

Chapitre 1

Généralités

L'optimisation (c'est-à-dire les techniques permettant de chercher les minima ou les maxima de fonctions ou de fonctionnelles) intervient dans pratiquement tous les processus de modélisation actuels. Qu'il s'agisse de problèmes directs (ajustement de données, contrôle optimal, résolution de systèmes linéaires par moindres carrés, etc . . .) ou inverses (identification de paramètres, contrôle de frontières libres etc..), il est rare qu'un problème d'optimisation plus ou moins complexe n'intervienne pas à un stade donné de la modélisation et/ou de la simulation. Avant de donner les définitions et principes de base de la théorie de l'optimisation, nous allons présenter quelques exemples simples permettant d'introduire et d'illustrer par anticipation notre propos.

1.1 Quelques exemples

1.1.1 Détermination de coefficients en combustion

On considère un mélange de gaz et on appelle T le taux d'introduction de chaleur dans ce mélange. On appelle α la variable (par exemple la proportion d'un des gaz dans le mélange). La loi donnant T en fonction de α est de la forme

$$T(\alpha) = f[1 - e^{-a_1 e^{\alpha b_1}}] + (1 - f)[1 - e^{-a_2 e^{\alpha b_2}}] \quad (1.1.1)$$

où les paramètres à déterminer sont $(f, a_1, b_1, a_2, b_2) \in \mathbb{R}^5$. Comme indiqué précédemment, on fait n mesures de T pour différentes valeurs de α de sorte que $T(\alpha_i) \simeq T_i$, $i = 1, \dots, n$. On obtient alors la formulation au sens des moindres carrés du problème :

$$\min \sum_{i=1}^n [f(1 - e^{-a_1 e^{\alpha_i b_1}}) + (1 - f)(1 - e^{-a_2 e^{\alpha_i b_2}}) - T_i]^2, (f, a_1, b_1, a_2, b_2) \in \mathbb{R}^5. \quad (1.1.2)$$

1.1.2 Un exemple en hydrologie

En hydrologie dans des problèmes de corrélation hydropluviométrique, le débit d'un bassin \mathcal{Y} est une variable aléatoire dépendant linéairement de n variables aléatoires $\mathcal{X}_1, \dots, \mathcal{X}_n$ (pluviosité

moyenne sur le bassin, différents indices de répartition de précipitation dans le temps etc...), selon la relation :

$$\mathcal{Y} = b_0 + b_1\mathcal{X}_1 + b_2\mathcal{X}_2 + \cdots + b_n\mathcal{X}_n. \quad (1.1.3)$$

On doit en général, chercher les coefficients b_0, b_1, \dots, b_n (dits de régression). Pour cela on effectue p observations (ou mesures) portant sur les variables \mathcal{X}_i et \mathcal{Y} ; on note $Y = [y_1, y_2, \dots, y_p]$ le vecteur des valeurs observées de \mathcal{Y} et $X = [x_{ij}]$, $i = 1, \dots, n$, $j = 1, \dots, p$, la matrice observée : x_{ij} est j -ième observation de la variable \mathcal{X}_i . Le problème d'identification des paramètres b_i se formule alors de la manière suivante

$$\min \sum_{j=1}^p \left[y_j - \left(b_0 + \sum_{i=1}^n x_{ij} b_i \right) \right]^2, \quad (b_0, b_1, \dots, b_n) \in \mathbb{R}^{n+1}. \quad (1.1.4)$$

Toutefois, si on résout le problème de minimisation ci-dessus on risque de trouver des débits négatifs ce qui n'a pas de sens ! On impose donc une condition (ou contrainte) supplémentaire sur les débits qui doivent être positifs. Le problème se modélise alors comme suit

$$\begin{cases} \min \sum_{j=1}^p \left[y_j - \left(b_0 + \sum_{i=1}^n x_{ij} b_i \right) \right]^2 \\ (b_0, b_1, \dots, b_n) \in \mathbb{R}^{n+1}, \\ b_0 + \sum_{i=1}^n b_i x_{ij} \geq 0. \end{cases} \quad (1.1.5)$$

1.1.3 Un exemple en chimie : problème de l'équilibre chimique

On considère un mélange de m éléments chimiques. On a prédéterminé que les m atomes différents peuvent se combiner pour produire n composés. Soit x_j le nombre de moles du composé j (i.e. Nx_j est le nombre de molécules du composé j , où N est le nombre d'Avogadro), a_{ij} est le nombre d'atomes d'un élément i dans une molécule de composé j et Nb_i le nombre d'atomes de l'élément i dans le mélange. On veut identifier x_j c'est-à-dire la composition exacte du mélange. L'équation de bilan des masses donne un premier type de contraintes

$$\sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, \dots, m$$

auxquelles s'ajoutent des contraintes naturelles

$$x_j \geq 0, \quad j = 1, \dots, n.$$

D'autre part, le second principe de la thermodynamique nous apprend qu'un mélange de composés chimiques à température et à pression constantes atteint son état d'équilibre lorsque l'énergie libre du système est minimale (Principe de GIBBS). Cette énergie libre est donnée par

$$f(x_1, \dots, x_n) = \sum_{j=1}^n x_j \left[c_j + \ln \left(\frac{x_j}{\sum_{j=1}^n x_j} \right) \right],$$

avec $c_j = \frac{F_j^0}{RT} + \ln P$, F_j^0 désignant l'énergie libre de GIBBS par mole du composé j à la température T et à la pression d'une atmosphère, P est la pression totale et R la constante des gaz parfaits.

Le problème revient donc à minimiser f sous contraintes

$$\left\{ \begin{array}{l} \min \sum_{j=1}^n x_j \left[c_j + \ln \left(\frac{x_j}{\sum_{j=1}^n x_j} \right) \right] \\ \sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, \dots, m, \\ x_j \geq 0, \quad j = 1, \dots, n \dots \end{array} \right. \quad (1.1.6)$$

Remarquons que ce problème n'est pas un problème formulé au sens des moindres carrés comme dans les exemples précédents. Toutefois, c'est encore un problème de minimisation d'une fonction de plusieurs variables.

1.2 Formulation mathématique

Les exemples précédents peuvent tous s'écrire sous la forme générale suivante

$$(\mathcal{P}) \quad \left\{ \begin{array}{l} \min J(x) \\ g(x) \leq 0, \\ h(x) = 0, \\ x \in \mathbb{R}^n \end{array} \right.$$

où

- $J : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction de plusieurs variables ($x = (x_1, \dots, x_n)$) **à valeurs réelles**. Cette fonction (que l'on minimise) est appelée indifféremment fonction **coût**, **objectif** ou **critère**.
- $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est une fonction de plusieurs variables $x \in \mathbb{R}^n$ à valeurs dans \mathbb{R}^p : elle a p composantes et on peut écrire

$$g(x) = (g_1(x), \dots, g_p(x)),$$

chaque fonction g_i étant définie sur \mathbb{R}^n et à valeurs dans \mathbb{R} . La fonction g représente les **contraintes en inégalité**. La notation $g(x) \leq 0$ signifie qu'on considère les inégalités composante par composante : $g(x) \leq 0 \stackrel{\text{def}}{\iff} \forall i = 1, \dots, p \quad g_i(x) \leq 0$.

- $h : \mathbb{R}^n \rightarrow \mathbb{R}^q$ est une fonction de plusieurs variables $x \in \mathbb{R}^n$ à valeurs dans \mathbb{R}^q : elle a q composantes et on peut écrire

$$h(x) = (h_1(x), \dots, h_q(x)),$$

chaque fonction h_i étant définie sur \mathbb{R}^n et à valeurs dans \mathbb{R} . La fonction h représente les **contraintes en égalité**.

Remarque 1.2.1 Plus généralement, on peut remplacer l'espace de dimension finie \mathbb{R}^n par un espace vectoriel topologique sur \mathbb{R} (de dimension a priori infinie). Nous nous bornerons à étendre (quand ce n'est pas trop compliqué) le cadre fonctionnel aux espaces de Hilbert réels.

Précisons maintenant que qu'on entend par minimisation (ou maximisation) d'une fonction. Soit \mathcal{C} l'ensemble des contraintes, c'est-à-dire par exemple dans le cas précédent

$$\mathcal{C} = \{ x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0 \}.$$

On suppose que \mathcal{C} est non vide ; un élément x de \mathcal{C} sera dit **réalisable**.

Définition 1.2.1 (Minimum (maximum) local)

Soient \mathcal{C} un ensemble non vide d'un espace de Hilbert réel \mathbb{H} et f une fonction de \mathcal{C} dans \mathbb{R} . On dit que $x^* \in \mathcal{C}$ réalise un **minimum local** de f sur \mathcal{C} si on peut trouver une boule de \mathbb{H} centrée en $x^* : \mathcal{B}(x^*)$ telle que

$$\forall x \in \mathcal{B}(x^*) \cap \mathcal{C} \quad f(x^*) \leq f(x).$$

On dit que $x^* \in \mathcal{C}$ réalise un **maximum local** de f sur \mathcal{C} si on peut trouver une boule de \mathbb{H} centrée en $x^* : \mathcal{B}(x^*)$ telle que

$$\forall x \in \mathcal{B}(x^*) \cap \mathcal{C} \quad f(x^*) \geq f(x).$$

On rappelle qu'une boule de \mathbb{H} centrée en x^* de rayon $\rho > 0$ est l'ensemble

$$\mathcal{B}(x^*, \rho) = \{ x \in \mathbb{H} \mid \|x - x^*\| \leq \rho \},$$

où $\|\cdot\|$ désigne la norme de \mathbb{H} .

Définition 1.2.2 (Minimum (maximum) global)

On dit que $x^* \in \mathcal{C}$ réalise un **minimum global** de f sur \mathcal{C} si $\forall x \in \mathcal{C} \quad f(x^*) \leq f(x)$.

On dit que $x^* \in \mathcal{C}$ réalise un **maximum global** de f sur \mathcal{C} si $\forall x \in \mathcal{C} \quad f(x^*) \geq f(x)$.

Nous donnons ci-dessous une illustration des différents cas de figure.

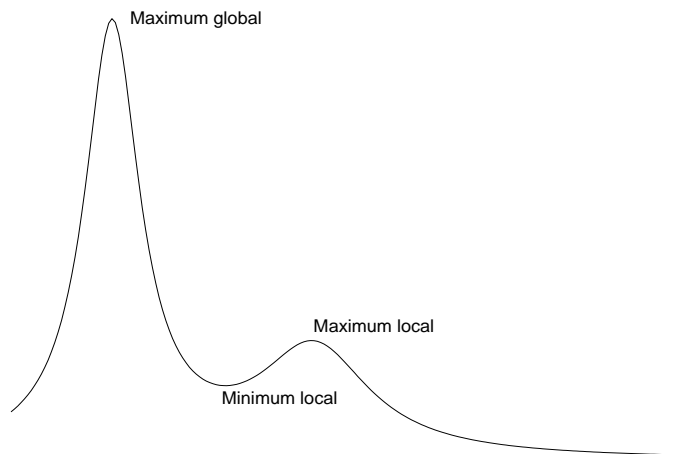


Figure 1.1a : Exemples de minima et de maxima locaux et globaux

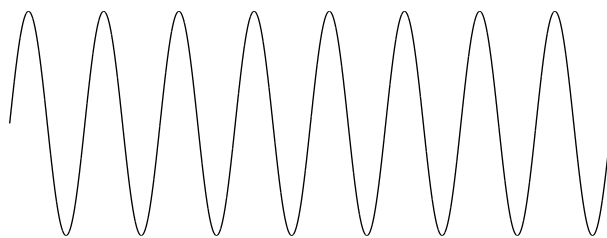


Figure 1.1b : Infinité de maxima et minima globaux

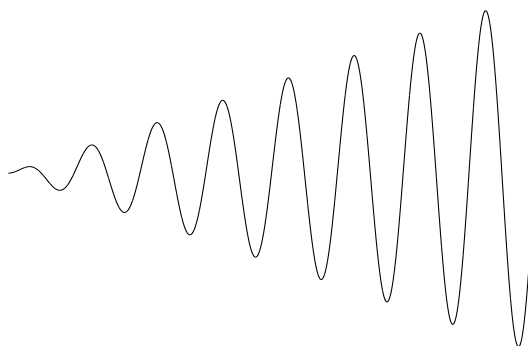


Figure 1.1c : Pas de maximum global - Infinité de maxima et minima locaux

Les minima et maxima sont dits **stricts** si les inégalités dans les définitions précédentes sont strictes. On s'intéressera essentiellement à la recherche des points réalisant des minima car la recherche des maxima peut se ramener à celle des minima comme le montre la proposition suivante :

Proposition 1.2.1 *Si x^* réalise un maximum (local ou global) de f sur \mathcal{C} , x^* réalise un minimum (local ou global) de $-f$ sur \mathcal{C} . Plus précisément*

$$\max \{ f(x), x \in \mathcal{C} \} = - \min \{ -f(x), x \in \mathcal{C} \}.$$

Démonstration - Donnons la preuve pour un maximum global : c'est exactement la même pour un maximum local.

Soit x^* tel que $f(x^*) = \max \{ f(x), x \in \mathcal{C} \}$. On a donc

$$\begin{aligned} \forall x \in \mathcal{C} \quad f(x) &\leq f(x^*) \\ \Leftrightarrow \forall x \in \mathcal{C} \quad -f(x) &\geq -f(x^*) \\ \Leftrightarrow -f(x^*) &= \min \{ -f(x), x \in \mathcal{C} \} \\ \Leftrightarrow f(x^*) &= - \min \{ -f(x), x \in \mathcal{C} \}. \end{aligned}$$

□

Remarque 1.2.2 *Par abus de langage, on dit souvent que x^* est un minimum pour la fonction J ou de la fonction J : il faudrait dire que x^* réalise un minimum pour J ou que $J(x^*)$ est une valeur minimale de J .*

On ne peut évidemment résoudre le problème général sans quelques hypothèses sur J , g et h qui permettront au moins d'assurer l'existence de solutions. Nous précisons ces hypothèses dans le prochain chapitre, mais nous avons besoin des définitions suivantes.

1.3 Notion de convexité

1.3.1 Définitions

Le cas où les données sont convexes est un cas très important car les problèmes quadratiques sont à la base de nombreux algorithmes non linéaires. Nous rappelons quelques définitions et propriétés. Dans tout ce qui suit \mathbb{H} désigne un espace de Hilbert réel. On désigne par $\|\cdot\|$ sa norme et (\cdot, \cdot) son produit scalaire.

Définition 1.3.1 (Ensemble convexe)

On dit que l'ensemble $\mathcal{C} \subset \mathbb{H}$ est **convexe** si

$$\forall (x, y) \in \mathcal{C} \times \mathcal{C}, \forall t \in [0, 1] \quad tx + (1 - t)y \in \mathcal{C}.$$

Autrement dit, \mathcal{C} est convexe s'il contient tout "segment" reliant deux quelconques de ses points.

Exemple 1.3.1 (Ensembles convexes)

1. Un intervalle $[a, b]$ est convexe dans \mathbb{R} .
2. Une réunion disjointe d'intervalles de \mathbb{R} n'est pas convexe. (\mathbb{R}^* par exemple).

Définition 1.3.2 (Fonction convexe)

On dit que la fonction $J : \mathcal{C} \subset \mathbb{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ est **convexe** si \mathcal{C} est convexe et si

$$\forall (x, y) \in \mathcal{C} \times \mathcal{C}, \forall t \in [0, 1] \quad J(tx + (1 - t)y) \leq tJ(x) + (1 - t)J(y).$$

Définition 1.3.3 (Domaine d'une fonction convexe)

Soit $J : \mathbb{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe. On appelle **domaine** de J l'ensemble

$$\text{dom } J = \{ x \in \mathbb{H} \mid J(x) < +\infty \}.$$

Cet ensemble est convexe.

Lorsque le domaine de J est non vide J est dite **propre**.

Définition 1.3.4 (Fonction strictement convexe)

On dit que la fonction $J : \mathcal{C} \rightarrow \mathbb{R} \cup \{+\infty\}$ est **strictement convexe** si \mathcal{C} est convexe et si

$$\forall (x, y) \in \mathcal{C} \times \mathcal{C} \text{ avec } x \neq y, \forall t \in]0, 1[\quad J(tx + (1 - t)y) < tJ(x) + (1 - t)J(y).$$

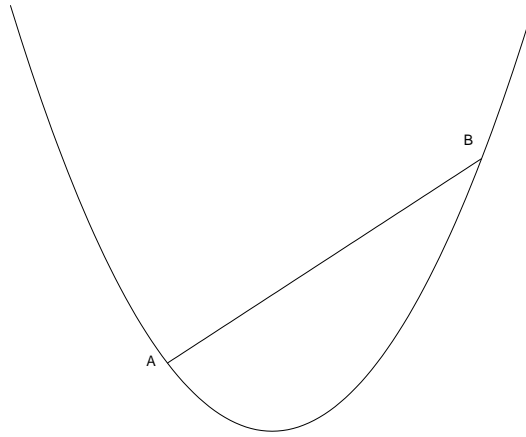


Figure 1.2 : Exemple de fonction convexe
(la corde AB est au-dessus de l'arc AB)

Exemples 1.3.1 Fonctions convexes et strictement convexes

- $J(x) = \|x\|^2$ est strictement convexe.
- Toute application affine, c'est-à-dire de la forme

$$J(x) = (b, x) + \beta ,$$

où b et x sont des éléments de \mathbb{H} et $\beta \in \mathbb{R}$ est convexe mais pas strictement.

- Soit A une matrice carrée symétrique d'ordre n semi-définie positive et b un vecteur de \mathbb{R}^n . Alors J définie par

$$J(x) = \frac{1}{2} (Ax, x)_n - (b, x)_n ,$$

est convexe. Si de plus A est définie positive, J est strictement convexe.

$(\cdot, \cdot)_n$ désigne le produit scalaire de \mathbb{R}^n .

Plus généralement si \mathcal{A} est un opérateur linéaire de \mathbb{H} (espace de Hilbert) dans \mathbb{H} , auto-adjoint et **monotone** c'est-à-dire

$$\forall (x, y) \in \mathbb{H} \times \mathbb{H} \quad (\mathcal{A}(x) - \mathcal{A}(y), x - y) \geq 0 ,$$

et $b \in \mathbb{H}$, alors J définie par

$$J(x) = \frac{1}{2} (\mathcal{A}x, x) - (b, x) ,$$

est convexe.

Les figures suivantes donnent quelques exemples de fonctions non convexes.

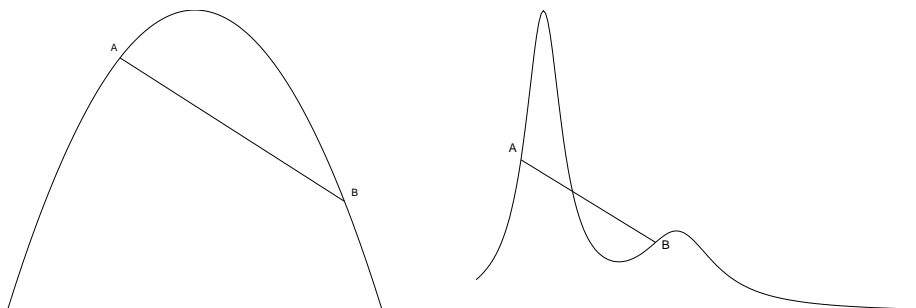


Figure 1.3 : Exemple de fonctions non convexes
(la corde AB n'est pas au-dessus de l'arc)

La fonction f apparaissant à gauche dans la figure 1.3 est telle que son opposée $-f$ est convexe : une telle fonction est dite **concave**. Le graphe de droite montre qu'une fonction peut n'être ni convexe, ni concave.



Figure 1.4 : Exemples de fonctions convexes, non strictement convexes

Définition 1.3.5 (Programmation convexe)

On dit que le problème (\mathcal{P}) est un problème de **programmation convexe** quand les fonctions $J, g_i, i = 1, \dots, p$ sont convexes et les fonctions $h_j, j = 1, \dots, q$ sont affines.

Définition 1.3.6 (Programmation linéaire)

On dit que le problème (\mathcal{P}) est un problème de **programmation linéaire** quand les fonctions $J, g_i, i = 1, \dots, p, h_j, j = 1, \dots, q$ sont affines.

Le cas de la programmation linéaire est certes un cas particulier de la programmation convexe mais il se présente plutôt comme un cas singulier pour lequel on n'est pas toujours sûr de trouver des solutions. De ce fait, les méthodes employées pour la résolution de ces problèmes sont des méthodes très spécifiques et non pas des cas particuliers des méthodes de programmation non linéaire que nous allons présenter. La résolution des problèmes de programmation linéaire relève de la Recherche Opérationnelle dont nous ne parlerons pas ici. On pourra par exemple consulter [10] à ce sujet.

1.3.2 Continuité des fonctions convexes

Donnons maintenant quelques propriétés (topologiques) importantes des fonctions convexes. Dans tout ce qui suit J est une fonction de \mathbb{H} vers $\mathbb{R} \cup \{+\infty\}$. On suppose que le domaine de J est non vide.

Théorème 1.3.1 (Continuité)

Soit $J : \mathbb{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ convexe. Il est équivalent de dire

- i) Il existe un ouvert non vide Ω sur lequel J est majorée par une constante a réelle et ne vaut pas constamment $-\infty$
- ii) J est propre, $\text{dom } J$ est d'intérieur non vide et J est continue sur l'intérieur de son domaine.

Démonstration - Il est clair que (ii) entraîne (i).

Pour démontrer le résultat réciproque nous aurons besoin du lemme suivant

Lemme 1.3.1 Si, au voisinage d'un point $u_o \in \mathbb{H}$, une fonction convexe J est majorée par une constante finie, alors J est continue en u_o .

Nous démontrerons ce lemme ensuite.

Supposons donc (i) vérifié : Ω est donc inclus dans l'intérieur du domaine de J qui est en particulier non vide. Donc J est propre. Soit $u \in \Omega$ tel que $J(u) > -\infty$. D'après le lemme (1.3.1), J sera continue en u , donc finie sur un voisinage de u . Pour tout $v \in \text{int}(\text{dom } J)$, il existe $\rho > 1$ tel que $w = u + \rho(v - u)$ appartienne encore à $\text{int}(\text{dom } J)$ car l'intérieur d'un convexe est convexe (ce que nous admettons). L'homothétie h de centre w et de rapport $1 - \frac{1}{\rho}$ transforme u en v et Ω en un ouvert $h(\Omega)$ contenant v . Pour tout v' de $h(\Omega)$, on a par convexité

$$J(v') \leq \frac{\rho - 1}{\rho} J(h^{-1}(v')) + \frac{1}{\rho} J(w) \leq \frac{\rho - 1}{\rho} a + \frac{1}{\rho} J(w).$$

Par conséquent : tout point v de $\text{int}(\text{dom } J)$ possède un voisinage $h(\Omega)$ sur lequel J est majorée par une constante finie. D'après le lemme (1.3.1), J est continue en v . \square

Démonstration du lemme (1.3.1)

On se ramène par translation au cas où $u_o = 0$ et $J(0) = 0$. Soit \mathcal{V} un voisinage de l'origine tel que $J(u) \leq a < +\infty$, pour tout u de \mathcal{V} . Posons $\mathcal{W} = \mathcal{V} \cap -\mathcal{V}$ et donnons nous $\varepsilon \in]0, 1[$. Si $u \in \varepsilon\mathcal{W}$, on par convexité

$$\begin{aligned} \frac{u}{\varepsilon} \in \mathcal{W}, \text{ donc } J(u) &\leq (1 - \varepsilon)J(0) + \varepsilon J\left(\frac{u}{\varepsilon}\right) \leq \varepsilon a, \\ -\frac{u}{\varepsilon} \in \mathcal{W}, \text{ donc } J(u) &\geq (1 + \varepsilon)J(0) - \varepsilon J\left(-\frac{u}{\varepsilon}\right) \geq \varepsilon a. \end{aligned}$$

Finalement

$$\forall u \in \varepsilon\mathcal{W} \quad |J(u)| \leq \varepsilon a,$$

d'où la continuité. \square

Corollaire 1.3.1 *Toute fonction convexe propre sur un espace de dimension finie ($\mathbb{H} = \mathbb{R}^n$) est continue sur l'intérieur de son domaine.*

Démonstration - Si l'intérieur du domaine de J est non vide, il contient $n + 1$ points affinement indépendants $u_i, 1 \leq i \leq n + 1$. D'après l'inégalité de convexité, J est majorée par $\max_{1 \leq i \leq n+1} J(u_i)$ sur l'ouvert

$$\left\{ \sum_{i=1}^{n+1} \lambda_i u_i \mid \sum_{i=1}^{n+1} \lambda_i = 1 \text{ et } \lambda_i > 0 \forall i \right\} .$$

□

Pour plus de résultats sur les fonctions convexes on peut se référer à [11].

1.3.3 Différentiabilité des fonctions convexes

Donnons maintenant quelques propriétés de différentiabilité.

Définition 1.3.7 *Soit J une fonction de \mathbb{H} dans $\mathbb{R} \cup \{+\infty\}$. On dit que J est **Gâteaux-différentiable** en $u \in \text{dom}(J)$ si la dérivée directionnelle*

$$J'(u; v) = \lim_{t \rightarrow 0^+} \frac{J(u + tv) - J(u)}{t} ,$$

existe dans toute direction v de \mathbb{H} et si l'application

$$v \mapsto J'(u; v)$$

est linéaire continue.

D'après le théorème de représentation de Riesz (voir [5] par exemple) on identifie \mathbb{H} et son dual ; on note alors

$$J'(u; v) = (\nabla J(u), v) ,$$

où $(,)$ désigne le produit scalaire de \mathbb{H} . L'élément $\nabla J(u)$ de \mathbb{H} est le **gradient** de J en u .

Il est clair que si J est différentiable au sens classique en u (on dit alors **Fréchet** - différentiable), alors J est Gâteaux-différentiable en u , et la dérivée classique et la dérivée au sens de Gâteaux coïncident.

La réciproque est fautive comme le montre le contre-exemple suivant : soit f de \mathbb{R}^2 dans \mathbb{R} définie par :

$$f(x, y) = \begin{cases} y & \text{si } x = y^2 , \\ 0 & \text{sinon} \end{cases}$$

La fonction f est continue en $(0,0)$ et Gâteaux-différentiable en $(0,0)$ mais pas Fréchet - différentiable en $(0,0)$.

Théorème 1.3.2 *Soit $J : \mathcal{C} \subset \mathbb{H} \rightarrow \mathbb{R}$, Gâteaux différentiable sur \mathcal{C} , avec \mathcal{C} convexe. J est convexe si et seulement si*

$$\forall (u, v) \in \mathcal{C} \times \mathcal{C} \quad J(v) \geq J(u) + (\nabla J(u), v - u) \quad (1.3.1)$$