ANALYSE CONVEXE

Prof. Adama COULIBALY
UFR Mathématiques et Informatique,
Université Félix HOUPHOUET-BOIGNY,
22 BP 582 Abidjan 22, Côte d'Ivoire.

28 mars 2023

Table des matières

Chapitre 1

Ensembles convexes

Le cadre général de ce cours est un espace vectoriel réel de dimension n. On peut donc sans perdre de généralités considérer l'espace vectoriel réel $\mathbb{R}^n$.

1.1 Ensembles convexes

1.1.1 Définitions

La notion de combinaison linéaire convexe est donnée dans la définition suivante.

Définition 1.1.1 *On appelle combinaison linéaire convexe de deux points x et y de $\mathbb{R}^n$, tout point $z = (1-\lambda)x + \lambda y$ avec $\lambda \in [0,1]$.*

De façon générale :

Définition 1.1.2 *On appelle combinaison linéaire convexe de k points $x_1, \cdots, x_k$ de $\mathbb{R}^n$, tout élément $x \in \mathbb{R}^n$ tel que*

$$x = \sum_{i=1}^{k} \lambda_i x_i \text{ avec } \quad \lambda_i \geq 0 \text{ et } \sum_{i=1}^{k} \lambda_i = 1.$$

On définit les notions suivantes :

Définition 1.1.3 *Soit x, $y \in \mathbb{R}^n$; on appelle segment "fermé" d'extrémités x et y, l'ensemble noté $[x, y]$ et défini par :*

$$[x,y] = \{z \in \mathbb{R}^n : z = (1-\lambda)\,x + \lambda y \ : \lambda \in [0,1]\}\,.$$

C'est l'ensemble de toutes les combinaisons linéaires convexes des points x et y.

De façon analogue, on définit :

Définition 1.1.4 *On appelle segment "ouvert" d'extrémités x et y, et on le note $]x, y[$, l'ensemble*

$$]x,y[= \{z \in \mathbb{R}^n : z = (1-\lambda)\,x + \lambda y : \ \lambda \in]0,1[\}\,.$$

On définit aussi $]x,y]$ et $[x,y[$ qui sont appelés segment semi ouvert en x respectivement en y.

$$]x,y] = \{z \in \mathbb{R}^n : z = (1-\lambda)\,x + \lambda y \ \ \lambda \in]0,1]\}\,.$$

$$[x,y[= \{z \in \mathbb{R}^n : z = (1-\lambda)\,x + \lambda y \ \ \lambda \in [0,1[\}\,.$$

Définition 1.1.5 *Soit C une partie de $\mathbb{R}^n$. C est convexe si seulement si pour tout x, $y \in C$, $(1-\lambda)\,x + \lambda y \in C$ pour tout $\lambda \in [0,1]$. Autrement dit, C est convexe si seulement si C contient tout segment fermé d'extrémités deux quelconques de ses points.*

Exemple 1.1.1 - *Dans $\mathbb{R}^n$, les ensembles suivants sont convexes. $\mathbb{R}^n$, l'ensemble vide, les singletons, les boules, les segments.*
- *Dans $\mathbb{R}$, les parties convexes sont les intervalles.*

On a la proposition :

Proposition 1.1.1 *Une partie C de $\mathbb{R}^n$ est convexe si seulement si elle contient toute combinaison linéaire convexe de toute famille finie d'éléments qui lui appartiennent.*

Preuve : Si C contient toute combinaison linéaire convexe de familles finies d'éléments qui lui appartiennent, en particulier, prenant une famille de deux éléments x et y de C, on a $[x,y] \subset C$ et donc C est convexe.

Réciproquement, soit C un ensemble convexe de $\mathbb{R}^n$. Alors C contient toute combinaison linéaire convexe de deux quelconques de ses éléments. Donc la propriété est vraie pour une famille comportant deux éléments. Supposons qu'elle est vraie pour une famille de $k-1$ éléments.

Soit

$$\mathcal{F} = \left\{x^1,\, x^2, \cdots, x^k\right\}$$

une famille de k éléments de C.

Soit

$$x = \sum_{i=1}^{k} \lambda_i x^i \text{ avec } \lambda_i \geq 0,\ \sum_{i=1}^{k} \lambda_i = 1.$$

On a

$$x = \sum_{i=1}^{k} \lambda_i x^i = \sum_{i=1}^{k-1} \lambda_i x^i + \lambda_k x^k.$$

Soit

$$\lambda = \sum_{i=1}^{k-1} \lambda_i.$$

On a $\lambda \in [0,1]$.

Si $\lambda = 0$ alors $\lambda_i = 0$ pour tout $i = 1, \cdots, k-1$ et donc $\lambda_k = 1$. Il vient alors que $x = \lambda_k x^k = x^k \in C$.

Si $\lambda \neq 0$, on peut écrire

$$x = \lambda \sum_{i=1}^{k-1} \left(\frac{\lambda_i}{\lambda}\right) x^i + \lambda_k x^k.$$

L'élément

$$y = \sum_{i=1}^{k-1} \left(\frac{\lambda_i}{\lambda} \right) x^i,$$

est une combinaison linéaire convexe de $k-1$ éléments de C. C'est donc un élément de C, par hypothèse de recurrence. Donc $x = \lambda y + \lambda_k x^k$. Or $\lambda_k = 1 - \lambda$ avec $\lambda \in [0,1]$. Donc x est combinaison linéaire convexe de deux éléments de C. Comme par hypothèse, C est convexe, on a alors $x \in C$. □

On a les propriétés suivantes

1.1.2 Propriétés algébriques

On rappelle les notions suivantes.

Définition 1.1.6 *Une application f de $\mathbb{R}^n$ dans $\mathbb{R}^m$ est dite affine si l'une des conditions suivantes est vérifiée.*
i) Pour tout x, y dans $\mathbb{R}^n$ et $\lambda \in \mathbb{R}$, on a

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

ii) Il existe une application linéaire $\mathcal{L}$ de $\mathbb{R}^n$ dans $\mathbb{R}^m$ et $a \in \mathbb{R}^m$ tels que :

$$\forall\, x \in \mathbb{R}^n,\ \ f(x) = \mathcal{L}(x) + a.$$

Les résultats suivants sont immédiats.

Proposition 1.1.2 1) *Si C_1 et C_2 sont convexes alors pour tous α_1 et α_2 dans $\mathbb{R}$, $\alpha_1 C_1 + \alpha_2 C_2$ est convexe.*
2) *Toute intersection de parties convexes est convexe.*
3) *Toute réunion d'une suite croissante de convexes est convexe.*
4) *Le produit cartésien de deux convexes est convexe.*
5) *L'image d'un convexe par une application affine est convexe.*
6) *L'image réciproque d'un convexe par une application affine est convexe.*

On a la proposition suivante

Proposition 1.1.3 *Si C est convexe alors pour tout α et β positifs ou nuls, on a*

$$\alpha C + \beta C = (\alpha + \beta)C.$$

Preuve : Comme les scalaires α et β sont positifs ou nuls, le cas où $\alpha + \beta = 0$ est trivial
Considérons α et β tels que $\alpha + \beta > 0$.
L'inclusion ci-dessous est immédiate :

$$(\alpha + \beta)C \subset \alpha C + \beta C.$$

Montrons à présent que

$$\alpha C + \beta C \subset (\alpha + \beta)C.$$

Soit $z \in \alpha C + \beta C$. Alors il existe x, y dans C tels que $z = \alpha x + \beta y$.
On peut écrire :

$$z = (\alpha + \beta) \left[\frac{\alpha}{\alpha + \beta} x + \frac{\beta}{\alpha + \beta} y \right].$$

On a

$$\frac{\alpha}{\alpha + \beta} \geq 0, \frac{\beta}{\alpha + \beta} \geq 0, \frac{\alpha}{\alpha + \beta} + \frac{\beta}{\alpha + \beta} = 1.$$

Comme C est convexe, alors

$$\frac{\alpha}{\alpha + \beta} x + \frac{\beta}{\alpha + \beta} y \in C.$$

D'où le résultat □

1.1.3 Propriétés topologiques

On rappelle les définitions suivantes. Bien avant notons que pour $x \in \mathbb{R}^n$, et $\varepsilon > 0$, $B(x, \varepsilon)$ désigne la boule euclidienne fermée de centre x et de rayon ε

On remarque qu'on a toujours $B(x, \varepsilon) = x + \varepsilon B(0, \varepsilon) = x + \varepsilon B(0, 1)$, $B(0, 1)$ étant la boule unité euclidienne fermée.

Définition 1.1.7 *Etant donné un sous ensemble C de $\mathbb{R}^n$, son adhérence son intérieur et sa frontière sont respectivement les ensembles :*

$\overline{C} = \{x \in \mathbb{R}^n : \exists \{x^k\} \subset C : x_k \to x\}$

$\mathrm{int}(C) = \{x \in \mathbb{R}^n : \exists \varepsilon > 0,\ B(x, \varepsilon) \subset C\}$,

$\mathrm{Fr}(C) = \overline{C} \setminus \mathrm{int}(C)$.

Proposition 1.1.4 *Soit C un convexe de $\mathbb{R}^n$. Alors :*

1) $\overline{C}$ est convexe.

2) $\mathrm{int}(C)$ est convexe.

Preuve : 1) Soit x, $y \in \overline{C}$ et $z = (1 - \lambda)x + \lambda y$, $\lambda \in [0, 1]$.

Comme $x \in \overline{C}$ il est équivalent de dire qu'il existe une suite $\{x^k\}$ de points de C convergent vers x. De même il existe une suite $\{y^k\}$ de points de C convergent vers y. L'ensemble C étant convexe, on a pour tout k, $z^k = (1 - \lambda)x^k + \lambda y^k \in C$. Comme la suite $\{z^k\}$ converge vers z alors $z \in \overline{C}$. Donc $\overline{C}$ est convexe.

2) Soient x et y deux éléments de $int(C)$, $\lambda \in [0, 1]$ et $z = (1 - \lambda)x + \lambda y$. D'après la définition de $\mathrm{int}(C)$, il existe ε_1 et ε_2 tels que $B(x, \varepsilon_1) \subset C$ et $B(y, \varepsilon_2) \subset C$. Donc pour $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$, on a

$$B(x, \varepsilon) \subset C,\ B(y, \varepsilon) \subset C.$$

Comme

$$\begin{aligned} B(z, \varepsilon) &= z + B(0, \varepsilon) \\ &= (1 - \lambda)x + \lambda y + B(0, \varepsilon) \\ &= (1 - \lambda)x + \lambda y + ((1 - \lambda) + \lambda)B(0, \varepsilon) \\ &= (1 - \lambda)\left[x + B(0, \varepsilon)\right] + \lambda \left[y + B(0, \varepsilon)\right] \\ &= (1 - \lambda)B(x, \varepsilon) + \lambda B(y, \varepsilon) \end{aligned}$$

et

$$(1 - \lambda)B(x, \varepsilon) + \lambda B(y, \varepsilon) \subset (1 - \lambda)C + \lambda C = C,$$

car C est convexe, on conclut que $B(z, \varepsilon) \subset C$. Par suite $z \in int(C)$. Donc $int(C)$ est convexe. □

1.2 Enveloppes convexes

Définition 1.2.1 *Soit S une partie de $\mathbb{R}^n$. On appelle enveloppe convexe de S, l'intersection de tous les convexes contenant S. C'est le plus petit convexe contenant S. On le note $conv(S)$.*

On a le résultat suivant.

Proposition 1.2.1 *Soit $S \subset \mathbb{R}^n$. L'enveloppe convexe de S, est l'ensemble de toutes les combinaisons linéaires convexes finies d'éléments de S. Autrement dit on a :*

$$conv(S) = \left\{ x = \sum_{i=1}^{k} \lambda_i x^i : k \in \mathbb{N}^*, \begin{array}{l} x^i \in S, \forall i = 1, \cdots, k, \\ \lambda_i \geq 0, \forall i = 1, \cdots, k, \sum_{i=1}^{k} \lambda_i = 1 \end{array} \right\}$$

Preuve :

Posons

$$\mathcal{C} = \left\{ x = \sum_{i=1}^{k} \lambda_i x^i : k \in \mathbb{N}^*, \begin{array}{l} x^i \in S, \forall i = 1, \cdots, k, \\ \lambda_i \geq 0, \forall i = 1, \cdots, k, \sum_{i=1}^{k} \lambda_i = 1 \end{array} \right\}$$

Soit C un convexe de $\mathbb{R}^n$ contenant S. Donc C contient toute combinaison linéaire convexe de toute famille finie d'éléments de C. Comme C contient S alors C contient toute combinaison linéaire convexe de toute famille finie d'éléments de S et donc C contient $\mathcal{C}$. Par suite $\mathcal{C}$ est contenu dans tous les convexes contenant S. On a alors $\mathcal{C} \subset conv(S)$.

D'autre part, on vérifie facilement que $\mathcal{C}$ est convexe et contient S. Et comme $conv(S)$ est le plus petit convexe contenant S, on a alors $conv(S) \subset \mathcal{C}$. D'où l'égalité $conv(S) = \mathcal{C}$.
□

Proposition 1.2.2 *S est convexe si et seulement si $conv(S) = S$*

Preuve : Si S est convexe alors il est le plus petit convexe contenant S. Donc $conv(S) = S$. Réciproquement si on a $conv(S) = S$ alors S est convexe. □

Proposition 1.2.3 *1) $conv(conv(S)) = conv(S)$*
2) Si on a $A \subset B$ alors $conv(A) \subset conv(B)$

Preuve : 1) Comme $conv(S)$ est convexe, on a $conv(conv(S)) = conv(S)$.

2) Soit $A \subset B$: le plus petit convexe contenant B contient aussi A donc $conv(B)$ contient $conv(A)$. □

Définition 1.2.2 *L'enveloppe convexe d'un nombre fini de points est appelée polytope.*

Avant de définir un polytope particulier qui est beaucoup utilisé en programmation linéaire, donnons la définition d'une famille de points affinement indépendants.

Définition 1.2.3 *Soit $x^0,\ x^1,\ \cdots,\ x^k$ $k+1$ points de $\mathbb{R}^n$. On dit qu'ils sont affinement indépendants si les vecteurs*

$$x^1 - x^0,\ x^2 - x^0,\ \cdots,\ x^k - x^0$$

sont linéairement indépendants

Définition 1.2.4 *L'enveloppe convexe de $k+1$ points affinement indépendants est appelée k-simplexe. Ces points sont les sommets du k-simplexe.*

On montre que

Théorème 1.2.1 (C. Carathéodory)
Soit S un sous ensemble non vide de $\mathbb{R}^n$ tel que $\dim(conv(S)) = k < +\infty$. Alors tout point $x \in conv(S)$ peut s'écrire comme une combinaison linéaire convexe d'au plus $k+1$ éléments de S.

Remarque 1.2.1 *Etant donné $S \subset \mathbb{R}^n$, pour obtenir son enveloppe convexe, il suffit de considérer les combinaisons linéaires convexes d'au plus $n+1$ éléments de S.*

On a la propriété topologique suivante ;

Proposition 1.2.4 *L'enveloppe convexe d'un ouvert est un ouvert.*

Preuve : Soit S un ouvert : montrons que $\mathrm{int}(conv(S)) = conv(S)$.
Supposons qu'il existe un élément x dans $S \cap \mathrm{Fr}(conv(S))$. Alors comme S est ouvert, il existerait un voisinage V de x contenu dans S et donc dans $conv(S)$. Or

$$x \in \mathrm{Fr}(conv(S)) = \overline{conv(S)} \setminus \mathrm{int}(conv(S)).$$

Ce qui est contradictoire.
Donc $S \cap \mathrm{Fr}(conv(S)) = \emptyset$. On a alors $S \subset \mathrm{int}(conv(S))$. Comme $conv(S)$ est le plus petit convexe contenant S, on a $conv(S) \subset \mathrm{int}(conv(S))$. On obtient donc $conv(S) = \mathrm{int}(conv(S))$. □

Comme le montre l'exemple ci-dessous, l'enveloppe convexe d'un fermé n'est pas en général un fermé.
Soit

$$S = \{(x, y) \in \mathbb{R}^2 : x \geq 0, xy \geq 1\} \cup \{(0, 0)\}$$

On a

$$conv(S) = \{(x, y) : x > 0, y > 0\} \cup \{(0, 0)\}$$

qui n'est pas un fermé.

Par contre on montre que :

Proposition 1.2.5 *Si S est compact alors $conv(S)$ est compact.*

Preuve :
On a :

$$conv(S) = \{x = \sum_{i=1}^{n+1} \lambda_i x_i : \lambda_i \geq 0, x_i \in S,\ \forall i = 1, \cdots, n+1, \sum_{i=1}^{n+1} \lambda_i = 1\}.$$

Posons

$$K = \{\lambda \in \mathbb{R}^{n+1} : \lambda_i \geq 0, \forall i = 1, \cdots, n+1, \sum_{i=1}^{n+1} \lambda_i = 1\}.$$

K est compact.

Soit

$$f: \begin{array}{l} S^{n+1} \times K \longrightarrow \mathbb{R}^n \\ (x_1, \cdots x_{n+1}, \lambda) \longmapsto \sum_{i=1}^{n+1} \lambda_i x_i \end{array}$$

f est continue et $S^{n+1} \times K$ est compact ce qui implique que $f(S^{n+1} \times K)$ est compact. Or $f(S^{n+1} \times K) = conv(S)$. D'où la proposition. □

Définition 1.2.5 *Soit S une partie de $\mathbb{R}^n$, l'enveloppe convexe fermée de S est l'intersection de tous les convexes fermés contenant S. C'est le plus petit convexe fermé contenant S. On le note $\overline{conv}(S)$.*

Proposition 1.2.6 *1) Si A et B sont deux sous ensembles de $\mathbb{R}^n$ avec $A \subset B$, alors $\overline{conv}(A) \subset \overline{conv}(B)$.*
2) Si S est une partie de $\mathbb{R}^n$, on a :

$$\overline{conv}(S) = \overline{conv}(\overline{S}) = \overline{conv(S)}.$$

Preuve : La preuve de 1) est immédiate.

2) L'ensemble des convexes fermés contenant S est égal à l'ensemble des convexes fermés contenant $\overline{S}$. Donc $\overline{conv}(S) = \overline{conv}(\overline{S})$.

Montrons que

$$\overline{conv}(S) = \overline{conv(S)}.$$

On a

$$S \subset conv(S) \subset \overline{conv(S)}.$$

Or $\overline{conv(S)}$ est un convexe fermé ; donc,

$$\overline{conv}(S) \subset \overline{conv(S)}.$$

D'autre part, on a

$$conv(\overline{S}) \subset \overline{conv}(S)$$

et $S \subset \overline{S}$: donc

$$conv(S) \subset conv(\overline{S}) \subset \overline{conv}(S).$$

On en déduit alors que

$$\overline{conv(S)} \subset \overline{\overline{conv}(S)} = \overline{conv}(S).$$

Ce qui donne la deuxième inclusion et donc l'égalité recherchée. □

1.3 Points extrêmes

Définition 1.3.1 *Soit C un sous ensemble convexe non vide de $\mathbb{R}^n$.*
Un point x de C est un point extrême si

$$\forall\, (x_1, x_2, \alpha) \in C^2 \times]0,1[,\ x = (1-\alpha)x_1 + \alpha x_2 \Longrightarrow x_1 = x_2 = x.$$

i.e. x ne peut pas s'écrire comme combinaison linéaire convexe stricte (coefficients non nuls) de deux points distincts de C, ou encore x n'appartient à l'intérieur relatif d'aucun segment fermé d'extrémités deux points distincts de C ou encore toute droite passant par x rencontre C suivant un segment ou une demi-droite dont x est une extrémité.

L'ensemble $\mathrm{ext}(C)$ *des points extrêmes de C est appelé profil de C.*

On a la caractérisation suivante

Proposition 1.3.1 *Soit C un sous ensemble convexe de $\mathbb{R}^n$. Les propositions suivantes sont équivalentes.*
i) $x \in \text{ext}(C)$
ii) $C \setminus \{x\}$ *est convexe.*

Preuve : Soit $x \in \text{ext}(C)$.
Si

$$(x_1, x_2, \alpha) \in (C \setminus \{x\}) \times (C \setminus \{x\}) \times]0, 1[$$

alors

$$(x_1, x_2, \alpha) \in C \times C \times]0, 1[\text{ et } x_1 \neq x \neq x_2.$$

Comme C est convexe, on a

$$(1 - \alpha)x_1 + \alpha x_2 \in C \text{ et } x_1 \neq x \neq x_2.$$

Alors

$$(1 - \alpha)x_1 + \alpha x_2 \in C \setminus \{x\}$$

et donc $C \setminus \{x\}$ est convexe.

Réciproquement, supposons que $C \setminus \{x\}$ est convexe et soit

$$(x_1, x_2, \alpha) \in C \times C \times]0, 1[\text{ avec } (1 - \alpha)x_1 + \alpha x_2 = x.$$

Donc

$$(1 - \alpha)x_1 + \alpha x_2 \notin C \setminus \{x\}.$$

Ce qui implique que :

$$x_1 \notin C \setminus \{x\} \text{ ou } x_2 \notin C \setminus \{x\}$$

C'est-à-dire que $x_1 = x$ ou $x_2 = x$. Comme $(1 - \alpha)x_1 + \alpha x_2 = x$ avec $\alpha \in]0, 1[$, on a nécessairement $x_1 = x = x_2$. D'où $x \in \text{ext}(C)$. □

On admettra que :

Théorème 1.3.1 *Tout convexe compact est l'enveloppe convexe de ses points extrêmes.*

1.4 Projection sur un convexe fermé

Dans ce chapitre $\mathbb{R}^n$ est muni de sa structure euclidienne.
Le théorême ci-dessous est fondamental en optimisation.

Théorème 1.4.1 *Soit S un convexe fermé de $\mathbb{R}^n$ et $x \in \mathbb{R}^n$. Il existe un unique point $p(x) \in S$ dont la distance à x est minimale. C'est-à-dire*

$$\|p(x) - x\| \leq \|y - x\| \quad \forall y \in S.$$

$p(x)$ est appelé projection de x sur S.

Preuve : Soit $d = \inf_y [\|y - x\| : y \in S]$, $B(x, d + 1)$ la boule fermée de centre x et de rayon $d + 1$. Considérons $K = S \cap B(x, d + 1)$. L'ensemble K est fermé et borné, c'est donc un compact. Alors l'application $y \mapsto \|y - x\|$ étant continue elle atteint ses bornes sur K. Donc il existe $p(x) \in K$ et donc $p(x) \in S$ tel que $\|p(x) - x\| = d$. D'où le résultat d'existence. Montrons maintenant l'unicité.

Soient a et b deux points de S tels que

$$\|a - x\| = \|b - x\| = d = \inf_y [\|y - x\| : y \in S].$$

Comme S est convexe alors $\frac{1}{2}a + \frac{1}{2}b \in S$. Donc

$$d \leq \|x - \frac{a + b}{2}\| = \frac{1}{2}\|2x - a - b\| = \frac{1}{2}\|(x - a) + (x - b)\|.$$

Ce qui implique que

$$d \leq \|x - \frac{a + b}{2}\| \leq \frac{1}{2}(\|x - a\| + \|x - b\|) = d.$$

On en déduit que $\|x - \frac{a+b}{2}\| = d$. Il s'ensuit que $\|x - a + x - b\| = \|x - a\| + \|x - b\|$ c'est-à-dire qu'il existe $\lambda \in \mathbb{R}_+$ tel que $x - a = \lambda(x - b)$. Donc on a $\lambda = 1$ c'est-à-dire $x - a = x - b$ et donc $a = b$. D'où l'unicité. $\square$

Une caractérisation de la projection est donnée ici.

Théorème 1.4.2 *La projection $p(x)$ de x sur S un convexe fermé de $\mathbb{R}^n$ est l'unique point z de S satisfaisant la condition*

$$\langle y - z, x - z\rangle \leq 0 \quad \forall y \in S.$$

Autrement dit, $p(x)$ est la projection de x sur S si et seulement si

$$\langle x - p(x), y - p(x)\rangle \leq 0 \quad \forall y \in S.$$

Preuve : Soit $p(x)$ la projection sur S un convexe fermé de $\mathbb{R}^n$. Supposons qu'il existe un point $y_0 \in S$ tel que $\langle y_0 - p(x), x - p(x)\rangle > 0$.
Posons

$$\varphi(t) = \|x - [(1 - t)p(x) + ty_0]\|^2, \quad t \in \mathbb{R}^+.$$

On a

$$\begin{aligned}\varphi(t) &= \|x - p(x)\|^2 + 2t\langle x - p(x), p(x) - y_0\rangle + t^2\|p(x) - y_0\|^2 \\ &= \|x - p(x)\|^2 + t\left(2\langle x - p(x), p(x) - y_0\rangle + t\|p(x) - y_0\|^2\right).\end{aligned}$$

Comme $\langle x - p(x), p(x) - y_0\rangle < 0$, pour t suffisamment petit, on aura $\varphi'(t) < 0$ et donc φ est strictement décroissante au voisinage de 0. On aura donc pour t suffisamment petit $\varphi(t) < \|x - p(x)\|^2$.

Or pour $t \in [0; 1]$, $z_t = (1 - t)p(x) + ty_0 \in S$ car S est convexe. Soit donc $t \in [0; 1]$ suffisamment petit tel que $\varphi(t) < \|x - p(x)\|^2$. Pour un tel t on aura $z_t \in S$ et $\|x - z_t\|^2 < \|x - p(x)\|^2$. Ce qui contredit le fait que $p(x)$ est la projection de x sur S.

Réciproquement supposons que

$$\langle y - z; x - z\rangle \leq 0 \quad \forall y \in S.$$

Montrons que z est la projection de x sur S.

Soit $y \in S$. On a

$$\|x-y\|^2 = \|x-z+z-y\|^2 = \|x-z\|^2 + 2\langle x-z; z-y\rangle + \|y-z\|^2.$$

Par suite on a $\|x-y\|^2 \geq \|x-z\|^2$, et donc $p(x) = z$. □

L'application projection est lipschitzienne sur $\mathbb{R}^n$.

Proposition 1.4.1 *Soit S un convexe fermé non vide de $\mathbb{R}^n$ et pour tout $x \in \mathbb{R}^n$, $p(x)$ la projection de x sur S. On a*

$$\|p(x) - p(y)\| \leq \|x-y\| \quad \forall x,\ y \in \mathbb{R}^n.$$

Preuve : Soient x et y fixés dans $\mathbb{R}^n$. Considérons $p(x)$ et $p(y)$ respectivement leurs projections sur S. On sait d'après la caractérisation de la projection que $p(x)$ est tel que

$$\langle z - p(x), x - p(x)\rangle \leq 0 \quad \forall z \in S.$$

En prenant $z = p(y)$ on obtient

$$\langle p(y) - p(x), x - p(x)\rangle \leq 0. \tag{1.1}$$

De même en considérant $p(y)$ on a

$$\langle z - p(y), y - p(y)\rangle \leq 0 \quad \forall z \in S.$$

Pour $z = p(x)$ on a

$$\langle p(x) - p(y), y - p(y)\rangle \leq 0. \tag{1.2}$$

En additionnant (1.1)et (1.2) on obtient

$$\langle p(x) - p(y), y - p(y) + p(x) - x\rangle \leq 0,$$

qui est équivalent à

$$\langle p(x) - p(y), p(x) - p(y) - (x-y)\rangle \leq 0,$$

c'est-à-dire

$$\langle p(x) - p(y), p(x) - p(y)\rangle \leq \langle p(x) - p(y), x - y\rangle.$$

On en déduit

$$\|p(x) - p(y)\|^2 \leq \|p(x) - p(y)\|\|x-y\|.$$

D'où le résultat. □

Corollaire 1.4.1 *l'application définie sur $\mathbb{R}^n$ qui à tout $x \in \mathbb{R}^n$ associe sa projection sur S un convexe fermé est continue.*

Chapitre 2

Fonctions convexes de plusieurs variables

2.1 Définitions, propriétés de base et exemples

Définition 2.1.1 *Soit C un convexe non vide de $\mathbb{R}^n$.*
Une fonction $f : C \to \mathbb{R}$ est convexe sur C, si

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

Définition 2.1.2 *Soit C un convexe non vide de $\mathbb{R}^n$.*
Une fonction $f : C \to \mathbb{R}$ est strictement convexe sur C, si

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

Définition 2.1.3 *Soit C un convexe non vide de $\mathbb{R}^n$.*
Une fonction $f : C \to \mathbb{R}$ est fortement convexe (on dit aussi elliptique ou encore uniformément convexe) sur C, s'il existe $r > 0$ tel que :

$$\forall\, x,\ y \in C, \forall\, \lambda \in [0,1],$$
$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) - \tfrac{1}{2} r\lambda(1-\lambda)\|y-x\|^2.$$

Dans ce cas on dit que f est fortement convexe (elliptique ou uniformément convexe) de module r sur C.

On définit aussi une fonction concave, strictement concave fortement concave de module $r > 0$ sur C.

Définition 2.1.4 *Soit C un convexe non vide de $\mathbb{R}^n$.*
Une fonction $f : C \to \mathbb{R}$ est concave (respectivement strictement concave, fortement concave de module $r > 0$ sur C, si

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

respectivement

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

$$\forall\, x,\ y \in C, \forall\, \lambda \in [0,1],$$
$$f((1-\lambda)x + \lambda y) \geq (1-\lambda)f(x) + \lambda f(y) + \tfrac{1}{2} r\lambda(1-\lambda)\|y-x\|^2.$$

Donnons ici quelques exemples.
1) Les fonctions suivantes sont convexes sur $\mathbb{R}$:
$f(x) = ax + b$ avec a et $b \in \mathbb{R}$;
$f(x) = e^x$;
$f(x) = x^2$;
$f(x) =| x |$.
2) Les fonctions suivantes sont convexes sur $\mathbb{R}^n$:
$f(x) = \langle a, x\rangle + \alpha$ avec $a \in \mathbb{R}^n$ et $\alpha \in \mathbb{R}$;
$f(x) = \|x\|$;
$f(x) = \|x\|^2$.
3) Les fonctions suivantes sont strictement convexes sur $\mathbb{R}$
$f(x) = x^2$
$f(x) = e^x$
4) La fonction $f(x) = x^2$ est fortement convexe sur $\mathbb{R}$:
5) Les fonction affines sont concaves sur $\mathbb{R}^n$. La fonction $f(x) = \ln x$ est concave sur $\mathbb{R}_+^*$

Proposition 2.1.1 *Une fonction $f : C \to \mathbb{R}$ est concave (respectivement strictement concave, fortement concave sur C si $-f$ est convexe, (respectivement strictement convexe, fortement convexe sur C).*

Dans tout ce quit nous allons considérer que des fonctions définies sur tout l'espace $\mathbb{R}^n$.

Définition 2.1.5 *Soit $f : \mathbb{R}^n \to \mathbb{R}$.*

On appelle épigraphe de f, l'ensemble :

$$\mathrm{epi}(f) = \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \lambda\}.$$

On appelle épigraphe strict de f, l'ensemble :

$$\widetilde{\mathrm{epi}}(f) = \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} : f(x) < \lambda\}.$$

On appelle hypographe de f, l'ensemble :

$$hyp(f) = \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} : f(x) \geq \lambda\}.$$

On appelle hypographe strict de f, l'ensemble :

$$\widetilde{hyp}(f) = \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} : f(x) > \lambda\}.$$

On appelle section de niveau λ de f, l'ensemble

$$S_\lambda(f) = \{x \in \mathbb{R}^n : f(x) \leq \lambda\}.$$

On appelle section stricte de niveau λ de f, l'ensemble

$$\widetilde{S}_\lambda(f) = \{x \in \mathbb{R}^n : f(x) < \lambda\}.$$

On a une caractérisation géométrique de la convexité d'une fonction.

Proposition 2.1.2 *Soit $f : \mathbb{R}^n \to \mathbb{R}$. Les propositions suivantes sont équivalentes :*
i) f est convexe ;
ii) L'épigraphe de f, $(\mathrm{epi}(f))$, est convexe ;
iii) L'épigraphe strict de f est convexe.

La démonstration est immédiate.

De façon analogue

Proposition 2.1.3 *Soit $f : \mathbb{R}^n \to \mathbb{R}$. Les propositions suivantes sont équivalentes :*
i) f est concave ;
ii) L'hypographe de f, $(hyp(f))$, est convexe ;
iii) L'hypographe strict de f est convexe.

On a les caractérisations suivantes :

Proposition 2.1.4 *Soit $f : \mathbb{R}^n \to \mathbb{R}$. On a les équivalences suivantes :*
i) f est convexe ;
ii) Pour toute combinaison linéaire convexe $x = \sum_{i=1}^k \lambda_i x^i$, on a :

$$f(\sum_{i=1}^k \lambda_i x^i) \leq \sum_{i=1}^k \lambda_i f(x^i).$$

Proposition 2.1.5 *$f : \mathbb{R}^n \to \mathbb{R}$ est convexe si et seulement si pour toute droite $D \subset \mathbb{R}^n$, la restriction de f à D est convexe. C'est-à-dire, pour tout a et d dans $\mathbb{R}^n$, la fonction $\varphi_{a,d}$ définie sur $\mathbb{R}$ par $\varphi_{a,d}(t) = f(a + td)$ est convexe.*

Preuve : Supposons f convexe. Soit $a,\ d \in \mathbb{R}^n$, montrons que φ est convexe.
Soient t_1 et t_2 deux réels et $\lambda \in [0, 1]$. On a

$$\begin{aligned}
\varphi_{a,d}(\lambda t_1 + (1-\lambda)t_2) &= f(a + (\lambda t_1 + (1-\lambda)t_2)d) \\
&= f(\lambda a + (1-\lambda)a + \lambda t_1 d + (1-\lambda)t_2 d) \\
&= f(\lambda(a + t_1 d) + (1-\lambda)(a + t_2 d)) \\
&\leq \lambda f(a + t_1 d) + (1-\lambda)f(a + t_2 d) \\
&= \lambda \varphi_{a,d}(t_1) + (1-\lambda)\varphi_{a,d}(t_2)
\end{aligned}$$

d'où la convexité de $\varphi_{a,d}$.

Réciproquement supposons que pour tous $a,\ d \in \mathbb{R}^n$, la fonction φ définie sur $\mathbb{R}$ par $\varphi(t) = f(a + td)$ est convexe. Montrons que f est convexe.
Soient $x,\ y \in \mathbb{R}^n$ et $\lambda \in [0,\ 1]$. On a

$$\begin{aligned}
f(\lambda x + (1-\lambda)y) = f(y + \lambda(x-y)) &= \varphi_{y,x-y}(\lambda) \\
&= \varphi_{y,x-y}(\lambda \times 1 + (1-\lambda) \times 0).
\end{aligned}$$

Comme $\varphi_{y,x-y}$ est convexe, on a

$$\begin{aligned}
\varphi_{y,x-y}(\lambda \times 1 + (1-\lambda) \times 0) &\leq \lambda \varphi_{y,x-y}(1) + (1-\lambda)\varphi_{y,x-y}(0) \\
&= \lambda f(x) + (1-\lambda)f(y)
\end{aligned}$$

d'où la convexité de f. $\square$

Proposition 2.1.6 *Si $f : \mathbb{R}^n \to \mathbb{R}$ est convexe alors pour $a \in \mathbb{R}^n$ et $d \in \mathbb{R}^n$, l'application définie sur $]0, +\infty[$ par*

$$\varphi(t) = \frac{f(a + td) - f(a)}{t}$$

est croissante.

Preuve : Soient $0 < t_1 \leq t_2$. On a alors $0 < \frac{t_1}{t_2} \leq 1$. Donc

$$f(a + t_1 d) = f((1 - \frac{t_1}{t_2})a + \frac{t_1}{t_2}(a + t_2 d)) \leq (1 - \frac{t_1}{t_2})f(a) + \frac{t_1}{t_2}f(a + t_2 d).$$

Il s'ensuit alors que

$$\frac{f(a + t_1 d) - f(a)}{t_1} \leq \frac{f(a + t_2 d) - f(a)}{t_2}.$$

Ce qui prouve la proposition. □

On a aussi la proposition suivante.

Proposition 2.1.7 *Si $f : \mathbb{R}^n \to \mathbb{R}$ est convexe alors les sections de niveau $S_\lambda(f)$, $\lambda \in \mathbb{R}$ sont convexes.*

Preuve : Soit $\lambda \in \mathbb{R}$ et x, y deux éléments de $S_\lambda(f)$ et $\alpha \in [0, 1]$.
La convexité de f et la définition de $S_\lambda(f)$ nous donne :

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

Donc $(1 - \alpha)x + \alpha y \in S_\lambda(f)$ qui est donc convexe. □

Remarque 2.1.1 *La réciproque de cette proposition n'est pas vraie.*

On a dans la proposition ci-dessous une caractérisation de la forte convexité.

Proposition 2.1.8 *Une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est fortement convexe de module r si et seulement si la fonction g définie sur $\mathbb{R}^n$ par $g(x) = f(x) - \frac{1}{2}r\|x - a\|^2 \quad (a \in \mathbb{R}^n)$ est convexe.*

Preuve : La fonction g est convexe si et seulement si

$$\forall\, x,\ y \in \mathbb{R}^n, \forall\, \lambda \in]0, 1[,$$

$$g((1 - \lambda)x + \lambda y) \leq (1 - \lambda)g(x) + \lambda g(y). \tag{2.1}$$

Posons

$$\mu = \|(1 - \lambda)x + \lambda y - a\|^2 - (1 - \lambda)\|x - a\|^2 - \lambda\|y - a\|^2.$$

La condition (2.1) est alors équivalente à :

$$f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y) + \frac{1}{2}r\mu.$$

Or

$$\begin{aligned}
\mu &= \|(1-\lambda)x+\lambda y-a\|^2-(1-\lambda)\|x-a\|^2-\lambda\|y-a\|^2\\
&= \|(1-\lambda)(x-a)+\lambda(y-a)\|^2-(1-\lambda)\|x-a\|^2-\lambda\|y-a\|^2\\
&= (1-\lambda)^2\|x-a\|^2+\lambda^2\|y-a\|^2+2\lambda(1-\lambda)\langle x-a,y-a\rangle\\
&\qquad -(1-\lambda)\|x-a\|^2-\lambda\|y-a\|^2\\
&= -\lambda(1-\lambda)\left(\|x-a\|^2+\|y-a\|^2-2\langle x-a,y-a\rangle\right)\\
&= -\lambda(1-\lambda)\|y-x\|^2.
\end{aligned}$$

La condition (2.1) est donc encore équivalente à :

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

Ce qui signifie que la fonction f est fortement convexe. □

2.2 Opérations sur les fonctions convexes

On montre facilement que

Proposition 2.2.1 *Si $f:\mathbb{R}^n\to\mathbb{R}$ est convexe et $\varphi:\mathbb{R}\to\mathbb{R}$ est convexe et croissante alors la fonction $h=\varphi\circ f$ est convexe.*

On en déduit alors que

Proposition 2.2.2 *Si $f:\mathbb{R}^n\to\mathbb{R}$ est concave et $\varphi:\mathbb{R}\to\mathbb{R}$ est concave et croissante alors la fonction $h=\varphi\circ f$ est concave.*

On peut aussi montrer que :

Proposition 2.2.3 *Si $f:\mathbb{R}^n\to\mathbb{R}$ est strictement convexe et $\varphi:\mathbb{R}\to\mathbb{R}$ est convexe et strictement croissante alors la fonction $h=\varphi\circ f$ est strictement convexe.*

On tire les conséquences suivantes :

Corollaire 2.2.1 *Si $f:\mathbb{R}^n\to\mathbb{R}$ est convexe alors la fonction e^f est convexe.*

Corollaire 2.2.2 *Si $f:\mathbb{R}^n\to\mathbb{R}_+^*$ est concave alors la fonction $\frac{1}{f}$ est convexe.*

Proposition 2.2.4 *Soient $f_i:\mathbb{R}^n\to\mathbb{R}$, $i=1,\cdots,p$ des fonctions convexes et $f=\sum_{i=1}^p\alpha_i f_i$ avec $\alpha_i>0$, pour $i=1,\cdots,p$, alors :*

1) la fonction f est convexe ;

2) si au moins l'une des fonctions $f_1,f_2,\cdots,f_p$ est strictement convexe, la fonction f est strictement convexe ;

3) si au moins l'une des fonctions $f_1,f_2,\cdots,f_p$ est fortement convexe, la fonction f est fortement convexe ;

Proposition 2.2.5 *L'enveloppe supérieure d'une famille de fonctions convexes est convexe. Autrement dit, si $\{f_i\}_{i\in I}$ est une famille quelconque de fonctions convexes définies sur $\mathbb{R}^n$ et à valeurs dans $\mathbb{R}$, alors la fonction f définie par $f(x) = \sup_{i\in I} f_i(x)$ est une fonction convexe.*

Preuve : Soient x et y deux éléments de $\mathbb{R}^n$ et $\lambda \in [0, 1]$. Comme pour tout $i \in I$, f_i est convexe, on a

$$\begin{aligned} \forall\, i \in I,\ f_i((1-\lambda)x + \lambda y) &\leq (1-\lambda)f_i(x) + \lambda f_i(y) \\ &\leq (1-\lambda)\sup_{k\in I} f_k(x) + \lambda \sup_{k\in I} f_k(y). \end{aligned}$$

Donc

$$\sup_{i\in I} f_i((1-\lambda)x + \lambda y) \leq (1-\lambda)\sup_{k\in I} f_k(x) + \lambda \sup_{k\in I} f_k(y).$$

Ce qui signifie que f est convexe. □

Remarque 2.2.1 *On peut démontrer ce résultat en vérifiant que*

$$\mathrm{epi}(f) = \cap_{i\in I}\mathrm{epi}(f_i).$$

Et comme les f_i sont convexes, leurs épigraphes sont convexes et donc l'épigraphe de f aussi. Par suite f est convexe.

2.3 Quelques fonctions convexes particulières

Définition 2.3.1 *Soit C un sous ensemble non vide de $\mathbb{R}^n$: on appelle fonction support de C la fonction notée σ_C ou $\sigma(.,C)$ définie sur $\mathbb{R}^n$ par :*

$$\sigma_C(x) = \sigma(x, C) = \sup_y \left[\langle x, y\rangle : y \in C\right].$$

Proposition 2.3.1 *La fonction support est une fonction convexe.*

Définition 2.3.2 *Soit C un convexe non vide de $\mathbb{R}^n$. On appelle fonction distance Euclidienne de C ou fonction distance par rapport à C, la fonction notée d_C ou $\mathrm{d}(.,C)$ définie sur $\mathbb{R}^n$ par :*

$$\mathrm{d}_C(x) = \mathrm{d}(x, C) = \inf_y \left[\|x - y\| : y \in C\right].$$

Proposition 2.3.2 *La fonction distance est une fonction convexe.*

Preuve : On considère la fonction h définie par : $h(x, y) = \|x - y\|$.
Elle est convexe en (x, y) et comme $\mathrm{d}_C(x) = \inf_y h(x, y)$, alors la fonction distance est convexe. □

On montre que

Proposition 2.3.3 *Si C est un sous-ensemble fermé non vide de $\mathbb{R}^n$, C est convexe si et seulement si La fonction distance par rapport à C est convexe.*

2.4 Caractérisation des fonctions convexes différentiables

Dans les résultats qui suivent, nous donnons une caractérisation de la convexité dans le cas differentiable.

Théorème 2.4.1 *Soit $f : \mathbb{R}^n \to \mathbb{R}$ differentiable.*
On a les équivalences suivantes :
1) f est convexe sur $\mathbb{R}^n$
2) $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \ \forall\, x, y \in \mathbb{R}^n$.
3) $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0, \ \forall\, x, y \in \mathbb{R}^n$.

Preuve : 1)$\Rightarrow$ 2) Soient x, y dans $\mathbb{R}^n$ et $\lambda \in]0, 1[$. Comme f est convexe, alors on a

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

Ce qui donne

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

En passant à la limite vers 0, on obtient :

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

D'où la proposition 2).

2)$\Rightarrow$ 1) On sait par hypothèse que

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \ \forall\, (x, y) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Soient x, et y dans $\mathbb{R}^n$ et $\lambda \in [0, 1]$. En considérant respectivement les couples $(x + \lambda(y - x), x)$ et $(x + \lambda(y - x), y)$, on a :

$$f(x) \geq f(x + \lambda(y - x)) - \lambda\langle \nabla f(x + \lambda(y - x)), y - x \rangle \tag{2.2}$$

et

$$f(y) \geq f(x + \lambda(y - x)) + (1 - \lambda)\langle \nabla f(x + \lambda(y - x), y - x \rangle \tag{2.3}$$

On multiplie (2.2) par $(1 - \lambda)$ et (2.3) par λ et on fait la somme des deux résultats. On obtient alors

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

Ce qui prouve que f est convexe.

2)$\Rightarrow$ 3) Soient x et y dans $\mathbb{R}^n$. On a

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \tag{2.4}$$

et

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \tag{2.5}$$

En considérant la somme de (2.4) et de (2.5), on obtient

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

Montrons à présent que la proposition 3) implique 2).

3)$\Rightarrow$ 2) Soient x et y dans $\mathbb{R}^n$. Comme f est differentiable alors :

$$\exists\, z \in]x,y[\ :\ f(y) - f(x) = \langle \nabla f(z), y - x\rangle \tag{2.6}$$

Comme $z \in]x,y[$, il existe $\lambda \in]0,1[$ tel que $z = x + \lambda(y - x)$.
D'après la proposition 3), on a :

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

Or $z - x = \lambda(y - x)$. Il vient donc

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

Soit

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

car $\lambda \in]0,1[$. C'est-à-dire

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

En utilisant (2.6), on obtient

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

D'où la proposition 2). Ce qui termine la démonstration □

On a des résultats similaires pour la stricte convexité.

Théorème 2.4.2 *Soit $f : \mathbb{R}^n \to \mathbb{R}$ differentiable.*
On a les équivalences suivantes :
1) f est strictement convexe sur $\mathbb{R}^n$
2) $f(y) > f(x) + \langle \nabla f(x), y - x\rangle,\ \forall\, x, y \in \mathbb{R}^n,\ x \neq y$.
3) $\langle \nabla f(y) - \nabla f(x), y - x\rangle > 0,\ \forall\, x, y \in \mathbb{R}^n,\ x \neq y$.

Donnons à présent les résultats concernant la forte convexité.

Théorème 2.4.3 *Soit $f : \mathbb{R}^n \to \mathbb{R}$ differentiable. On a les équivalences suivantes :*
1) f est fortement convexe de module $r > 0$ sur $\mathbb{R}^n$
2) $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{r}{2}\|y - x\|^2,\ \forall\, x, y \in \mathbb{R}^n$.
3) $\langle \nabla f(y) - \nabla f(x), y - x\rangle \geq r\|y - x\|^2,\ \forall\, x, y \in \mathbb{R}^n$.

Preuve : 1)$\Rightarrow$ 2) Soient x, y dans $\mathbb{R}^n$ et $\lambda \in]0,1[$. Comme f est fortement convexe de module r, alors on a

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

Ce qui donne

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

En passant à la limite, on obtient :

$$\langle \nabla f(x), y - x\rangle \leq f(y) - f(x) - \frac{r}{2}\|y - x\|^2.$$

D'où la proposition 2).

2)$\Rightarrow$ 1) On sait par hypothèse que

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{r}{2}\|y - x\|^2, \ \forall\, (x, y) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Soient x, et y dans $\mathbb{R}^n$ et $\lambda \in [0, 1]$. En considérant respectivement les couples $(x + \lambda(y - x), x)$ et $(x + \lambda(y - x), y)$, on a :

$$f(x) \geq f(x + \lambda(y - x)) - \lambda\langle \nabla f(x + \lambda(y - x)), y - x \rangle + \frac{r}{2}\lambda^2\|y - x\|^2 \tag{2.7}$$

et

$$f(y) \geq f(x + \lambda(y - x) + (1 - \lambda)\langle \nabla f(x + \lambda(y - x)), y - x \rangle + \frac{r}{2}(1 - \lambda)^2\|y - x\|^2 \tag{2.8}$$

On multiplie (2.7) par $(1 - \lambda)$ et (2.8) par λ et on fait la somme des deux résultats. On obtient alors

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

Ce qui prouve que f est fortement convexe de module r.

2)$\Rightarrow$ 3) Soit x et y dans $\mathbb{R}^n$. On a

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{r}{2}\|y - x\|^2 \tag{2.9}$$

et

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{r}{2}\|y - x\|^2 \tag{2.10}$$

En considérant la somme de (2.9) et de (2.10), on obtient

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq r\|y - x\|^2.$$

3)$\Rightarrow$ 2) Soient x et y dans $\mathbb{R}^n$. Considérons la fonction φ définie sur $[0, 1]$ et à valeurs dans $\mathbb{R}$ par :

$$\varphi(t) = f(x + t(y - x)) - f(x) - t\langle \nabla f(x), y - x \rangle - \frac{t^2}{2}r\|y - x\|^2.$$

Par hypothèse, on a :

$$\langle \nabla f(x + t(y - x)) - \nabla f(x), x + t(y - x) - x \rangle \geq r\|x + t(y - x) - x\|^2.$$

Soit

$$t\langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle \geq rt^2\|y - x\|^2.$$

Donc pour $t > 0$, on a :

$$\langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle \geq rt\|y - x\|^2. \tag{2.11}$$

Comme f est differentiable, la fonction φ est derivable et pour $t \in]0, 1[$, on a :

$$\varphi'(t) = \langle \nabla f(x + t(y - x)) - \nabla f(x), y - x \rangle - rt\|y - x\|^2.$$

En considérant (2.11), il vient que $\varphi'(t)$ est positif pour tout $t \in]0,1]$. Donc la fonction φ est croissante sur cet intervalle. On a alors $\varphi(1) \geq \varphi(0) = 0$. Soit

$$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle + \frac{r}{2}\|y-x\|^2.$$

D'où le théorème. □

Dans le cas où la fonction est deux fois differentiable, on a aussi les caractérisations suivantes.

Théorème 2.4.4 *Soit $f : \mathbb{R}^n \to \mathbb{R}$ deux fois differentiable. On a les équivalences suivantes :*
1) f est convexe sur $\mathbb{R}^n$
2) Pour tout $x \in \mathbb{R}^n$, $\langle \nabla^2 f(x)h, h \rangle \geq 0 \; \forall\, h \in \mathbb{R}^n$.

Preuve : Soit $x \in \mathbb{R}^n$. On sait que pour tout $h \in \mathbb{R}^n$ et $t \in \mathbb{R}$, suffisamment petit, on a

$$f(x+th) = f(x) + t\langle \nabla f(x), h \rangle + \frac{t^2}{2}\langle \nabla^2 f(x)h, h \rangle + t^2\|h\|^2\varepsilon(t).$$

Par hypothèse, la fonction f est convexe, donc on a pour tout $h \in \mathbb{R}^n$ et $t \in \mathbb{R}$,

$$f(x+th) \geq f(x) + t\langle \nabla f(x), h \rangle.$$

Donc

$$f(x) + t\langle \nabla f(x), h \rangle + \frac{t^2}{2}\langle \nabla^2 f(x)h, h \rangle + t^2\|h\|^2\varepsilon(t) \geq f(x) + t\langle \nabla f(x), h \rangle.$$

Ce qui donne

$$\langle \nabla^2 f(x)h, h \rangle + \|h\|^2\varepsilon(t) \geq 0.$$

Comme la fonction ε tend vers 0 quand t tend vers 0, on obtient :

$$\langle \nabla^2 f(x)h, h \rangle \geq 0.$$

On conclut que la matrice hessienne $\nabla^2 f(x)$ est semi définie positive.

Réciproquement supposons que pour tout $x \in \mathbb{R}^n$, $\nabla^2 f(x)$ est semi définie positive.

Soit $x \in \mathbb{R}^n$. On sait que pour tout $y \in \mathbb{R}^n$, on a

$$f(y) = f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2}\langle \nabla^2 f(z)(y-x), y-x \rangle,$$

avec $z \in]x,y[$. Comme par hypothèse la matrice $\nabla^2 f(z)$ est semi définie positive, alors on a

$$\langle \nabla^2 f(z)(y-x), y-x \rangle \geq 0.$$

Ce qui implique que

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

Par suite la fonction f est convexe. □

Le résultat qui suit concerne la stricte convexité.

Théorème 2.4.5 *Soit $f : \mathbb{R}^n \to \mathbb{R}$ deux fois differentiable. Si pour tout $x \in \mathbb{R}^n$, on a $\langle \nabla^2 f(x)h, h\rangle > 0 \; \forall\, h \in \mathbb{R}^n$, $h \neq 0$ alors f est strictement convexe sur $\mathbb{R}^n$.*

Preuve : Soit x, $y \in \mathbb{R}^n$ avec $x \neq y$. On a

$$f(y) = f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}\langle \nabla^2 f(z)(y-x), y-x\rangle,$$

avec $z \in]x, y[$. Comme par hypothèse la matrice $\nabla^2 f(z)$ est définie positive, alors on a $\langle \nabla^2 f(z)(y-x), y-x\rangle > 0$ car $x \neq y$. On obtient alors

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

Ce qui signifie que f est strictement convexe. □

Remarque 2.4.1 *Il faut signaler que la réciproque de ce résultat n'est pas vraie. On peut considérer la fonction φ définie sur $\mathbb{R}$ suivante : $\varphi(t) = t^4$. Cette fonction est strictement convexe mais sa dérivée seconde en 0 est nulle.*

On a ici une caractérisation de la forte convexité.

Théorème 2.4.6 *Soit $f : \mathbb{R}^n \to \mathbb{R}$ deux fois differentiable. On a les équivalences suivantes :*
1) f est fortement convexe de module $r > 0$ sur $\mathbb{R}^n$
2) Pour tout $x \in \mathbb{R}^n$, $\langle \nabla^2 f(x)h, h\rangle \geq r\|h\|^2 \; \forall\, h \in \mathbb{R}^n$.

Preuve : Supposons f fortement convexe de module r. Alors comme f est differentiable, on a pour tout x, $y \in \mathbb{R}^n$,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{r}{2}\|y - x\|^2, \; \forall\, x, y \in \mathbb{R}^n.$$

Soit $x \in \mathbb{R}^n$; pour tout $h \in \mathbb{R}^n$ et t suffisamment petit, on a

$$f(x + th) = f(x) + t\langle \nabla f(x), h\rangle + \frac{t^2}{2}\langle \nabla^2 f(x)h, h\rangle + t^2\|h\|^2\varepsilon(t).$$

avec ε continue et tendant vers 0 quand t tend vers 0. Donc

$$f(x) + t\langle \nabla f(x), h\rangle + \frac{t^2}{2}\langle \nabla^2 f(x)h, h\rangle + t^2\|h\|^2\varepsilon(t) \geq f(x) + t\langle \nabla f(x), h\rangle + \frac{r}{2}t^2\|h\|^2.$$

Soit

$$\frac{1}{2}\langle \nabla^2 f(x)h, h\rangle + \|h\|^2\varepsilon(t) \geq \frac{r}{2}\|h\|^2.$$

En passant à la limite, on obtient $\langle \nabla^2 f(x)h, h\rangle \geq r\|h\|^2$.

Réciproquement, supposons que pour tout $x \in \mathbb{R}^n$,

$$\langle \nabla^2 f(x)h, h\rangle \geq r\|h\|^2 \; \forall\, h \in \mathbb{R}^n.$$

On a pour tout x, $y \in \mathbb{R}^n$, la relation

$$f(y) = f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}\langle \nabla^2 f(z)(y-x), y-x\rangle,$$

avec $z \in]x, y[$. Donc on a

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}r\|y - x\|^2.$$

Ce qui signifie que f est fortement convexe de module r sur $\mathbb{R}^n$. □

Chapitre 3

Annexes

3.1 Rappels et compléments de calcul différentiel

Dans cette partie on se placera toujours dans un espace vectoriel réel de dimension finie n que l'on identifie à $\mathbb{R}^n$.

3.1.1 Cadre et notation

n et m sont des entiers naturels supérieurs ou égaux à 1. Par convention les vecteurs de $\mathbb{R}^n$ sont des vecteurs colonnes. On note $\langle \cdot, \cdot \rangle$ le produit scalaire canonique et $\| \cdot \|$ la norme euclidienne associée. On note $\mathcal{M}_{m,n}(\mathbb{R})$ l'ensemble des matrices de taille $m \times n$ à coefficients réels et $\mathcal{M}_n(\mathbb{R}) = \mathcal{M}_{n,n}(\mathbb{R})$ l'ensemble des matrices carrées d'ordre n à coefficients réels. La transposée d'une matrice A est notée A^T. On a donc pour tous x, $y \in \mathbb{R}^n$,

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i, \quad \langle Ax, y \rangle = \langle x, A^T y \rangle = x^T A^T y.$$

On désigne par $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ l'ensemble des applications linéaires continues de $\mathbb{R}^n$ dans $\mathbb{R}^m$.

On considère dans cette partie Ω un ouvert de $\mathbb{R}^n$.

3.1.2 Dérivée directionnelle

Définition 3.1.1 *Soit $f : \Omega \longrightarrow \mathbb{R}^m$. Soit $a \in \Omega$ et $v \in \mathbb{R}^n$. On dit que f admet une dérivée au point a suivant a direction v si :*

$$\exists\, f'(a; v) \in \mathbb{R}^m \text{ tel que } \quad \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t} = f'(a; v).$$

Dans ce cas $f'(a; v)$ est appelée la dérivée de f au point a suivant la direction v.

Définition 3.1.2 *Soit $f : \Omega \longrightarrow \mathbb{R}^m$. Si f admet en $a \in \Omega$ une dérivée suivant le $i^{\text{ème}}$ vecteur e_i de la base canonique de $\mathbb{R}^n$, on dit que f admet une dérivée partielle par rapport à la variable x_i et on note $\frac{\partial f}{\partial x_i}(a)$ ou $D_i f(a)$ la dérivée de f au point a suivant la direction e_i.*

Lorsque f ne depend que de 2 ou 3 variables, on utilise souvent la notation (x, y) (resp. (x, y, z)) au lieu de (x_1, x_2) (resp (x_1, x_2, x_3)) et les dérivées partielles sont notées $\frac{\partial f}{\partial x}(a)$, $\frac{\partial f}{\partial y}(a)$ (resp. $\frac{\partial f}{\partial x}(a)$, $\frac{\partial f}{\partial y}(a)$, $\frac{\partial f}{\partial z}(a)$)

Définition 3.1.3 *Soit $f : \Omega \longrightarrow \mathbb{R}^m$, $a \in \Omega$ et $i \in \{1, \cdots, n\}$. Soit*

$$\Omega_i^a = \{t \in \mathbb{R} : (a_1,\, a_2,\, \cdots,\, a_{i-1},\, t,\, a_{i+1},\, \cdots,\, a_n) \in \Omega\}.$$

On appelle $i^{\text{ème}}$ application partielle de f en a l'application

$$\begin{aligned} \Omega_i^a &\longrightarrow \mathbb{R}^m \\ t &\longmapsto f(a_1,\, a_2,\, \cdots,\, a_{i-1},\, t,\, a_{i+1},\, \cdots,\, a_n) \end{aligned}$$

Signalons au passage que si f est continue alors toute les applications partielles qui lui sont associées le sont également. La réciproque est fausse.

Remarque 3.1.1 *Le vecteur $\frac{\partial f}{\partial x_i}(a)$ n'est autre que la dérivée en a_i de la $i^{\text{ème}}$ application partielle de f en a.*

3.1.3 Différentiabilité

Définition 3.1.4 *Soit $f : \Omega \longrightarrow \mathbb{R}^m$. On dit que f est dite différentiable au point $a \in \Omega$ si $\exists\, L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ tel que :*

$$\lim_{u \to 0} \frac{f(a+u) - f(a) - L(u)}{\|u\|} = 0;$$

c'est-à-dire :

$$\lim_{x \to a} \frac{f(x) - f(a) - L(x-a)}{\|x-a\|} = 0.$$

On montre que si f est différentiable en a, l'application linéaire L est unique : on la note $f'(a)$ ou df_a et elle est appelée la différentielle de f au point a et $Df(a)$ la matrice associée est appelée matrice jacobienne.

On montre facilement les équivalences suivantes.

Proposition 3.1.1 *f est différentiable au point a si et seulement si on a l'une des conditions équivalentes suivantes.*

$$a) \begin{cases} \exists\, f'(a) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) : \forall \varepsilon \in \mathbb{R}_+^* \ \exists \rho \in \mathbb{R}_+^* : \forall\, u \in B(0, \rho) \\ \|f(a+u) - f(a) - f'(a)(u)\| < \varepsilon \|u\|. \end{cases}$$

$$b) \begin{cases} \exists\, f'(a) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) : \ f(a+u) = f(u) + f'(a)(u) + \|u\|\varepsilon(u) \\ \text{avec } \lim_{u \to 0} \varepsilon(u) = 0. \end{cases}$$

Si f est différentiable en a, la fonction affine $f(x) = f(a) + f'(a)(x-a)$ est l'approximation à l'ordre 1 de f au point a. La différentielle peut être calculée à partir des dérivées partielles des composantes de f. On montre que :

Proposition 3.1.2 *Si $f : \Omega \longrightarrow \mathbb{R}^m$ est différentiable en $a \in \Omega$, alors f admet des dérivées suivant toutes les directions et on a :*

$$\forall\, v \in \mathbb{R}^n \quad f'(a)(v) = Df(a)(v) = f'(a; v) = \lim_{t\to 0} \frac{f(a+tv) - f(a)}{t}.$$

Si on note $f_1, \cdots, f_n$ les applications coordonnées ou applications composantes de f, la matrice jaobienne s'écrit :

$$Df(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_i n}(a) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix} \in \mathcal{M}_{m,n}(\mathbb{R}).$$

Remarque 3.1.2 *La réciproque est fausse. Il existe des fonctions pour lesquelles toutes les dérivées directionnelles existent et qui ne sont pas différentiables.*

Définition 3.1.5 *Si $n = m$, le déterminant de la matrice jacobienne est appelé le jacobien. On le note $Jac_f(a)$ ou $\frac{D(f_1,\cdots f_m)}{D(x_1,\cdots x_n)}(a)$.*

Définition 3.1.6 *Pour une fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, différentiable, la matrice jacobienne est une matrice ligne*

$$Dfa) = \left(\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a)\right).$$

La transposée de cette matrice ligne est un vecteur appelé gradient de f. On le note ∇f.

On a $\nabla f(a) = Df(a)^T$ c'est-à-dire que :

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a)\right)^T.$$

On a donc pour tout $v \in \mathbb{R}^n$:

$$f'(a; v) = f'(a)(v) = Df()v = \nabla f(a)^T v = \langle \nabla f(a), v\rangle.$$

$\nabla f(x)$ s'interprête comme le vecteur de la plus forte augmentation de f au voisinage de x. En particulier, $\nabla f(x)$ est orthogonal au ligne de niveaux de la fonction f.

On a le résultat suivant :

Proposition 3.1.3 *Soit $f : \Omega \longrightarrow \mathbb{R}^m$ Si au point $a \in \Omega$ toutes les dérivées partielles $\frac{\partial f}{\partial x_i}(a)$, $1 \le i \le i \le n$, existent et si les fonctions $x \longmapsto \frac{\partial f}{\partial x_i}(x)$, $1 \le i \le i \le n$, sont continues au voisinage de a alors f est différentiable en a.*

Définition 3.1.7 *On dit que $f : \Omega \longrightarrow \mathbb{R}^m$ est continûment différentiable sur Ω on dit aussi que f est de classe C^1 sur Ω et on note $f \in \mathcal{C}^1(\Omega, \mathbb{R}^m)$ si elle est différentiable en tout point de Ω et si l'application $f' : \Omega \longrightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ est continue sur Ω.*

Proposition 3.1.4 *La fonction $f : \Omega \longrightarrow \mathbb{R}^m$ est continûment différentiable sur Ω si et seulement si f admet en tout point de Ω des dérivées partielles continues sur Ω.*

On a les propriétés suivantes :

Proposition 3.1.5 *Si f et g sont des applications de $\Omega \longrightarrow \mathbb{R}^m$ différentiable en $a \in \Omega$ (resp. continûment différentiables sur Ω) alors pour tout α, $\beta \in \mathbb{R}$, $\alpha f + \beta g$ est différentiable en $a \in \Omega$ (resp. continûment différentiables sur Ω) et on a $(\alpha f + \beta g)'(a) = \alpha f'(a) + \beta g'(a)$ (resp. $(\alpha f + \beta g)'(x) = \alpha f'(x) + \beta g'(x)$ pour tout $x \in \Omega$).*

Proposition 3.1.6 *Soient $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ et $g : \mathbb{R}^m \longrightarrow \mathbb{R}^p$ des applications différentiables. Alors $h = g \circ f : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ est différentiable et on a :*

$$\forall\, a \in \mathbb{R}^n, \quad h'(a) = g'(f(a)) \circ f'(a) \text{ et } Dh(a) = Dg(f(a))Df(a).$$

Proposition 3.1.7 *(Différentielle d'un produit) Soient f et g deux applications de Ω dans $\mathbb{R}$ différentiables en un point a de Ω. Alors l'application h définie sur Ω par $h(x) = f(x)g(x)$ est différentiable en a, et on a :*

$$h'(a) = f(a)g'(a) + g(a)f'(a).$$

Proposition 3.1.8 *(Différentielle d'un quotient) Soient f et g deux applications de Ω dans $\mathbb{R}$ différentiables en un point a de Ω. Si g ne s'annule pas au voisinage de a, alors l'application h définie sur Ω par $h(x) = \frac{f(x)}{g(x)}$ est différentiable en a, et on a :*

$$h'(a) = \frac{g(a)f'(a) - f(a)g'(a)}{g(a)^2}.$$

3.1.4 Différentielle d'ordre deux

Si f est différentiable sur Ω, sa différentielle est une application de Ω dans $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. On peut donc étudier la différentiabilité de f', et en itérant ce processus, définir les différentielles successives.

a) Définitions

Définition 3.1.8 *Supposons différentiable sur Ω. On dit que f est deux fois différentiable en $a \in \Omega$ si f' est différentiable en a. En d'autres termes, s'il existe une application linéaire continue notée $f''(a)$ (ou $D^2 f(a)$) avec $f''(a) : \Omega \subset \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, telle que :*

$$\lim_{h \to 0} \frac{f'(a+u) - f'(a) - f''(a)(u)}{\|u\|} = 0.$$

Remarque 3.1.3 *On a : $f''(x) \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)) = \mathcal{L}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m))$ et*

$$\forall\,(u,v) \in \mathbb{R}^n \times \mathbb{R}^n, \quad f''(x)(u,v) = \lim_{t \to 0} \frac{f'(a+tu)(v) - f'(a)(v)}{t}.$$

On définit :

Définition 3.1.9 *$f : \Omega \longrightarrow \mathbb{R}^m$ est dite :*

i) deux fois différentiable sur Ω si f est 2 fois différentiable en tout point de Ω

ii) deux fois continûment différentiable sur Ω ou classe C^2 sur Ω on dit aussi que $f \in \mathcal{C}^2(\Omega, \mathbb{R}^m)$ si f est 2 fois différentiable en tout point de Ω et $f'' : \Omega \longrightarrow \mathcal{L}(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}^m))$ est continue.

On montre que :

Proposition 3.1.9 *Si f est deux fois différeniable au point $a \in \Omega$, alors $f''(a)$ est bilinéaire et on a :*

$$\forall (u,v) \in \mathbb{R}^n \times \mathbb{R}^n, f''(a)(u,v) = f''(a)(v,u)$$

c'est-à-dire que $f''(a)$ est symétrique.

b) Dérivées partielles d'ordre deux

Dans cette section nous allons nous limiter, pour simplifier, à l'étude des applications définies sur un ouvert de $\mathbb{R}^n$ et à valeurs dans $\mathbb{R}$.

Nous avons déjà vu que $f : \Omega \longrightarrow \mathbb{R}$ est de classe C^1 sur Ω si et seulement si toutes les dérivées partielles $\frac{\partial f}{\partial x_j}$ sont des fonctions continues sur Ω.

Si la $i^{\text{ème}}$ application partielle associée à $\frac{\partial f}{\partial x_j}$ est elle même dérivable en a, sa dérivée est notée $\frac{\partial^2 f}{\partial x_i \partial x_j}(a)$ (ou plus simplement $\frac{\partial^2 f}{\partial x_i^2}$ si $i = j$).

Proposition 3.1.10 *Soit $f : \Omega \longrightarrow \mathbb{R}$ différentiable sur Ω et deux fois différentiable en $a \in \Omega$. Alors toutes les dérivées partielles d'ordre deux e f sont définies et l'on a :*

$$\forall (u,v) \in \mathbb{R}^n \times \mathbb{R}^n, f''(a)(u,v) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) u_i v_j.$$

De cette formule et de la symétrie de $f''(a)$, on déduit :

Théorème 3.1.1 *(de Schawrz) Soit $f : \Omega \longrightarrow \mathbb{R}$ deux fois différentiable en $a \in \Omega$. Alors toutes les dérivées partielles d'ordre deux sont définies en a et l'on a*

$$\frac{\partial^2 f(a)}{\partial x_i \partial x_j} = \frac{\partial^2 f(a)}{\partial x_j \partial x_i}.$$

Définition 3.1.10 *Soit $f : \Omega \longrightarrow \mathbb{R}$ une application dont toutes les dérivées partielles d'ordre deux sont définies au voisinage de $a \in \Omega$ et continues. On appelle matrice hessienne de f au point a, la matrice carrée d'ordre n symétrique :*

$$\nabla^2 f(a) = \left(\frac{\partial^2 f(a)}{\partial x_i \partial x_j} \right)_{1 \leq i,j \leq n}.$$

On a pour tous vecteurs h, $k \in \mathbb{R}^n$,

$$f''(a)(h,k) = \langle \nabla^2 f(a) h, k \rangle = h^T \nabla^2 f(a) k = k^T \nabla^2 f(a) h.$$

Dans ce cours on aura constamment besoin de calculer le gradient et la matrice hessienne de fonctionnelles deux fois différentiables et à valeurs dans $\mathbb{R}$. En pratique on utilise la proposition suivante.

Proposition 3.1.11 *Soit $f : \Omega \longrightarrow \mathbb{R}$ différentiable sur Ω et deux fois différentiable en $x \in \Omega$. Alors la matrice hessienne $\nabla^2 f(x)$ de f au point x est la différentielle de l'application gradient $x \mapsto \nabla f(x)$ au point x : $D(\nabla f) = \nabla^2 f$.*

Exemple 3.1.1

1) Si $f : \mathbb{R}^m \to \mathbb{R}$ est deux fois différentiable alors l'application $g : \mathbb{R}^n \to \mathbb{R}$ définie par $g(x) = f(Ax+b)$ où A est une matrice de $\mathcal{M}_{m,n}(\mathbb{R})$, $b \in \mathbb{R}^m$ est deux fois différentiable, et on a :

$$g'(x) = f'(Ax+b)A, \quad \nabla g(x) = g'(x)^T = A^T f'(Ax+b)^T = A^T \nabla f(Ax+b),$$

$$\nabla^2 g(x) = A^T \nabla^2 f(Ax+b)A$$

2) Si $f : \Omega \longrightarrow \mathbb{R}$ est deux fois différentiable sur Ω et $g : \mathbb{R} \longrightarrow \mathbb{R}$ est deux fois dérivable, alors, $h = g \circ f$ est deux fois différentiable et :

$$h'(x) = g'(f(x))f'(x), \quad \nabla^2 h(x) = g'(f(x))\nabla^2 f(x) + g''(f(x))\nabla f(x)\nabla f(x^T.$$

c) Formules de Taylor

On considère dans ce qui suit la notation suivante

Notation 3.1.1 *Supposons que f est m fois différentiable au point x de Omega. On note :*

$$\forall u \in \mathbb{R}^n, \; f^{(k)}(x)\underbrace{(u, u, \cdots, u)}_{k \text{ fois}} = f^{(k)}(x)u^k.$$

Proposition 3.1.12 *(Formule de Taylor-Lagrange)*

Supposons que $f : \Omega \longrightarrow \mathbb{R}$ est $m+1$ fois différentiable sur Ω, x et y sont tels que $[x, y]$ est inclus dans Ω. Alors il existe $\theta \in]0,1[$ tel que

$$f(y) = f(x) + \sum_{k=1}^{m} \frac{1}{k!} f^{(k)}(x)(y-x)^k + \frac{1}{(m+1)!} f^{(m+1)}(x + \theta(y-x))(y-x)^{m+1}.$$

En particulier,

a) pour $m = 0$ on a :

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

b) pour $m = 1$, on a :

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 f(x + \theta(y-x))(y-x), y-x \rangle.$$

Corollaire 3.1.1 *Supposons que $f : \Omega \longrightarrow \mathbb{R}$ est $m+1$ fois différentiable sur Ω, x et y sont tels que $[x, y] \subset \Omega$ et que $f^{(m+1)}$ soit bornée par un réel $M \geq 0$ sur $[x, y]$. Alors on a l'inégalité suivante :*

$$\left| f(y) - f(x) - \sum_{k=1}^{m} \frac{1}{k!} f^{(k)}(x)(y-x)^k \right| \leq M \frac{\|y-x\|^{m+1}}{(m+1)!}.$$

Proposition 3.1.13 *(Formule de Taylor avec reste intégral)*

Supposons que $f : \Omega \longrightarrow \mathbb{R}$ est de classe $\mathcal{C}^{m+1}$ sur Ω avec $m \geq 0$. Alors pour tout x, $y \in \Omega$ tels que $[x, y] \subset \Omega$, on a

$$f(y) = \sum_{k=0}^{m} \frac{1}{k!} f^{(k)}(x)(y-x)^k + \int_0^1 \frac{(1-t)^m}{m!} \left[f^{(m+1)}(x + t(y-x))(y-x)^{m+1} \right] dt.$$

En particulier

a) pour $m = 0$ on a

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

b) pour $m = 1$, on a

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \langle \nabla^2 f(x + t(y-x))(y-x), y - x \rangle (1-t) dt.$$

Proposition 3.1.14 *(Formule de Taylor-Young)*

Supposons que $f : \Omega \longrightarrow \mathbb{R}$ est de classe $\mathcal{C}^m$ sur Ω avec $m \geq 0$ et $m+1$ fois différentiable au point x de Ω. Alors il existe un voisinage V de 0 et une fonction $\varepsilon ; V \longrightarrow \mathbb{R}$ avec $\lim_{u \to 0} \varepsilon(u) = 0$ tels que

$$\forall\, u \in V, \quad f(x+u) = \sum_{k=0}^{m+1} \frac{1}{k!} f^{(k)}(x) u^k + \|u\|^{m+1} \varepsilon(u).$$

Pour $m = 0$, la formule de Taylor-Young s'écrit :

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \|h\| \varepsilon(h).$$

Pour $m = 1$ la formule de Taylor-Young s'écrit :

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \nabla^2 f(x) h, h \rangle + \|h\|^2 \varepsilon(h).$$

3.2 Matrices symétriques semi-définies positives

3.2.1 Définitions

Dans les résultats qui suivent, nous donnons des caractérisations de la convexité dans le cas différentiable.

Avant de donner ces résultats rappelons la notion de matrice définie positive, semi définie positive. On considère les notations suivantes

$\mathcal{M}_n(\mathbb{R})$ désigne l'ensemble des matrices carrées d'ordre n et $\mathcal{S}_n(\mathbb{R})$ l'ensemble des matrices carrées d'ordre n symétriques à coefficients réels.

Définition 3.2.1 *Soit A une matrice de $\mathcal{M}_n(\mathbb{R})$.*

- On dit que A est semi définie positive, si on a $\langle Ax, x \rangle \geq 0$ pour tout $x \in \mathbb{R}^n$

- On dit que A est définie positive, si on a $\langle Ax, x \rangle > 0$ pour tout $x \in \mathbb{R}^n$ non nul.

Définition 3.2.2 *Soit A une matrice de $\mathcal{M}_n(\mathbb{R})$.*

- On dit que A est semi définie négative, si on a $\langle Ax, x \rangle \leq 0$ pour tout $x \in \mathbb{R}^n$.

- On dit que A est définie négative, si on a $\langle Ax, x \rangle < 0$ pour tout $x \in \mathbb{R}^n$ non nul.

3.2.2 Propriétés

On montre que

Proposition 3.2.1 *Une matrice $A \in \mathcal{M}_n(\mathbb{R})$ est semi définie négative (resp. définie négative) si et seulement si $-A$ semi définie positive (resp. définie positive).*

on montre que

Proposition 3.2.2 *Une matrice $A \in \mathcal{S}_n(\mathbb{R})$ est semi définie positive si et seulement si toutes ses valeurs propres sont positives ou nulles.*

Une matrice $A \in \mathcal{S}_n(\mathbb{R})$ est définie positive si et seulement si toutes ses valeurs propres sont strictement positives.

Proposition 3.2.3 *Une matrice $A \in \mathcal{S}_n(\mathbb{R})$ est semi définie négative si et seulement si toutes ses valeurs propres sont négatives ou nulles.*

Une matrice $A \in \mathcal{S}_n(\mathbb{R})$ est définie négative si et seulement si toutes ses valeurs propres sont strictement négatives.

On sait que le déterminant d'une matrice carrée est égal au produit de ses valeurs propres et sa trace est égale à la somme de ses valeurs propres. Alors on a le corollaire.

Corollaire 3.2.1 *Soit A une matrice carrée d'ordre 2 symétrique et à coefficients réels.*

1) A est semi définie positive si et seulement si son déterminant et sa trace sont positifs ou nuls.

2) A est définie positive si et seulement si son déterminant et sa trace sont strictement positifs.

Définition 3.2.3 *Soit A une matrice de $\mathcal{M}_n(\mathbb{R})$. Les mineurs principaux diagonaux de A sont les n sous-matrices carrées $A_{(p)}$ d'ordre p obtenues en supprimant les $n-p$ dernières lignes et colonnes, $p = 1 \cdots, n$. Ce sont les matrices*

$$A_{(1)} = (a_{11}), A_{(2)} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \cdots, A_{(p)} = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pp} \end{pmatrix}, \cdots, A_{(n)} = A.$$

Proposition 3.2.4 *Une matrice $A \in \mathcal{S}_n(\mathbb{R})$ est définie positive si et seulement si les déterminants des n mineurs principaux diagonaux sont strictement positifs.*

$$\det(A_{(p)}) > 0 \quad \textit{pour} \quad 1 \leq p \leq n.$$

Proposition 3.2.5 *Une matrice $A \in \mathcal{S}_n(\mathbb{R})$ est définie négative si et seulement si on a :*

$$(-1)^p \det(A_{(p)}) > 0 \quad \textit{pour} \quad 1 \leq p \leq n.$$

Si A est semi définie positive alors on a $\det(A_{(p)}) \geq 0$ pour $1 \leq p \leq n$. Mais cette condition nécessaire n'est pas suffisante. Par exemple la matrice symétrique d'ordre trois

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

vérifie $\det(A_{(1)}) > 0$, $\det(A_{(2)}) = \det(A_{(3)}) = 0$. Ses valeurs propres sont 0, 2 et -1. Elle n'est donc semi definie positive ni semi definie négative.

Définition 3.2.4 *Soit A une matrice de $\mathcal{S}_n(\mathbb{R})$. Un mineur principal d'ordre p de A est la sous-matrice de A d'ordre p obtenue en supprimant $n-p$ lignes et les $n-p$ colonnes correspondantes dans A.*

Proposition 3.2.6 *Soit $A \in \mathcal{S}_n(\mathbb{R})$.*

A est semi définie positive si et seulement si les déterminants de tous ses mineurs principaux sont positifs ou nuls.

Proposition 3.2.7 *Soit $A \in \mathcal{S}_n(\mathbb{R})$.*

A est semi définie négative si et seulement si les déterminants de ses mineurs principaux d'ordre k sont alternativement ≤ 0 pour k impair et ≥ 0 pour k pair.

Bibliographie

[1] Bazaraa, Mokhtar S. and Shetty, C. M., 1979. Nonlinear Programming Theory and Algorithms, *John Wiley and Sons.*

[2] Bazaraa, Mokhtar S. and Shetty, C. M., 1976. Foundations of Optimization, *Lecture Notes in Economic and Mathematical Systems, No 122, Springer-Verlag New-York.*

[3] Bergounioux Maïtine, 2001. Optimisation et Contrôle des systèmes linéaires, *Dunod.*

[4] Bertsekas, Dimiri P. 1995. Non Linear Programming, *Athena Scientific.*

[5] Bonnans, J. Fréderic and Shapiro, Alexander 2000. Perturbations Analysis of Opimization Problems, *Springer.*

[6] Culioli, Jean-Christophe, 1994. Introduction à l'optimisation, *Ellipses.*

[7] Hiriart-Urruty, Jean-Baptiste, 1998. Optimisation et Analyse Convexe, *Presse Universitaire de France.*

[8] Hiriart-Urruty, Jean-Baptiste and Lemaréchal, Claude, 1993. Convex Analysis and Minimization algorithms, *Vol I et II Grundlehren der mathematichen Wissenschaften 305 and 306, Springer-Verlag.*

[9] Hiriart-Urruty, Jean-Baptiste, 1996. L'Optimisation, in collection "Que sais-je ?", *Presse Universitaire de France.*

[10] Minoux, Michel, 1983. Programmation mathématique : Théorie et Algorithmes, Vol I, *Dunod.*

[11] Rockafellar, R. Tyrrel, 1970. Convex Analysis, *Princeton University Press, Princeton N. J..*

[12] Roberts, A. Wayne and Varberg, Dale E., 1973. Convex Functions, *Academic Press.*