle Ax, x\rangle - \langle b, x\rangle$, pour tout $i < k$, on a

$$\langle Aw_k, w_i\rangle = \langle \nabla f(x_k + 1) - \nabla f(x_k), w_i\rangle = 0.$$

Ainsi, la famille (w_k) est orthogonal pour le produit scalaire associé) A. La résolution se fait donc explicitement :

— On construit v_k en utilisant la méthode de Gram-Schmidt, on détermine une direction v_k dans F_k, orthogonal aux $w_0, \ldots, w_{k-1}$

$$v_k = \nabla f(x_k) - \sum \frac{\langle A\nabla f(x_k), w_i\rangle}{\langle Av_i, v_i\rangle} v_i.$$

— On détermine le pas optimal en minimisant $t \mapsto f(x_k - tv_k)$

$$\rho_k = \frac{\langle \nabla f(x_k), v_k\rangle}{\langle Av_k, v_k\rangle}.$$

Cette méthode a une vitesse de convergence meilleure que la descente à pas optimal :

$$\|x_k - x^*\| \leq \sqrt{\kappa(A)} \left(\frac{\sqrt{\kappa(A)}-1}{\sqrt{\kappa(A)}+1} \right)^k \|x_0 - x^*\|,$$

mais toujours sensible au conditionnement.

4.3 Méthodes newtoniennes

Les méthodes précédente tendent à faire décroitre la fonction objectif entre deux itérations. Une démarche différente consiste à rechercher un point critique de la fonction objectif en utilisant une méthode de recherche de zéro, appliqué à la différentielle de la fonction objectif.

Theorem 103 (Méthode de Newton)**.** *Soit $g : \mathbb{R}^d \to \mathbb{R}^d$, de classe $\mathcal{C}^2$, $x^* \in \mathbb{R}^d$ tel que $g(x^*) = 0$ et $dg(x^*) \in GL_d(\mathbb{R})$. Pour $x_0 \in \mathbb{R}^d$, on définit la suite récurrente*

$$x_{k+1} = x_k - dg(x_k)^{-1} g(x_k).$$

Alors, il existe $r > 0$, tel que pour tout $x_0 \in B(x^*, r)$, la suite converge vers x^*, à vitesse quadratique.

Application 104. *Dans le cadre de la minimisation de $f : \mathbb{R}^d \to \mathbb{R}$, α-convexe, de classe $\mathcal{C}^3$, en appliquant la méthode précédente à $g = \nabla f$, on obtient une suite qui converge vers l'unique minimum.*

La convergence quadratique signifie que le nombre de chiffres significatifs exactes double à chaque itération. C'est beaucoup plus rapide que les méthode de gradient mais cela demande plus de régularité. En revanche, il faut à chaque étape inverser une matrice de taille d, ou plus intelligemment, résoudre le système $dg(x_k)(x_{k+1} - x_k) = -g(x_k)$.

Un autre inconvénient de cette méthode est qu'il faut avoir accès à la différentielle de g. Une manière de palier à ce problème est d'utiliser une méthode de type méthode de la sécante. Cette méthode de dimension 1 s'écrit

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})} f(x_k).$$

Elle peut être généralisée en dimension supérieure. Il faut alors déterminer une suite de matrice B_k telle que $B_k(x_k - x_{k-1}) = f(x_k) - f(x_{k-1})$, système sous-déterminé. La méthode de Broyden consiste à chercher une telle suite vérifiant $B_k v = B_{k-1} v$ pour tout $v \perp (x_k - x_{k-1}$. Il existe alors une expression explicite de B_k^{-1} en fonction de B_{k-1}^{-1}. Cette méthode converge à vitesse superlinéaire (mais pas à vitesse quadratique).

Dans certains cas, on peut aussi avoir un candidat naturel pour approximer dg^{-1}. C'est le cas de la méthode quasi-newtonienne suivante.

Application 105 ([4])**.** *Soit $A \in \mathbb{S}_d^+(\mathbb{R})$. On cherche à approcher A^{-1} en minimisant la fonction*

$$f : X \in \mathcal{M}_d(\mathbb{R}) \mapsto \frac{1}{2} \operatorname{Tr}(AXX^t) - \operatorname{Tr}(X).$$

Le problème, pour appliquer la méthode de Newton, c'est que $dg(X) = A$ et ainsi, inverser dg c'est déjà résoudre le problème. En revanche, si l'on a une suite $(X_k)_k$ qui converge vers A^{-1} alors elle converge vers dg^{-1}. Cela suggère la méthode quasi-newtonienne suivante :

$$X_{k+1} = X_k - X_k(AX_k - \mathrm{id}).$$

On peut alors montrer qu'elle converge pour la condition initiale $X_0 = rA^t$, avec r (explicitement) suffisamment petit.

4.4 Algorithmes sous contraintes

Les méthodes de résolution algorithmique des problème de minimisation sans contraintes ne se généralise pas nécessairement bien aux problèmes sous contraintes. Dans ces cas, deux stratégies permettent une généralisation : la projection et la pénalisation.

4.4.1 Méthode de relaxation sur un pavé

La méthode naïve de relaxation se généralise naturellement sur les pavés. On suppose ici que $K = \prod[a_i, b_i]$, non nécessairement borné. On peut alors définir la suite récurrente $x_{k+1} = x_k - \rho_k v_k$ de la manière suivante : $v_k = e_{i_k}$ avec $i_k = k + 1[d]$ et

$$\rho_k = \mathrm{argmin}_{t \in I_k} f(x_k - tv_k), \quad I_k = [x_k^{(i_k)} - b_{i_k}, x_k^{(i_k)} - a_{i_k}].$$

Theorem 106. *Si K est un pavé et f est de classe C^1 et fortement convexe, alors la méthode de relaxation converge pour toute condition initiale.*

Contre-exemple 107. *Si K n'est pas un pavé, la convergence n'est pas garantie. Par exemple, pour $K = \{x + y \geq 2\} \subset \mathbb{R}^2$, $f(x,y) = x^2 + y^2$ et $(x_0, y_0) = (2,2)$ alors la suite stationne dès la première itération sur $(0,2)$ qui n'est pas le minimum de f, en l'occurrence $(1,1)$.*

4.4.2 Méthode de projection

On suppose que K est convexe fermé non vide et f convexe sur K. Alors x^* est un minimum de f, si et seulement si on a $\langle \nabla f(x^*), y - x^* \rangle \geq 0$ pour tout $y \in K$. Par caractérisation de la projection, x^* est un minimum si et seulement si $x^* = \pi_K(x^* - \rho \nabla f(x^*))$. On est ainsi ramené à rechercher un point fixe. On définit alors une méthode de type itéré de Picard :

$$x_0 \in K, \quad x_{k+1} = \pi_K \left(x_k - \rho \nabla f(x_k) \right).$$

Cette méthode généralise la méthode de gradient à pas fixe.

Theorem 108. *On suppose K convexe fermé non vide, $f : K \to \mathbb{R}$ fortement convexe de gradient localement lipschitzien. Alors il existe $\rho^* > 0$ tel que pour tout $0 < \rho < \rho^*$, la méthode converge vers l'unique minimum.*

En pratique, K est un pavé et la projection se calcul facilement.

4.4.3 Méthode de pénalisation

On cherche un méthode applicable lorsque K n'est pas un pavé. L'idée est de transformé le problème de minimisation sous contrainte en un problème sans contrainte en modifiant la fonction objectif. Pour cela, on va perturber la fonction par la distance à K.

Proposition 109. *Soit $K \subset \mathbb{R}^d$ non vide, $f : \mathbb{R}^d \to \mathbb{R}$ L-lipschitzienne. Pour $R > 0$, on pose $f_R : x \in \mathbb{R}^d \mapsto f(x) + Rd(x, K)$.*

(i) Si $R > L$ et si x^* est un minimum global de f sur K alors x^* est un minimum global de f_R.

(ii) Si $R > L$ et K fermé, alors tout minimum de f_R est dans K.

Theorem 110. *Soit K convexe fermé non vide et $f : \mathbb{R}^d \to \mathbb{R}$ fortement convexe. Soit $\varphi : \mathbb{R}^d \to \mathbb{R}_+$, convexe, telle que $K = \varphi^{-1}(0)$. Alors pour tout $k \geq 1$ $f_k = k + k\varphi$ admet un unique minimum $x_k \in \mathbb{R}^d$ et $x_k \to x^* = \mathrm{argmin}_K f$.*

Application 111 (Programmation convexe)**.** *Soit $f : \mathbb{R}^d \to \mathbb{R}$ fortement convexe et $K = \{g_i \leq 0\}$ avec g_i des contrainte convexes. La fonction $\varphi(x) = \sum \max(g_i(x), 0)$ convient.*

Bibliographie

[1] D. Azé, G. Constans, and J.-B. Hiriart-Urruty. *Calcul différentiel et équations différentielles*. EDP Science, 2010.

[2] V. Beck, J. Malick, and G. Peyré. *Objectif Agregation*. H&K, 2005.

[3] H. Brézis. *Analyse fonctionnelle. Théorie et applications*. Paris : Masson, 1994.

[4] P. Caldero and Q. Chamussy. *Carnet de voyage en Analystan*. Calvage & Mounet, 2024.

[5] P. Ciarlet and J.-L. Lions. *Introduction à l'analyse numérique matricielle et à l'optimisation*. Dunod, 2007.

[6] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. Part 1 : Fundamentals*, volume 305 of *Grundlehren Math. Wiss.* Berlin : Springer-Verlag, 1993.

[7] F. Hirsh and G. Lacombe. *Élément d'analyse fonctionnelle*. Dunod, 2009.