Chapitre 3

FONCTIONS CONVEXES

3.1 Notations et définitions préliminaires

L'étude des fonctions convexes montrera que celles ci sont continues sur tout l'intérieur de leur domaine de définition et qu'elles sont presque partout différentiables. Une propriété suffisante pour étudier les fonctions convexes est la *semi-continuité inférieure* :

Définition 3.1 *Soit une fonction $f : C \mapsto R$, où C est un ouvert de $I\!R^n$. f est dite semi-continue inférieurement (sci) en $x \in C$ si, pour toute suite $\{x_k\} \subset C$ convergente vers x, on a* $\liminf f(x_k) \geq f(x)$.

Une notion fondamentale est celle de la *dérivée directionnelle* dont on rappelle les définitions :

Définition 3.2 *Soit une fonction $f : C \mapsto R$, où C est un ouvert de $I\!R^n$. La dérivée directionnelle de f en $x \in C$ dans la direction $d \in I\!R^n$ est définie par la limite quand elle existe de :*

$$f'(x; d) = \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}$$

La fonction f sera dite *différentiable* en $x \in C$ si elle possède des dérivées directionnelles dans toutes les directions et si f' est linéaire par rapport à d, i.e. s'il existe un vecteur $\nabla f(x)$ appelé le *gradient* de f en x tel que :

$$f'(x; d) = \langle \nabla f(x), d \rangle$$

Les composantes du gradient sont alors les dérivées partielles de f par rapport à chaque variable $x_j, j = 1, \ldots, n$ notées $\partial f / \partial x_j(x)$.

Si $F = [f_1 \cdots f_p]^T$ est un vecteur de fonctions $f_i : C \mapsto R, i = 1, \ldots, p$ différentiables, le *Jacobien* de la fonction vectorielle de $I\!R^n$ dans $I\!R^p$ est la matrice $JF(x)$ de p lignes et n colonnes dont les lignes sont les gradients $\nabla f_i(x), i = 1, \ldots, p$.

Si une fonction f est deux fois différentiable en x, le *Hessien* est la matrice symétrique $(n \times n)$ dont les éléments sont les dérivées secondes partielles par rapport aux variables x_j et x_k notées $\partial^2 f / \partial x_j \partial x_k, j, k = 1, \ldots, n$.

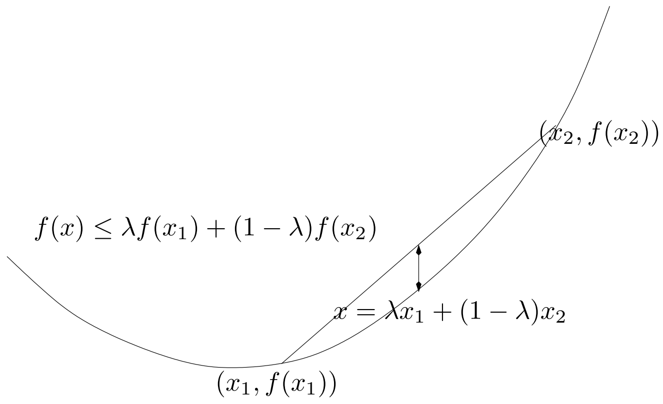

FIGURE 3.1 – Fonction convexe

3.2 Définitions et propriétés

On décrit dans ce chapitre les propriétés des fonctions convexes de $\mathbb{R}^n$.

Définition 3.3 *: Soit une fonction $f : S \mapsto R$, où S est un ensemble convexe non vide de $\mathbb{R}^n$. f est dite convexe sur S si :*

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2), \; \forall x_1 \text{ et } x_2 \in S, \; \lambda \in (0,1)$$

f est dite strictement convexe si l'inégalité précédente est stricte pour $x_1 \neq x_2$.

Définition 3.4 *: f est fortement convexe de constante $\alpha > 0$ si :*

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) - \frac{\alpha}{2}\lambda(1-\lambda)\|x_1 - x_2\|^2$$

On montre facilement qu'une fonction fortement convexe est strictement convexe. On a aussi la caractérisation suivante :

Proposition 3.1 *Soit C un convexe de $\mathbb{R}^n$ et $a \in \mathbb{R}^n$. La fonction $f : C \mapsto \mathbb{R}^n$ est fortement convexe sur C si et seulement si la fonction g définie ci-dessous est convexe :*

$$g(x) = f(x) - \frac{\alpha}{2}\|x - a\|^2$$

Démonstration : Exercice

Si f est convexe, la fonction $g : S \mapsto R$ telle que $g = -f$ est dite *concave* sur S.

Les fonctions affines de $\mathbb{R}^n$ sont bien sûr convexes (elles sont aussi concaves). Comme on le vérifiera plus loin, les fonctions quadratiques convexes de $\mathbb{R}^n$ sont celles qui sont associées à une matrice semi-définie positive. Dans $\mathbb{R}$, des exemples courants de fonctions convexes sont :

$f(x) = x^2$
$f(x) = e^x$
$f(x) = -\log x$, sur $x > 0$
$f(x) = -\sqrt{x}$ sur $x \geq 0$
$f(x) = 1/x$ sur $x > 0$

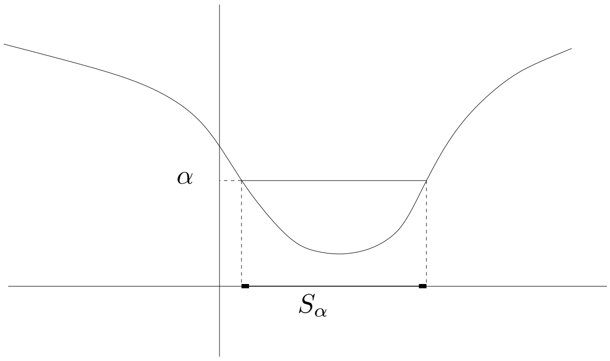

FIGURE 3.2 – Fonction quasiconvexe

$f(x) = |x|$
Le *domaine* de définition de f sera noté $\mathrm{Dom}(f)$, c.a.d. :

$$\mathrm{Dom}(f) = \{x \in I\!R^n | f(x) < +\infty\}$$

On appelle *propres* les fonctions qui ne prennent jamais la valeur $-\infty$ et qui ne sont pas identiques à $+\infty$. Par convention, la fonction convexe f prendra la valeur $+\infty$ en dehors de $\mathrm{Dom}(f)$ et l'opération $\infty - \infty$ est interdite.

On utilisera deux outils de représentation des fonctions de $I\!R^n$:
L'*épigraphe* qui décrit f dans $I\!R^{n+1}$ et les *ensembles de niveaux* qui sont des surfaces de $I\!R^n$.

Définition 3.5 *: L'épigraphe d'une fonction f de $I\!R^n$ est noté epi(f) et est défini par :*

$$epi(f) = \{\begin{pmatrix} x \\ z \end{pmatrix} \in I\!R^{n+1} \mid z \geq f(x)\}$$

L'ensemble de niveau α est l'ensemble des x de $I\!R^n$ tels que $f(x) = \alpha$. On lui associe la section, notée S_α :

$$S_\alpha = \{x \in I\!R^n \mid f(x) \leq \alpha\}$$

L'épigraphe sera particulièrement utile pour transférer les propriétés des fonctions sur les ensembles de $I\!R^{n+1}$. Par exemple, un épigraphe fermé signifie que la fonction est semi-continue inférieurement (sci).

On a la caractérisation fondamentale suivante :

Une fonction est convexe si et seulement si son épigraphe est convexe.

Par contre, s'il est vrai qu'une fonction convexe possède des sections convexes (par convention, l'ensemble vide est convexe), il existe des fonctions non convexes dont toutes les sections sont convexes.

Définition 3.6 *: Une fonction dont toutes les sections sont convexes est dite* quasiconvexe *(Fig. 3.2). Une fonction convexe est donc quasiconvexe.*

Proposition 3.2 *f est quasiconvexe sur un ensemble convexe S si et seulement si :*

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \max\{f(x_1), f(x_2)\}, \forall x_1, x_2 \in S \ \textit{et}\ \lambda \in (0,1)$$

Propriétés des fonctions convexes :
Soient f_1 et f_2 deux fonctions convexes telles que $\mathrm{Dom}(f_1) \cap \mathrm{Dom}(f_2) \neq \emptyset$. Alors :

i) $f_1 + f_2$ est convexe
ii) af_1 est convexe $\forall a \geq 0$
iii) $\sup\{f_1, f_2\}$ est convexe
iv) $\inf_z\{f_1(z) + f_2(x-z)\}$ est convexe

Démonstration :

i) et ii) sont des conséquences immédiates de la définition.

iii) équivaut à $\mathrm{epi}(f_1) \cap \mathrm{epi}(f_2)$ est un ensemble convexe de $I\!R^{n+1}$.

iv) équivaut à $\mathrm{epi}(f_1) + \mathrm{epi}(f_2)$ est un ensemble convexe de $I\!R^{n+1}$. Cette dernière opération est appelée l'*inf-convolution* de f_1 et f_2, notée $f_1 \square f_2$.

$$f_1 \square f_2(x) = \inf_z\{f_1(z) + f_2(x-z)\}$$

En général, la composition de deux fonctions convexes n'est pas convexe. On a par contre le résultat suivant :

Lemme 3.1 *Soit f une fonction convexe sur le sous-ensemble C de $I\!R^n$, et ϕ une fonction convexe non décroissante de $f(C)$ dans $I\!R$. Alors $h = \phi \circ f$ est convexe sur C.*

Démonstration : Exercice.

Composition par une transformation affine :

Soit A une transformation affine de $I\!R^n$ dans $I\!R^p$ et f une fonction convexe de $I\!R^p$ dans $I\!R$. Alors la fonction $f \circ A$:

$$(f \circ A)(x) = f(Ax)$$

est convexe. En effet, l'épigraphe de $f \circ A$ est l'image réciproque de l'épigraphe de f par la transformation affine de $I\!R^{n+1}$ dans $I\!R^{p+1}$: $(x,z) \mapsto (Ax, z)$.

De même, si f est une fonction convexe sur $I\!R^n$, alors la fonction Af définie par :

$$(Af)(x) = \inf\{f(y) \mid Ay = x\}$$

est convexe sur $I\!R^p$. Cette fois-ci, l'épigraphe de Af est l'image directe de l'épigraphe de f par la transformation affine précédente.

Application : Soit une fonction f de $I\!R^n \times I\!R^p$, convexe en $(x,y) \in I\!R^n \times I\!R^p \mapsto f(x,y)$. Alors la fonction

$$x \mapsto h(x) = \inf_{y \in I\!R^p} f(x,y)$$

est convexe sur $I\!R^n$. En effet, elle s'écrit Af où A est l'opérateur de projection de $I\!R^n \times I\!R^p$ sur $I\!R^n$.

3.3 Continuité et différentiabilité

Dans la suite f est une fonction convexe propre définie sur un ouvert convexe C de $\mathbb{R}^n$

Lemme 3.2 (Monotonie des accroissements) *Soit $x \in C$ et d une direction de $\mathbb{R}^n$. Alors, la fonction $q(t) = \frac{f(x+td)-f(x)}{t}$, définie pour des réels t suffisamment petits, est une fonction croissante de t.*

Démonstration Prenons un intervalle $[h, k]$ avec $0 < h \leq k$. On a : $x + hd = (1-\alpha)x + \alpha(x + kd)$ avec $\alpha = \frac{h}{k} < 1$. La convexité implique $f(x + hd) \leq (1 - \frac{h}{k})f(x) + \frac{h}{k}f(x + kd)$ qui s'écrit :

$$\frac{f(x+hd)-f(x)}{h} \leq \frac{f(x+kd)-f(x)}{k}$$

Le même raisonnement s'applique pour des pas négatifs. □
Ce résultat implique, en étudiant les limites à gauche et à droite de $q(t)$ quand t tend vers 0, l'existence de dérivées directionnelles en tout point de C, définies par :

$$f'(x; d) = \lim_{t \downarrow 0} q(t) = \inf_{t>0} q(t)$$

Observez que l'infimum est borné car, C étant un ouvert, il existe $t < 0$ tel que $f'(x; d) \geq q(t)$.

Théorème 3.1 *: Une fonction convexe est continue sur rint$(Dom(f))$, c'est-à-dire en tout point de l'intérieur de $Dom(f)$ relativement à la topologie induite dans aff$\{Dom(f)\}$.*

Démonstration et résultats connexes : cf. Rockafellar, Convex Analysis, chap. 10.

En résumé, une fonction convexe est continue et différentiable presque partout sur l'intérieur de son domaine. Si f est différentiable en x, on a :

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \forall y \in C$$

En effet, la dérivée directionnelle satisfait $f'(x; y - x) \leq f(y) - f(x)$ grâce à la relation de monotonie précédente et, dans le cas différentiable $f'(x; d) = \langle \nabla f(x), d \rangle$. La réciproque est vraie comme l'indique le théorème suivant :

Théorème 3.2 *Si f est différentiable sur l'ouvert convexe C, alors les 3 affirmations suivantes sont équivalentes :*

$$f \textit{ convexe sur } C$$

$$\forall x, y \in C \quad \langle \nabla f(x), y - x \rangle \leq f(y) - f(x) \tag{3.1}$$

$$\forall x, y \in C \quad \langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0 \tag{3.2}$$

Démonstration Montrons que (3.1) implique que f est convexe. Cette inégalité peut s'écrire :

$$\forall y \in C, \sup_{x \in C}\{f(x) + \langle \nabla f(x), y - x \rangle \leq f(y)$$

qui est en fait une égalité car on peut échanger les rôles de x et y. Donc, f est convexe en tant que supremum d'une famille de fonctions affines.

Montrons maintenant que (3.2) implique (3.1) par l'absurde. Supposons un segment $[x, y]$ tel que $\langle \nabla f(x), y - x \rangle > f(y) - f(x)$. Grâce au théorème de Rolle, il existe un $z = (1 - \lambda)x + \lambda y, \lambda \in [0, 1]$, satisfaisant $f(y) = f(x) + \langle \nabla f(z), y - x \rangle$. On en déduit :

$$\langle \nabla f(z) - \nabla f(x), z - x \rangle = \langle \nabla f(z) - \nabla f(x), \lambda(y - x) \rangle < 0$$

ce qui contredit (3.2) pour le couple (x, z).
Les autres implications sont immédiates. □
Monotonie de l'opérateur Gradient : la relation (3.2) signifie que l'opérateur Gradient $\nabla f : C \mapsto I\!R^n$ est monotone sur C.
Fonctions fortement convexes différentiable :

Proposition 3.3 *: Soit une fonction f de $I\!R^n$ dans $I\!R$ fortement convexe de constante α sur un ensemble convexe C et différentiable sur C ; alors :*

$$\forall x, y \in C, \langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \alpha \|x - y\|^2$$

Corollaire 3.1 (Fonctions convexes deux fois différentiables) *Si f est deux fois différentiable sur C, f est convexe si et seulement si le Hessien $\nabla^2 f(x)$ est semi-définie positif sur C.*

Démonstration : Exercice.
Exercice : Montrer le résultat suivant :
Si f est deux fois différentiable et fortement convexe (de paramètre α), alors la matrice $\nabla^2 f(x) - \alpha I$ est définie positive.

3.4 Supports affines et fonctions conjuguées

On note Γ_0 l'ensemble des fonctions convexes propres et sci. Ce sont les fonctions dont l'épigraphe est un ensemble convexe fermé. D'aprés la représentation externe de ces ensembles abordée au chapitre 1, ces épigraphes coïncident avec l'intersection de tous les demi-espaces qui les contiennent. On peut remarquer sur la figure 3.3 ci-dessus qu'un tel demi-espace (associé à un hyperplan non vertical) est l'épigraphe d'une fonction affine qui minore f.

On a d'ailleurs le théorème fondamental suivant :

Théorème 3.3 *: Une fonction f appartient à Γ_0 si et seulement si f est l'enveloppe supérieure des fonctions affines qui la minorent.*

Démonstration Il est clair qu'une fonction qui est l'enveloppe supérieure de fonctions affines de $I\!R^n$ est dans Γ_0.

Soit $f \in \Gamma_0$; puisque f est propre, il existe au moins un x tel que $f(x)$ est fini. On en déduit que le point $\begin{pmatrix} x \\ z \end{pmatrix}$ avec $z < f(x)$ n'appartient pas à epi(f). On peut donc le séparer par un hyperplan qui est clairement non vertical. Il correspond donc à l'épigraphe d'une fonction affine qui minore f. Supposons que f ne soit pas l'enveloppe supérieure de ses minorantes affines. Cela implique qu'il existe $\bar{x} \in \mathrm{Dom}(f)$ tel que $f(\bar{x}) > \bar{z}$, où $\bar{z} = \sup\{\langle a_i, \bar{x} \rangle + b_i, i \in I\}$, I représentant ici les indices des fonctions affines qui minorent

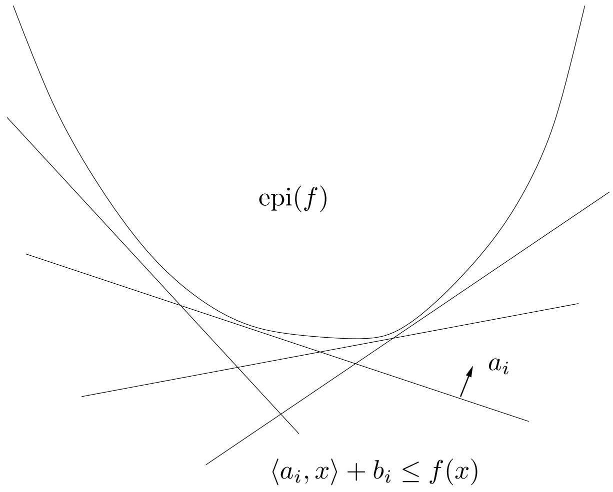

FIGURE 3.3 – Fonctions affines minorantes d'une fonction convexe

f. A nouveau, on peut séparer $\begin{pmatrix} \bar{x} \\ \bar{z} \end{pmatrix}$ et l'épigraphe de f pour obtenir une minorante de f contredisant la définition de $\bar{z}$. $\square$

Si $f \notin \Gamma_0$, cette enveloppe supérieure notée cl(conv$f$) est la plus grande fonction convexe sci qui minore f.

On remarque que ces demi-espaces définissent des hyperplans frontières non verticaux à condition de se placer à l'intérieur du domaine ; on va maintenant les construire en utilisant l'outil suivant :

Définition 3.7 *: On appelle fonction support d'un ensemble C de $\mathbb{R}^n$ la fonction σ_C définie par :*

$$\sigma_C(y) = \sup_{x \in C} \langle y, x\rangle$$

Remarquons que σ_C est convexe et semi-continue inférieurement, en tant que supremum d'une famille de fonctions linéaires. σ_C est aussi positivement homogène, c'est-à-dire $\sigma_C(ay) = a\sigma_C(y), \forall a \geq 0$. Observez que son épigraphe est un cône convexe. On citera de plus deux propriétés remarquables dont les démonstrations sont laissées en exercice :

1. $C_1 \subset C_2 \Longrightarrow \sigma_{C_1}(y) \leq \sigma_{C_2}(y), \forall y$
2. $\sigma_C = \sigma_{\mathrm{conv}\{C\}} = \sigma_{\mathrm{cl}(\mathrm{conv}(C))}$

Considérons pour une certaine fonction f de Γ_0 le calcul de $\sigma_{\mathrm{epi}(f)}$ dans $\mathbb{R}^{n+1}$ pour la direction (non horizontale) $Y = (y_0, -1)^T$. Supposons que $Y \in \mathrm{Dom}(\sigma_{\mathrm{epi}(f)})$; il existe alors $x_0 \in \mathrm{Dom}(f)$ tel que :

$$\sigma_{\mathrm{epi}(f)}(Y) = \langle y_0, x_0\rangle - f(x_0)$$

Cette correspondance définit une fonction de $\mathbb{R}^n$ appelée *fonction conjuguée* de f :

Définition 3.8 *: On appelle fonction conjuguée d'une fonction f de $\mathbb{R}^n$ la fonction f^* définie par :*

$$f^*(y) = \sup_{x \in Dom(f)} \langle y, x\rangle - f(x)$$

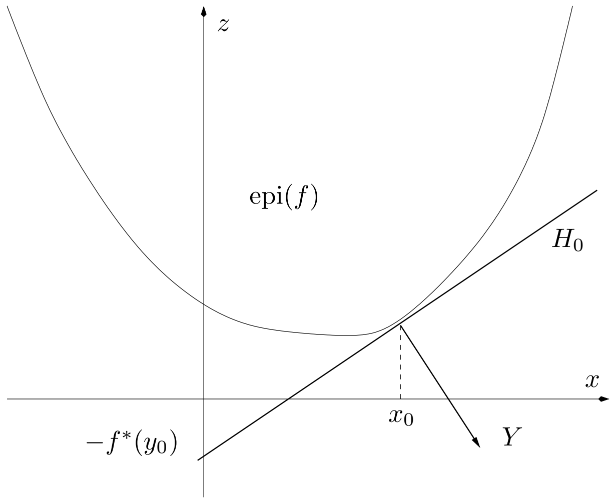

FIGURE 3.4 – Conjuguée et support de l'épigraphe

Donc, $\sigma_{\text{epi}(f)}(Y) = \langle y_0, x_0\rangle - f(x_0) = f^*(y_0)$ (Fig. 3.4).

Pour chaque $x \in \text{Dom}(f)$, $\langle y, x\rangle - f(x)$ définit une fonction affine de la variable y. On en déduit que $f^* \in \Gamma_0$. La dualité inhérente à la propriété de convexité consiste à identifier les fonctions telles que $(f^*)^* = f$:

Théorème 3.4 *: $f \in \Gamma_0$ si et seulement si $f^{**} = f$*

Démonstration On doit seulement démontrer l'implication directe : l'hyperplan associé au calcul de $f^*(y)$, H_0, est un hyperplan support de l'épigraphe de f :

$$H_0 = \{\begin{pmatrix} x \\ z \end{pmatrix} \mid \langle y_0, x\rangle - z = \langle y_0, x_0\rangle - f(x_0)\}$$

Donc, $f^*(y_0) \geq \langle y_0, x\rangle - f(x), \forall x \in \text{Dom}(f)$.

Cette inégalité, appelée inégalité de Fenchel peut s'écrire :

$$f(x) \geq \langle x, y\rangle - f^*(y), \forall y \in \text{Dom}(f^*) \text{ et } \forall x \in \text{Dom}(f)$$

En prenant le supremum des termes de droite, on trouve que $f^{**} = \text{cl}(\text{conv} f)$. Donc, d'après le théorème précédent, si $f \in \Gamma_0, f^{**} = f$. $\square$

Exemples de fonctions conjuguées

a) Soit χ_C la fonction indicatrice d'un ensemble C de $\mathbb{R}^n$ définie par :

$$\chi_C(x) = \begin{cases} 0 & \text{si } x \in C \\ +\infty & \text{sinon} \end{cases}$$

χ_C est convexe si et seulement si C est convexe.

$$(\chi_C)^*(y) = \sup\{\langle y, x\rangle - \chi_C(x)\} = \sup_{x \in C}\langle y, x\rangle = \sigma_C(y)$$

La fonction support est la conjuguée de l'indicatrice.

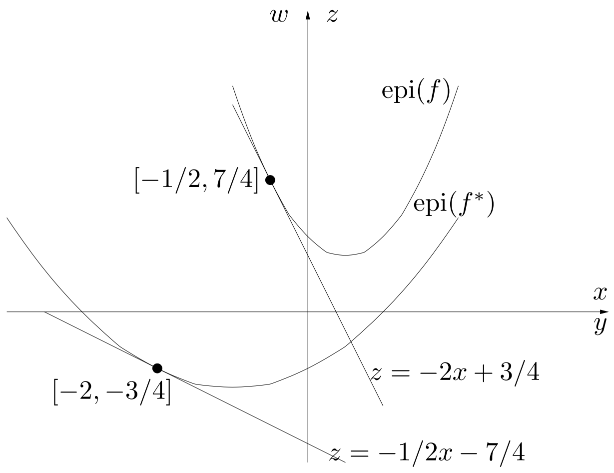

FIGURE 3.5 – Exemple e)

b) $f(x) = 1/p\|x\|^p \longrightarrow f^*(y) = 1/q\|y\|^q$, où $\frac{1}{p} + \frac{1}{q} = 1$

c) Dans $I\!R^n$, $f(x) = \langle a, x\rangle + b \Longrightarrow f^*(y) = -b + \chi_{\{a\}}(y)$

d) Soient f_1 et f_2 deux fonctions convexes de $I\!R^n$ et soit $f = f_1 \Box f_2$. Alors

$$f^* = f_1^* + f_2^*$$

En effet,
$$\begin{array}{rcl}(f_1 \Box f_2)^*(y) & = & \sup_x\{\langle y, x\rangle - \inf\{f_1(x_1) + f_2(x_2) \mid x_1 + x_2 = x\}\} \\ & = & \sup_x \sup\{\langle y, x\rangle - f_1(x_1) - f_2(x_2) \mid x_1 + x_2 = x\} \\ & = & \sup\{\langle y, x_1\rangle + \langle y, x_2\rangle - f_1(x_1) - f_2(x_2) \mid x_1, x_2\} \\ & = & f_1^*(y) + f_2^*(y)\end{array}$$

e) Dans $I\!R$, $f(x) = x^2 - x + 1 \Longrightarrow f^*(y) = \sup\{x.y - x^2 + x - 1\} = 1/4(y + 1)^2 - 1$

3.5 Sous-différentiabilité

Si on reprend l'équation de l'hyperplan support H_0, on a la relation :

$$\forall x \in \mathrm{Dom}(f), \langle y_0, x\rangle - f(x) \leq \langle y_0, x_0\rangle - f(x_0)$$

On dit alors que y_0 est un sous-gradient de f en x_0.

Définition 3.9 *On appelle sous-gradient d'une fonction convexe f en x_0 tout vecteur γ de $I\!R^n$ tel que :*

$$\forall x \in Dom(f), f(x) \geq f(x_0) + \langle \gamma, x - x_0\rangle$$

L'ensemble des sous-gradients en x_0 est le sous-différentiel de f en x_0, noté $\partial f(x_0)$ et on dira que f est sous-différentiable en x_0 s'il existe au moins un sous-gradient.

Exemple :

Dans $I\!R$, $f(x) = \max\{ax + b, cx + d\}$ où $a < 0$ et $c > 0$.

On notera $x_0 = \frac{d-b}{a-c}$ le point qui minimise f, et $f_0 = f(x_0)$ (Figure 3.6) L'ensemble des sous-gradients est le segment $[a, c]$

Des résultats précédents, on tire donc que les 3 conditions suivantes sont équivalentes :

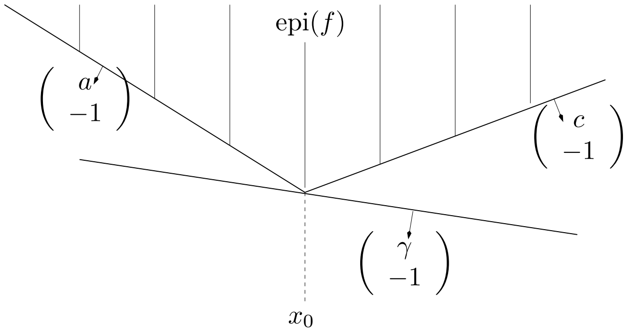

FIGURE 3.6 – Support affine et sous-gradient : $\gamma \in [a, c]$

i) $y_0 \in \partial f(x_0)$

ii) $x_0 \in \partial f^*(y_0)$

iii) $f(x_0) + f^*(y_0) = \langle x_0, y_0 \rangle$

On va voir maintenant que ces sous-gradients généralisent le concept de gradient dans certains cas non différentiables.

On a vu plus haut que les dérivées directionnelles existent en tout point intérieur au domaine et dans toute direction. Observons que $f'(x; \cdot)$ est une fonction positivement homogène et sous-additive, i.e. :

$$\begin{array}{rcl} f'(x; ad) & = & af'(x; d) \ \text{ si } a \geq 0 \\ f'(x; d + d') & \leq & f'(x; d) + f'(x; d') \end{array}$$

Elle est donc la fonction support d'un ensemble qui n'est autre que le sous-différentiel.

En effet, soit $g \in \partial f(x_0)$. On a donc $f(x_0 + td) \geq f(x_0) + t\langle g, d \rangle$, ce qui implique :

$$f'(x_0; d) \geq \langle g, d \rangle, \ \forall g \in \partial f(x_0), \ \text{ et } \forall d \in \text{aff}\{\text{Dom}(f)\} - \{x_0\}$$

Le sous-différentiel est donc l'ensemble (convexe compact) dont la fonction support est la dérivée directionnelle, soit :

$$f'(x_0; d) = \max\{\langle g, d \rangle \mid g \in \partial f(x_0)\}$$

Une conséquence importante est que, si f est différentiable en x_0, le sous-différentiel $\partial f(x_0)$ se réduit à un élément, le gradient de f en x_0.

$$f \textbf{ différentiable en } x_0 \quad \Longleftrightarrow \partial f(x_0) = \{\nabla f(x_0)\}$$

On rejoint alors l'inégalité (3.1) trouvée à la section 3.2 qui exprime que la tangente se situe au dessous du graphe.

Exemples de calcul du sous-différentiel :

a) Soit un ensemble C convexe fermé de $\mathbb{R}^n$ et soit χ_C sa fonction indicatrice. On a alors : si $x \in C, \partial \chi_C(x) = \{g \in \mathbb{R}^n | \sigma_C(g) = \langle g, x \rangle\} = N_C(x)$, le cône normal à C en x.

b) Dans $\mathbb{R}$, $f(x) = \begin{cases} x^2 & \text{si } x < 0 \\ 2x & \text{si } x \geq 0 \end{cases}$

f n'est pas différentiable en 0 et $\partial f(0) = [0, 2]$.
c) Soit $f(x) = \max\{\langle a_i, x\rangle + b_i,\ i \in I\}$ une fonction affine par morceaux qui est bien convexe et sci. On se propose de calculer le sous-différentiel en un point x_0 et soit $I(x_0) = \{i \in I \mid f(x_0) = \langle a_i, x_0\rangle + b_i\}$. On peut alors réécrire f comme :

$$f(x) = f(x_0) + \max\{\langle a_i, x - x_0\rangle - e_i, i \in I\}$$

où $e_i = f(x_0) - \langle a_i, x_0\rangle - b_i \geq 0$ (donc $e_i = 0, \forall i \in I(x_0)$). On en déduit, pour t suffisamment petit :

$$f(x_0 + td) = f(x_0) + t \max\{\langle a_i, d\rangle,\ i \in I(x_0)\}$$

Donc $f'(x_0; d) = \max_{i \in I(x_0)} \langle a_i, d\rangle$ et d'après les résultats précédents sur la dérivée directionnelle et le fait que $\sigma_{a_i, i \in I(x_0)} = \sigma_{\text{conv}\{a_i, i \in I(x_0)\}}$, on obtient :

$$\partial f(x_0) = \text{conv}\{a_i, i \in I(x_0)\}$$

On peut généraliser ce résultat au cas d'une fonction 'max' comme $f(x) = \max\{f_i(x), i \in I\}$ où les f_i sont convexes, différentiables. En $x_0 \in \bigcap_{i \in I} \text{Dom}(f_i)$, on a :

$$\partial f(x_0) = \text{conv}\{\nabla f_i(x_0), i \in I(x_0)\}$$

Ce dernier exemple est un cas particulier du théorème suivant, dû à Danskin :

Théorème 3.5 *Soit Z un sous-ensemble compact de $\mathbb{R}^m$ et $\Phi : \mathbb{R}^n \times Z \mapsto \mathbb{R}$ une fonction continue, différentiable et convexe par rapport à son premier argument $x \in \mathbb{R}^n$ et ce, pour tout $z \in Z$. Soit :*

$$f(x) = \max_{z \in Z} \ \Phi(x, z)$$

et soit $Z(x)$ l'ensemble des z qui maximisent $\Phi(x, z)$ sur Z. Alors :

i) f est convexe et $f'(x; d) = \max_{z \in Z(x)} \Phi'(x, z; d)$

ii) $\partial f(x) = conv\{\nabla_x \Phi(x, z), z \in Z(x)\}$

Démonstration : cf. Hiriart-Urruty et Lemarechal, 1993

3.6 Optimalité

Quand on cherche à minimiser une fonction de $\mathbb{R}^n$, on distingue minimum global de minimum local : un minimum global de f est un point $m \in \text{Dom}(f)$ tel que $f(x) \geq f(m), \forall x \in \text{Dom}(f)$; pour définir un minimum local, on remplace $x \in \text{Dom}(f)$ par $x \in \text{Dom}(f) \cap V_\epsilon(m)$, où $V_\epsilon(m)$ est un voisinage de m de rayon ϵ, i.e. $V_\epsilon(m) = \{x \mid \|x - m\| \leq \epsilon\}$ pour un $\epsilon > 0$.

Théorème 3.6 *Soit f une fonction convexe de $\mathbb{R}^n$ et m un point où f est sous-différentiable. Une condition nécessaire et suffisante pour qu'un point x soit un minimum global d'une fonction convexe est $0 \in \partial f(m)$.*

Démonstration Conséquence immédiate de la définition d'un sous-gradient.

$$f(x) \geq f(m) \Longleftrightarrow f(x) \geq f(m) + \langle 0, x - m \rangle$$

□

On retrouve dans le cas différentiable la condition nécessaire d'optimalité du premier ordre : si m minimise f, le gradient de f en m est nul. Cette relation est locale, alors que **tout minimum local d'une fonction convexe est un minimum global**. Dans le cas convexe, la condition d'optimalité du premier ordre est nécessaire et suffisante (Observez que la condition du deuxième ordre est satisfaite en tout point car le Hessien d'une fonction convexe deux fois différentiable est semi-défini positif).

Références :

C. Berge et A. Ghouila-Houri, Programmes, Jeux et Transport, Dunod, 1962

J.B. Hiriart-Urruty et C. Lemarechal, Convex Analysis and Minimization Algorithms, Springer V., 1993

P.J. Laurent, Approximation et Optimisation, Hermann, 1972

R.T Rockafellar, Convex Analysis, Princeton U. , 1970

J. Stoer et C. Witzgall, Convexity and Optimization, Springer V., 1970